Parsing the heterogeneity of Major Depression

Beijers, Lian

DOI:
10.33612/diss.177477548

Parsing the heterogeneity of Major Depression

Biological subtyping and other statistical approaches
to unravel the causes of Major Depression

Lian Beijers

Lian Beijers
Parsing the heterogeneity of Major Depression
Biological subtyping and other statistical approaches to unravel the causes of Major Depression

Cover image & lay-out: Esther Scheide (www.proefschriftomslag.nl)
Printing: ProefschriftMaken
CV photo: Gerard Scheers (www.simpelconcept.nl)

Paranymphs: Hanne Beijers and Loes Beijers

Publication of this dissertation was financially supported by the University Medical Center Groningen, the University of Groningen, and the Graduate School SHARE of the University Medical Center Groningen.

This thesis is printed on 100% recycled and FSC-certified paper.

university of groningen

Parsing the heterogeneity of Major Depression

Biological subtyping and other statistical approaches
to unravel the causes of Major Depression

PhD thesis

to obtain the degree of PhD at the University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

26 August 2021 at 09:00 hours

by

Lian Beijers

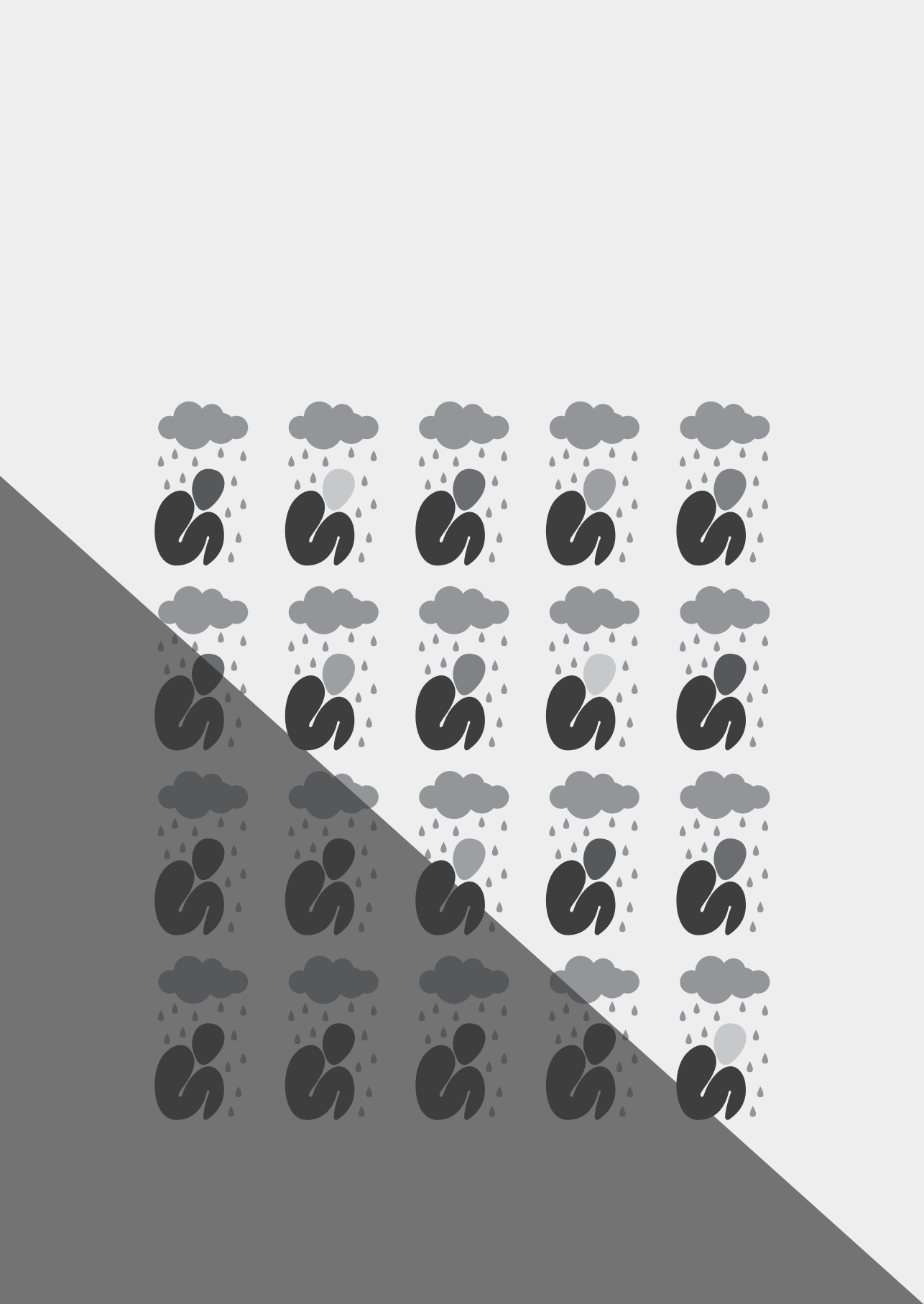
born on 12 September 1990
in Winterswijk

Supervisor
Prof. R. A. Schoevers

Co-supervisors
Dr. K. J. Wardenaar
Dr. H. M. van Loo

Assessment committee
Prof. D. C. Cath
Prof. P. de Jonge
Prof. D. Borsboom

Table of contents



Chapter 1

Introduction

Preface

Suffering is an integral component of the human condition. Nothing in life is permanent, and since even the happiest of moments must come to an end at some point, loss is unavoidable. None of us are exempt from negative emotions like fear, sadness, and grief. Fortunately, these emotional states generally are impermanent as well. As we recover from whatever experience life throws at us, the negative emotions will pass. However, this kind of emotional resilience seems to come more difficult to some people. Through the ages, a certain percentage of the population has been affected by unusually prolonged states of depressed mood - a condition which the ancient Greeks called melancholia.[1,2] It is important to note that feeling melancholic or depressed is not necessarily a sign of pathology or mental illness. These days it is not unusual to hear someone say the news was really depressing in a casual conversation, or that they have been depressed after a recent break-up.

Major Depression

It is difficult to determine exactly where normal variation in mood ends and pathology starts.[3,4] Since there is a continuum of severity and pervasiveness from ordinary sadness to clinical depression, it makes sense for the boundary to be fixed on pragmatic grounds, i.e., giving priority to clinical utility.[5] This is what the American Psychiatric Association (APA) attempts to achieve in their Diagnostic and Statistical Manual of Mental Disorders (DSM), regarding depression as a 'disorder' when it reaches a given threshold in terms of severity, duration and degree of suffering or functional impairment (see Box 1, criterion B), thus deserving clinical attention.[6] The quality of life is often low for people diagnosed with MD, because MD impacts all aspects of a person's life, limiting their ability to function at work and manage daily tasks like cleaning and shopping, and slowly spoiling their social lives and close relationships.[7,8] Indeed, MD is experienced as more disabling than even many physical disorders such as chronic pain, heart disease or even cancer.[9–11] MD is currently the single largest contributor to the global burden of disease, and MD patients are most likely to commit suicide of all patients diagnosed with mental disorders (see Box 2 for some more key figures).[12–14] Currently, Dutch patients presenting with MD at their general practitioner's (GP) office receive problem solving treatment (PST) or guided self-help interventions, or, if their symptoms are too severe, they are referred to a specialist for psychotherapeutic interventions such as cognitive-behavioral therapy (CBT)

or interpersonal psychotherapy (IPT).[15] If the patient does not want psychotherapy or the therapy is not effective enough, the GP can also offer antidepressant medication such as selective serotonin reuptake inhibitors (SSRIs) or tricyclic antidepressants (TCA).[15]

Box 1. Major Depression - diagnostic criteria according to the DSM-5[6]

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others e.g., appears tearful). (Note: In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., A change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (Note: In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear dying), recurrent suicidal ideation without a specific plan, or suicide attempt or specific plan for committing suicide.

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The episode is not attributable to the physiological effects of a substance or to another medical condition.

Note: Criteria A-C represent a major depressive episode.

Note: Responses to a significant loss (e.g., bereavement, financial ruin, losses from a natural disaster, a serious medical illness or disability) may include the feelings of intense sadness, rumination about the loss, insomnia, poor appetite, and weight loss noted in criteria a, which may resemble a depressive episode. Although such symptoms may be understandable or considered appropriate to the loss, the presence of major depressive episode in addition to normal response to a significant loss should also be carefully considered. This decision inevitably requires the exercise of clinical judgment based on individual's history and cultural norms for the expression of distress in the context of loss.

D. The occurrence of the major depressive episode is not better explained by seasonal affective disorder, schizophrenia, schizophrenic form disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.

E. There has never been a manic episode or hypomanic episode.

Note: This exclusion does not apply if all the manic-like or hypomanic-like episodes are substance-induced or are attributable to the physiological effects of another medical condition.

Unfortunately, most MD treatments are associated with moderate effect sizes (see Box 2). Better understanding of patient-specific causal mechanisms is expected to facilitate the development of biologically informed, patient-specific diagnoses, which in turn should enable psychiatrists to provide treatments that are tailored to a patients' etiological and pathophysiological background.[16,17]

Box 2. Major Depression - key figures

MD is:
- the second most common mental disorder in the world[18,19]
- the single largest contributor to the global burden of disease[12]

Worldwide:
- 300 to 350 million people suffer from MD[20-22]
- 1 in 5 people will experience MD in their lifetime[23-25]
- Women suffer from MD roughly twice as often as men[20,26,27]

MD is highly comorbid with:
- anxiety disorders (50-60%, lifetime)[28]
- substance use disorder (30-40%, lifetime SUD among treatment-seeking MD patients)[29]
- coronary heart disease (80% higher chance)[30]
- overweight and diabetes (40-60% higher chance)[31-33]

Death by suicide:
- affects 800,000 people every year[34,35]
- involves a mood disorder in 43-59% of cases[36,37]

Of all MD patients:
- about 50-70% recover within a year[38]
- about 20% develop a chronic course[39,40]
- about 30-60% seek treatment.[41]

Based on meta-analyses, treatment effect sizes are:
- 0.34-0.40 for PST (Cohen's *d*)[42,43]
- 0.22 for CBT (Hedge's g, Cohen's *d*)[44,45]
- 0.60 for IPT (Hedge's *g*)[46]
- 0.32 for SSRIs/SNRIs (Hedge's *g*)[47]
- 0.42 for TCAs (Cohen's *d*)[48]

CBT, cognitive-behavioral therapy; IPT, interpersonal therapy; MD, Major Depression; PST, problem-solving therapy; SSRI, selective serotonin reuptake inhibitor; SNRI, serotonin-norepinephrine reuptake inhibitor; SUD, Substance abuse disorder; TCA, tricyclic antidepressant

Causes of MD

Throughout history, theories about what exactly ails depressed people have varied widely, ranging from ideas about an excess of black bile to a theories about a conflict between the id and the superego.[1,2] Today, we recognize that MD is an exceedingly complex multifactorial disorder, involving a wide range of interacting risk factors from different levels (see Figure 1).[49–51] According to the diathesis-stress model, MD is not caused by one biological or psychological factor in isolation.[51,52] Rather, in a vulnerable patient, who is predisposed to a negative response to stress, repeated stressors can cause that pre-existing vulnerability (i.e., diathesis) to manifest itself.

This development from pre-existing vulnerability to full-blown MD can be investigated at different levels (see Figure 1). At the phenotypic level, MD includes symptoms like depressive affect, feelings of worthlessness, motor symptoms, and suicidal ideation. However, these symptoms do not arise in all patients. Box 1 shows that there are many different symptom profiles that could fit the diagnostic classification of MD. A number of the additional symptom criteria of MD concern changes (either an increase or a decrease) in appetite, weight, amount of sleep, amount of motor activity, and arousal, which means that some MD patients may have almost opposite symptom profiles. For example, one patient may suffer from weight loss, insomnia, and psychomotor agitation where another patient is plagued by weight gain, hypersomnia, and psychomotor retardation. Most of these symptom profiles are shared by a small percentage of the population only.[54] Other sources of heterogeneity on the clinical level are the severity and course of the disorder. For example, the duration of episodes varies. Based on 15 years of clinical observations, the US National Institute of Mental Health estimates that 67% of patients recover within a year.[55] The recovery is estimated to be 88% after 5 years, and 93% by 10 years.[55] A more recent study in the general population supports these findings, showing that about 50% of new MD patients recover without further episodes, but there is also a sizeable portion (~35%) that experiences recurrent episodes, and about 15% of patients suffer from a chronic course.[56] It should be pointed out that phenotypical heterogeneity by itself is no cause for concern, since this phenomenon also occurs in many somatic disorders.[57,58] However, because the heterogeneity of MD extends to the pathophysiological and etiological levels, identifying the pathophysiological processes leading to this disorder has proven to be more difficult.[49,59,60]

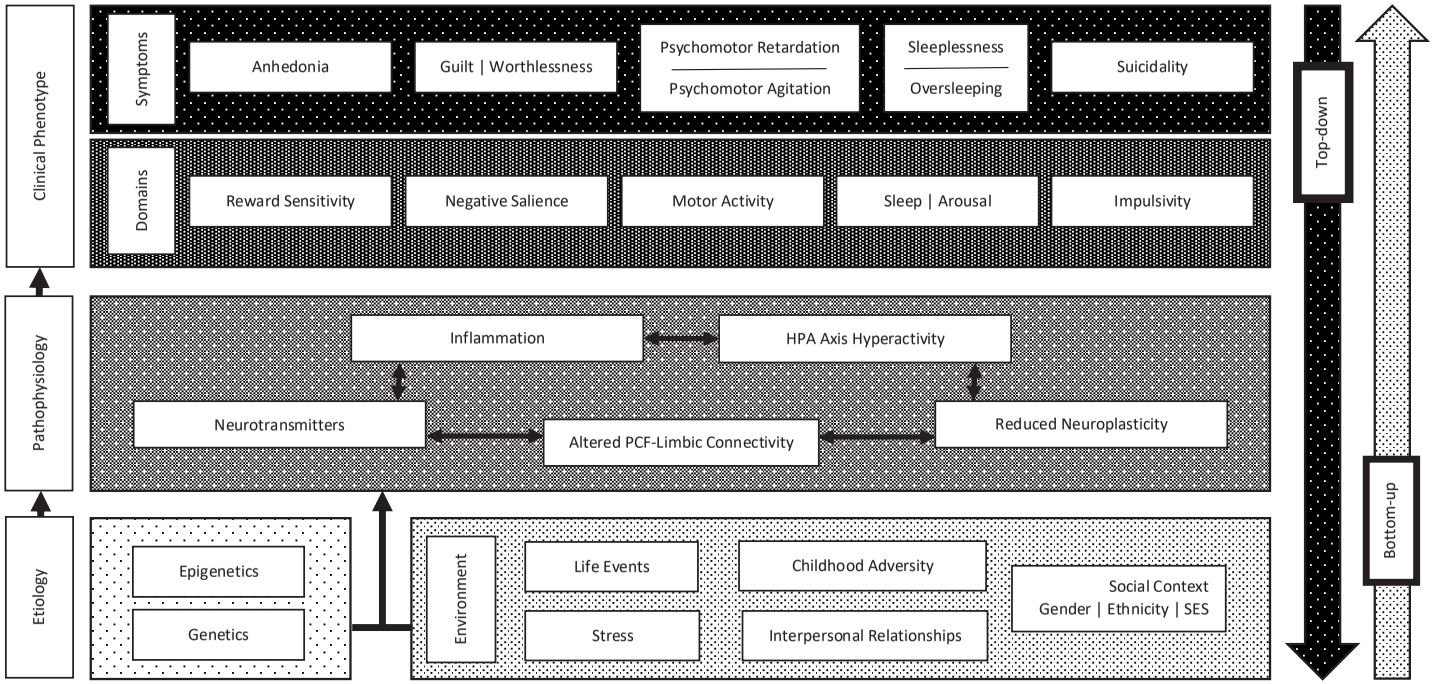


Figure 1. A unified theory of depression

Numerous genetic and environmental etiological factors interact to cause one or more reciprocally interactive pathophysiological mechanisms, which in turn cause symptomatic phenotypic expression, organized here according to domains from the Research Domain Criteria[53]. Adapted from "The neurobiology of depression: An integrated view" by Dean & Keshavan (2017), with permission.

At the pathophysiological level, the putative underlying mechanisms of MD can be organized according to different categories like neurochemistry[61], tissue- and organ-level pathology (e.g., inflammation[62] or an increased stress response[63]), and altered neurocircuitry[64,65].[59] The most commonly prescribed medication (i.e., SSRIs/TCAs) is based on the proposition that MD is a result of diminished activity of serotonin pathways (i.e., the serotonin hypothesis) or both serotonin and catecholamine pathways (i.e., the monoamine hypothesis).[66–68] However, there are some problems with these hypotheses, chief among which is the modest effect size of antidepressant medication[47,48], which suggests that either these drugs do not actually increase serotonin levels, or that many of the people currently taking SSRIs might not have dysfunctional serotonin pathways in the first place.[47,69] The latter seems more likely, because although a transient lowering in brain serotonin activity can be induced in multiple ways, this has failed to induce depression in healthy subjects like it does in people with a history of MD.[68] Furthermore, it has been shown that some MD patients have lower monoamine levels, but there are also patients with similar levels compared to healthy controls.[70–73] Treatments based on other theories suffer from similar issues. For example, the effect size of anti-inflammatory treatments for MD is estimated to be about 0.34-0.55 (Cohen's *d*) on average, but research suggests that this might be the result of averaging over patients with and without immunological dysregulations.[74–79] It is important to note that it is often difficult to tell whether the biological differences between MD patients and controls are really part of the pathophysiology of the disorder. They could also be a result of psychopathology-induced lifestyle changes, or they might be better categorized part of the etiology of the disorder, because they are a result of a genetic predisposition.

At the etiological level, the predisposition to a negative response to stress is thought to be a result of genetic[80,81] and epigenetic factors related to the pathophysiological mechanisms mentioned above[82,83], interacting with environmental factors such as traumatic life experiences[84,85] or socioeconomic status[86]. None of these factors are by themselves sufficient to cause MD, and none of them are absolutely required for the development of the disorder.[87] Indeed, the correlations between single risk factors and the presence of MD tend to be weak, which could mean that the total risk of MD consists of many small effects.[88,89] It could also be that each risk factor is more strongly related to the development of MD for some patients and less so in others, which would result in a low average observed correlation in the complete patient group.[90–92]

Overall, this evidence suggests that there is not one biological disturbance underlying depression in all patients (i.e., impaired serotonin functioning) that underlies MD in all patients, which can be targeted with one type of treatment.[67,93,94]

Investigating the etiology of MD

Since the effectiveness of current therapies relative to placebo is modest other approaches are necessary to address the public health burden of MD.[47,69,95] Preventive interventions for both first onsets and recurrent episodes of MD seem like a promising avenue.[96,97] Identifying key risk factors for MD will help us provide focal points for preventive interventions.[96–98] As described above, potential risk factors for MD range from genetic and environmental variables to different types of biological disturbances.[49]

One method to identify key risk factors for MD is relative importance analysis, which calculates the proportion explained variance of each variable, by comparing the statistical fit of all possible models that include the variable in question to that of the complete collection of possible models.[99,100] This means that putative risk factors should be investigated together in a large general population study. Unfortunately, studies that would enable such analyses are rare, because collecting data on a large group of variables in a sizeable group of participants is expensive and time-consuming. Almost all longitudinal general population studies investigating onset and/or recurrence of MD include either: (a) a sample with a limited age range, (b) only males or females, (c) a limited sample size or (d) a limited number of risk factors.[101–112] Furthermore, the computational power required of this type of analysis is large, and increases exponentially with each additional variable. Therefore, most studies report models including individual risk factors instead, or opt to specify a single multivariable model including all variables that are significant in univariable analyses.[56,113,114]

Other limitations of commonly used models include their inability to investigate patterns more complex than a u-curve. Linear regression models suffice to study general trends, but these models are unable to accommodate different patterns.[115] For example, it is common knowledge that women suffer from depression more often than men, but how exactly MD varies across age and sex has been subject of debate for a long time.[115–117] More insight into these patterns could be used to improve opportunities for public health interventions by identifying sub-populations with higher MD prevalence or incidence because of specific life phases, such as parenthood or menopause.[117–119]

The heterogeneity of Major Depression

Elucidating the etiology of MD is complicated further by the heterogeneity of the patient population. Good classifications group individual in such a way that all the members have roughly the same chance for some relevant characteristic or outcome, i.e., the intra-class homogeneity should be high.[120] When it comes to MD, this means that in the ideal scenario, patients share similar genetic or environmental risk factors, similar pathophysiology, and respond well to similar specific treatments like SSRIs or CBT. Ideally, there is also high inter-class heterogeneity, meaning these characteristics should vary widely between MD and other classes like Generalized Anxiety Disorder (GAD).[120] If that is the case, class membership can be established with high certainty, and optimal treatment for an individual patient can be determined by referring back to the classification.[120] One way to improve the intra-class homogeneity of the MD diagnosis is to look for more homogeneous groups within the population of patients diagnosed with MD (i.e., subtypes). The first clinical subtypes of MD were largely based on clinical consensus, not unlike the MD classification itself.[121] Unfortunately, clinical subtypes of MD have not performed much better with regards to prediction of onset, course, and treatment response than the original MD classification.[122]

Data-driven subtyping of Major Depression

Data-driven approaches address the issue of intra-class heterogeneity by using computational methods to identify patterns in data, which might have been missed by clinical observeration.[123] Although data-driven approaches to psychiatric diagnostics have long been used in psychiatric research, they have recently gained more popularity.[123–132] This is likely a result of the growing realization that better-specified phenotypes are needed, but also due to the increasing availability of suitable datasets and ongoing advances in statistics and machine learning.[123,133,134] However, because of several methodological issues, it is still unclear how much of an improvement can be made with data-driven subtypes.

The influence of methodological variation

Data-driven subtyping is usually performed using some form of unsupervised learning (i.e., finite mixture models and clustering algorithms such as k-means clustering, hierarchical clustering, and community detection[123]). Whereas supervised learning either succeeds at predicting a predefined outcome (e.g., onset of MD, treatment response, chronic course)

or not, unsupervised learning aims to detect previously hidden structures in data, which means there is no straightforward way to judge the quality of unsupervised learning results. Because of this, there is a plethora of different unsupervised learning methods, and the amount of different model specifications available is a lot larger compared to supervised learning.[135] This means that, since the specifics of a chosen analytical method can have a significant influence on research outcomes, variations across studies are a realistic risk when it comes to unsupervised learning algorithms.[136–138] And while a statistical model is not necessarily valid just because it is robust to methodological variation, significant changes to the model as a result of a different set of methodological decisions leads to serious doubts about its validity.[139] Thus, increased insight into the effects of methodological variation on unsupervised clustering results could help us prevent overinterpretation of the results from our data-driven subtyping models. In addition, it could provide leads for data-driven subtypes of MD by identification of patterns that are robust to methodological variation.

Top-down vs. bottom-up

It is also unknown which type of data will deliver the best results when it comes to data-driven subtyping. Even though cluster algorithms do not prioritize any explanatory level over the other *a priori*, research into diagnostic subtypes of depression has thus far predominantly focused on subtyping based on higher-level data such as symptom patterns, comparing lower-level etiological and pathophysiological differences post-hoc (i.e., top-down subtyping, see Figure 1).[120,123,131] However, there is little evidence showing that heterogeneity in etiology and treatment response are best explained by variations at the level of symptoms, as data-driven subtype classifications based on cluster analyses of symptoms have been shown to have limited value when it comes to prediction of course and treatment response.[123,131,140] In fact, there is no obvious reason to assume that similar symptoms will always be caused by similar pathophysiology or similar etiology, as there are plenty of examples in medicine where different pathologies lead to similar symptoms and biomedical tests are required to differentiate between them (i.e., equifinality).[141] For example, a fever can be caused by some kind of viral or bacterial infection, by inflammatory conditions like rheumatoid arthritis, but also by a heat stroke.[142] Therefore, it would be very interesting to perform subtyping based on other sources of heterogeneity, including clinical risk factors, biochemical markers, genetic variations, and brain region activity/connectivity. Indeed, research initiatives such as Research Domain Criteria and large-scale projects, such as the Roadmap for Mental Health Research in Europe have emphasized the need to incorporate multiple levels when investigating psychiatric disorder mechanisms.[143–145]

Instead of using a top-down approach of comparing differences between higher-level symptom-based subtypes on lower explanatory levels such as pathophysiology or etiology, it might be worthwile to apply a bottom-up approach, starting with lower-level data and working our way up from there (see Figure 1). In this way, we might be able to identify groups of people that share a similar etiology and/or similar pathophysiology, which means there are more likely to respond to similar treatments. Symptom profiles might differ between these groups, or they might not – this is of lesser importance than predicting treatment response. Still, recent top-down subtypes might provide an interesting guide mark for bottom-up subtyping based on lower-level data. For example, Latent Class Analysis (LCA) resulted in one moderate and two severe depression subtypes in the Netherlands Study of Depression and Anxiety (NESDA).[146,147] The severe subtypes mainly differed on the probabilities of diurnal variation, early morning awakening, hypersomnia vs. insomnia, and increased vs. decreased appetite and weight. Subsequent studies showed that the subtype with increased weight and appetite had, among other things, higher leptin, insulin, and fatty-acid-binding protein scores, higher metabolic syndrome risk, and a higher probability of carrying a genetic variant of obesity-associated protein (FTO; rs9939609).[147–151] They also had higher inflammation marker levels (e.g., C-reactive protein, interleukin-6, complement C3).[151,152] If subtypes based on biomarkers such as the metabolic and inflammation-related markers mentioned above include similar patients as these top-down subtypes, this might imply that the differences in symptoms do, in this case, indeed reflect different pathologies.

This thesis

In summary, the MD classification captures a group of patients with a high burden of disease, but MD patients are a heterogeneous group in many ways, which constitutes a major challenge for research into the underlying etiological and pathophysiological processes, as well as the development of more effective, tailored. Based on this heterogeneity, it seems that theories stating that there is one biological disturbance underlying depression in all patients (i.e., impaired serotonin functioning) that underlies MD in all patients are unlikely to be valid. In fact, previous studies have identified many potential risk factors, but in order to figure out which of these are the most important for predicting MD onset and recurrence, they need to be investigated together in a large general population study, and understanding the relationship between specific risk factors such as age and sex requires more sophisticated non-linear models. Since it is possible that different risk

factors are more strongly related to the development of MD in some patients and less so in others, looking for subtypes of MD is another promising avenue for improving prevention and treatment efforts. However, bottom-up subtyping based on etiological or pathophysiological data is as of yet largely unexplored. The aims of this thesis are to gain more insight into the etiology of MD by (1) using rich datasets and novel methodology to take a more detailed look at MD risk factors and (2) to investigate if and how well bottom-up subtyping approaches might enable the discovery of more homogeneous subtypes of MD.

The first part of this thesis (Chapters 2-3) describes studies that use sophisticated statistical models in combination with a large and rich dataset to refine our understanding of the etiology of MD. In Chapter 2, relative importance analysis was used to investigate which risk factors are most important, using rich set of risk factors for the incidence and chronicity of MD in a large population study. In order to take a more detailed look at the relationship between sex, age, and internalizing psychopathology, the next study applied advanced non-linear modelling to a large population sample, investigating the prevalence of MD and other internalizing disorders as well as mean scores for internalizing symptoms and traits over the lifetime (Chapter 3).

The second part of this thesis (Chapter 4-6) focuses on methodological and empirical questions about bottom-up MD subtyping. Chapter 4 aimed to gain insight into existing knowledge about the role of biological factors in MD heterogeneity by means of a systematic review of current evidence available for data-driven biological subtypes of MD from studies that identified (1) data-driven subtypes of MD based on biological variables, or (2) data-driven subtypes based on clinical features such as symptom patterns and validated these with biological variables post-hoc. In order to investigate whether it was possible to successfully apply clustering techniques commonly used in studies based on clinical data to a set of biochemical biomarkers, Chapter 5 attempted to identify biochemical subtypes of MD using Latent Class Analysis. The final Chapter describes the use of Specification-Curve Analysis to gain more insight into the influence of methodological variation on biomarker-based cluster-analysis results (Chapter 6).

References

1. Radden, J. *The Nature of Melancholy: From Aristotle to Kristeva*. *The Nature of Melancholy: From Aristotle to Kristeva* (Oxford University Press, 2002).
2. Jackson, S. W. *Melancholia and Depression: From Hippocratic Times to Modern Times.* (Yale University Press, 1986).
3. Maj, M. When does depression become a mental disorder? *Br. J. Psychiatry* **199**, 85–86 (2011).
4. Lewinsohn, P. M., Solomon, A., Seeley, J. R. & Zeiss, A. Clinical implications of 'Subthreshold' depressive symptoms. *J. Abnorm. Psychol.* **109**, 345–351 (2000).
5. Kendell, R. & Jablensky, A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am. J. Psychiatry* **160**, 4–12 (2003).
6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). (2013)
7. Papakostas, G. I. *et al.* Quality of life assessments in major depressive disorder: A review of the literature. *Gen. Hosp. Psychiatry* **26**, 13–17 (2004).
8. Berlim, M. T. & Fleck, M. P. A. Quality of Life and Major Depression. in *Quality of Life Impairment in Schizophrenia, Mood and Anxiety Disorders: New Perspectives on Research and Treatment* 241–252 (Springer, 2007).
9. Ormel, J. *et al.* Disability and treatment of specific mental and physical disorders across the world. *Br. J. Psychiatry* **192**, 368–375 (2008).
10. Andrews, G. & Titov, N. Depression is very disabling. *Lancet* **370**, 808–809 (2007).
11. Moussavi, S. *et al.* Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851–858 (2007).
12. Friedrich, M. J. Depression Is the Leading Cause of Disability Around the World. *JAMA* **317**, 1517 (2017).
13. Isometsä, E. Suicidal behaviour in mood disorders-Who, When, and Why? *Can. J. Psychiatry* **59**, 120–130 (2014).
14. Turecki, G. *et al.* Suicide and suicide risk. *Nat. Rev. Dis. Prim.* **5**, 74 (2019).
15. NHG-werkgroep Depressie. *NHG-Richtlijnen, M44, Versie 3.1*. (2019).
16. Simon, G. E. & Perlis, R. H. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry* **167**, 1445–1455 (2010).
17. Kendler, K. S. The structure of psychiatric science. *Am. J. Psychiatry* **171**, 931–938 (2014).
18. Angst, J. *et al.* The epidemiology of common mental disorders from age 20 to 50: Results from the prospective Zurich cohort Study. *Epidemiol. Psychiatr. Sci.* **25**, 33–34 (2016).
19. Ferrari, A. J. *et al.* Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychol. Med.* **43**, 471–481 (2013).
20. Wittchen, H. U. *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).
21. Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T. & Kessler, R. C. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J. Clin. Psychiatry* **76**, 155–62 (2015).

22. Patel, V. *et al.* Addressing the burden of mental, neurological, and substance use disorders: Key messages from Disease Control Priorities, 3rd edition. *Lancet* **387**, 1672–1685 (2016).

23. Bijl, R. V, Ravelli, A. & Van Zessen, G. Prevalence of psychiatric disorder in the general population: results of The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc. Psychiatry Psychiatr. Epidemiol.* **33**, 587–595 (1998).

24. Kessler, R. C. *et al.* Results from the National Comorbidity Survey Replication (NCS-R). *J. Am. Med. Assoc.* **289**, 3095–3105 (2003).

25. Hasin, D. S. *et al.* Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry* **75**, 336–346 (2018).

26. Baxter, A. J. *et al.* Challenging the myth of an 'epidemic' of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depress. Anxiety* **31**, 506–516 (2014).

27. Kuehner, C. Why is depression more common among women than among men? *The Lancet Psychiatry* **4**, 146–158 (2017).

28. Kaufman, J. & Charney, D. Comorbidity of mood and anxiety disorders. *Depress. Anxiety* **12**, 69–76 (2000).

29. Davis, L., Uezato, A., Newell, J. M. & Frazier, E. Major depression and comorbid substance use disorders. *Curr. Opin. Psychiatry* **21**, 14–18 (2008).

30. Nicholson, A., Kuper, H. & Hemingway, H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *Eur. Heart J.* **27**, 2763–2774 (2006).

31. Knol, M. J. *et al.* Depression as a risk factor for the onset of type 2 diabetes mellitus. A meta-analysis. *Diabetologia* **49**, 837–845 (2006).

32. Mezuk, B., Eaton, W. W., Albrecht, S. & Golden, S. H. Depression and type 2 diabetes over the lifespan: A meta-analysis. *Diabetes Care* **31**, 2383–2390 (2008).

33. Luppino, F. S. *et al.* Overweight, Obesity, and Depression: A Systematic Review and Meta-analysis of Longitudinal Studies. *Arch. Gen. Psychiatry* **67**, 220–229 (2010).

34. World Health Organization. *Preventing Preventing suicide suicide - A global imperative*. (2014).

35. Roth, G. A. *et al.* Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1736–1788 (2018).

36. Arsenault-Lapierre, G., Kim, C. & Turecki, G. Psychiatric diagnoses in 3275 suicides: A meta-analysis. *BMC Psychiatry* vol. 4 37 (2004).

37. Cavanagh, J. T. O., Carson, A. J., Sharpe, M. & Lawrie, S. M. Psychological autopsy studies of suicide: A systematic review. *Psychol. Med.* **33**, 395–405 (2003).

38. Richards, D. Prevalence and clinical course of depression: A review. *Clin. Psychol. Rev.* **31**, 1117–1125 (2011).

39. Klein, D. N. & Santiago, N. J. Dysthymia and chronic depression: Introduction, classification, risk factors, and course. *J. Clin. Psychol.* **59**, 807–816 (2003).

40. Hölzel, L., Härter, M., Reese, C. & Kriston, L. Risk factors for chronic depression - A systematic review. *J. Affect. Disord.* **129**, 1–13 (2011).

41. Bristow, K. & Patten, S. Treatment-seeking rates and associated mediating factors among individuals with depression. *Can. J. Psychiatry* **47**, 660–665 (2002).

42. Cuijpers, P., van Straten, A. & Warmerdam, L. Problem solving therapies for depression: A meta-analysis. *Eur. Psychiatry* **22**, 9–15 (2007).

43. Bell, A. C. & D'Zurilla, T. J. Problem-solving therapy for depression: A meta-analysis. *Clin. Psychol. Rev.* **29**, 348–353 (2009).

44. Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D. & Andersson, G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol. Med.* **40**, 211–23 (2010).

45. Santoft, F. *et al.* Psychological Medicine Cognitive behaviour therapy for depression in primary care: systematic review and meta-analysis. **49**, 1266–1274 (2019).

46. Cuijpers, P., Donker, T., Weissman, M. M., Ravitz, P. & Cristea, I. A. Interpersonal psychotherapy for mental health problems: A comprehensive meta-analysis. *Am. J. Psychiatry* **173**, 680–687 (2016).

47. Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A. & Rosenthal, R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* **358**, 252–260 (2008).

48. Arroll, B. *et al.* Efficacy and tolerability of tricyclic antidepressants and SSRIs compared with placebo for treatment of depression in primary care: A meta-analysis. *Ann. Fam. Med.* **3**, 449–456 (2005).

49. Kendler, K. S. The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Mol. Psychiatry* **17**, 377–388 (2012).

50. Zachar, P. & Kendler, K. S. General Introduction. in *Levels of Analysis in Psychopathology* 1–16 (Cambridge University Press, 2020).

51. Dean, J. & Keshavan, M. The neurobiology of depression: An integrated view. *Asian J. Psychiatr.* **27**, 101–111 (2017).

52. Patten, S. B. Major depression epidemiology from a diathesis-stress conceptualization. *BMC Psychiatry* **13**, 19 (2013).

53. Morris, S. E. & Cuthbert, B. N. Research domain criteria: Cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin. Neurosci.* **14**, 29–37 (2012).

54. Fried, E. I. & Nesse, R. M. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR???D study. *J. Affect. Disord.* **172**, 96–102 (2015).

55. Mueller, T. I. *et al.* Recovery after 5 years of unremitting major depressive disorder. *Arch. Gen. Psychiatry* **53**, 794–799 (1996).

56. Eaton, W. W. *et al.* Population-based study of first onset and chronicity in major depressive disorder. *Arch. Gen. Psychiatry* **122**, 184–191 (2008).

57. Capobianco, E. Systems and precision medicine approaches to diabetes heterogeneity: a Big Data perspective. *Clin. Transl. Med.* **6**, 1–10 (2017).

58. Kravitz, R. L., Duan, N. & Braslow, J. Evidence-based medicine, heterogeneity of treatment effects, and the trouble with averages. *Milbank Q.* **82**, 661–687 (2004).

59. Jentsch, M. C. *et al.* Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses. *Biomark. Med.* **9**, 277–297 (2015).

60. Monroe, S. M. & Anderson, S. F. Depression: The Shroud of Heterogeneity. *Curr. Dir. Psychol. Sci.* **24**, 227–231 (2015).

61. Albert, P. R., Benkelfat, C. & Descarries, L. The neurobiology of depression-revisiting the serotonin hypothesis. I. cellular and molecular mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2378–2381 (2012).

62. Valkanova, V., Ebmeier, K. P. & Allan, C. L. CRP, IL-6 and depression: A systematic review and meta-analysis of longitudinal studies. *J. Affect. Disord.* **150**, 736–744 (2013).

63. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).

64. Ferrari, F. & Villa, R. F. The Neurobiology of Depression: an Integrated Overview from Biological Theories to Clinical Evidence. *Mol. Neurobiol.* **54**, 4847–4865 (2017).

65. Nestler, E. J. *et al.* Neurobiology of depression. *Neuron* **34**, 13–25 (2002).

66. Hirschfeld, R. M. A. History and evolution of the monoamine hypothesis of depression. *J. Clin. Psychiatry* **61**, 4–6 (2000).

67. Cowen, P. J. & Browning, M. What has serotonin to do with depression? *World Psychiatry* **14**, 158–160 (2015).

68. Liu, B., Liu, J., Wang, M., Zhang, Y. & Li, L. From Serotonin to Neuroplasticity: Evolvement of Theories for Major Depressive Disorder. *Front. Cell. Neurosci.* **11**, 305 (2017).

69. Coe, R. It's the effect size, stupid. What effect size is and why it is important. in *British Educational Research Association annual conference* 1–18 (2002).

70. Asberg, M., Bertilsson, L., Tuck, D., Cronholm, B. & Sjöqvist, F. Indoleamine metabolites in the cerebrospinal fluid of depressed patients before and during treatment with nortriptyline. *Clin. Pharmacol. Ther.* **14**, 277–86 (1973).

71. Asberg, M., Thoren, P., Traskman, L., Bertilsson, L. & Ringberger, V. 'Serotonin depression'--a biochemical subgroup within the affective disorders? *Science (80-.).* **191**, 478–480 (1976).

72. Gibbons, R. D. & Davis, J. M. Consistent evidence for a biological subtype of depression characterized by low CSF monoamine levels. *Acta Psychiatr. Scand.* **74**, 8–12 (1986).

73. Westenberg, H. G. M. & Verhoeven, W. M. A. CSF monoamine metabolites in patients and controls: support for a bimodal distribution in major affective disorders. *Acta Psychiatr. Scand.* **78**, 541–549 (1988).

74. Köhler, O. *et al.* Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects a systematic review and meta-analysis of randomized clinical trials. *JAMA Psychiatry* **71**, 1381–1391 (2014).

75. Kohler, O., Krogh, J., Mors, O. & Eriksen Benros, M. Inflammation in Depression and the Potential for Anti-Inflammatory Treatment. *Curr. Neuropharmacol.* **14**, 732–742 (2016).

76. Bai, S. *et al.* Efficacy and safety of anti-inflammatory agents for the treatment of major depressive disorder: A systematic review and meta-analysis of randomised controlled trials. *J. Neurol. Neurosurg. Psychiatry* **91**, 21–32 (2020).

77. Ioannou, M. *et al.* Towards precision medicine: What are the stratification hypotheses to identify homogeneous inflammatory subgroups. *Eur. Neuropsychopharmacol.* **45**, 108–121 (2021).

78. Osimo, E. F. *et al.* Inflammatory markers in depression: A meta-analysis of mean differences and variability in 5,166 patients and 5,083 controls. *Brain, Behavior, and Immunity* vol. 87 901–909 (2020).

79. Chamberlain, S. R. *et al.* Treatment-resistant depression and peripheral C-reactive protein. *Br. J. Psychiatry* **214**, 11–19 (2019).

80. Jansen, R. *et al.* Gene expression in major depressive disorder. *Mol. Psychiatry* **21**, 1–9 (2015).

81. Levinson, D. F. The Genetics of Depression: A Review. *Biol. Psychiatry* **60**, 84–92 (2006).

82. Archer, T., Oscar-Berman, M., Blum, K. & Gold, M. Epigenetic Modulation of Mood Disorders. *J. Genet. Syndr. Gene Ther.* **4**, 1000120 (2013).

83. Dalton, V. S., Kolshus, E. & McLoughlin, D. M. Epigenetics and depression: Return of the repressed. *J. Affect. Disord.* **155**, 1–12 (2014).

84. Jaworska-Andryszewska, P. & Rybakowski, J. K. Childhood trauma in mood disorders: Neurobiological mechanisms and implications for treatment. *Pharmacol. Reports* **71**, 112–120 (2019).

85. Wiersma, J. E. *et al.* The importance of childhood trauma and childhood life events for chronicity of depression in adults. *J. Clin. Psychiatry* **70**, 983 (2009).

86. Reiss, F. Socioeconomic inequalities and mental health problems in children and adolescents: A systematic review. *Soc. Sci. Med.* **90**, 24–31 (2013).

87. Rothman, K. J. & Greenland, S. Causation and causal inference in epidemiology. *Am. J. Public Health* **95**, S144–S150 (2005).

88. Kendler, K. S., Gardner, C. O. & Prescott, C. A. Toward a comprehensive developmental model for major depression in women. *Am. J. Psychiatry* **159**, 1133–1145 (2002).

89. Kendler, K. S., Gardner, C. O. & Prescott, C. A. Toward a comprehensive developmental model for major depression in men. *Am. J. Psychiatry* **163**, 115–124 (2006).

90. Moffitt, T. E. & Caspi, A. Gene–environment interactions in psychiatry: joining forces with neuroscience. *Nat. Rev. Neurosci.* **7**, 583–590 (2006).

91. Jaffee, S. R. & Price, T. S. Gene-environment correlations: A review of the evidence and implications for prevention of mental illness. *Mol. Psychiatry* **12**, 432–442 (2007).

92. Molenaar, P. C. M. On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Dev. Psychobiol.* **50**, 60–69 (2008).

93. Prins, J., Olivier, B. & Korte, S. M. Triple reuptake inhibitors for treating subtypes of major depressive disorder: the monoamine hypothesis revisited. *Expert Opin. Investig. Drugs* **20**, 1107–1130 (2011).

94. Baumeister, A. A., Hawkins, M. F. & Uzelac, S. M. The myth of reserpine-induced depression: Role in the historical development of the monoamine hypothesis. *J. Hist. Neurosci.* **12**, 207–220 (2003).

95. Cuijpers, P., Andersson, G., Donker, T. & Van Straten, A. Psychological treatment of depression: Results of a series of meta-analyses. *Nord. J. Psychiatry* **65**, 354–364 (2011).

96. Beshai, S., Dobson, K. S., Bockting, C. L. H. & Quigley, L. Relapse and recurrence prevention in depression: Current research and future prospects. *Clin. Psychol. Rev.* **31**, 1349–1360 (2011).

97. Ormel, J., Cuijpers, P., Jorm, A. F. & Schoevers, R. Prevention of depression will only succeed when it is structurally embedded and targets big determinants. *World Psychiatry* **18**, 111 (2019).

98. van Zoonen, K. *et al.* Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *Int. J. Epidemiol.* **43**, 318–329 (2014).

99. Burnham, K. P. & Anderson, D. R. Multimodel Inference: understanding AIC and BIC in Model Selection, Amsterdam Workshop on Model Selection. *Sociol. Methods Res.* **33**, 261–304 (2004).

100. Tonidandel, S. & LeBreton, J. M. Relative Importance Analysis: A Useful Supplement to Regression Analysis. *J. Bus. Psychol.* **26**, 1–9 (2011).

101. Murphy, J. M., Laird, N. M., Monson, R. R., Sobol, A. M. & Leighton, A. H. Incidence of depression in the Stirling County Study: Historical and comparative perspectives. *Psychol. Med.* **30**, 505–514 (2000).

102. Bromberger, J. T. *et al.* Predictors of first lifetime episodes of major depression in midlife women. *Psychol. Med.* **39**, 55–64 (2009).

103. Schoevers, R. Prevention of Late-Life Depression in Primary Care: Do We Know Where to Begin? *Am. J. Psychiatry* **163**, 1611–1621 (2006).

104. Schoevers, R. A. *et al.* Risk factors for depression in later life; results of a prospective community based study (AMSTEL). *J. Affect. Disord.* **59**, 127–137 (2000).

105. Grant, B. F. *et al.* Sociodemographic and psychopathologic predictors of first incidence of DSM-IV substance use, mood and anxiety disorders: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Mol. Psychiatry* **14**, 1051–66 (2009).

106. Mattisson, C., Bogren, M., Brådvik, L. & Horstmann, V. Mortality of subjects with mood disorders in the Lundby community cohort: A follow-up over 50 years. *J. Affect. Disord.* **178**, 98–106 (2015).

107. Mattisson, C. *et al.* Risk factors for depressive disorders in the Lundby cohort - A 50 year prospective clinical follow-up. *J. Affect. Disord.* **113**, 203–215 (2009).

108. Bromberger, J. T., Schott, L., Kravitz, H. M. & Joffe, H. Risk factors for major depression during midlife among a community sample of women with and without prior major depression: Are they the same or different? *Psychol. Med.* **45**, 1653–1664 (2015).

109. Lukat, J., Becker, E. S., Lavallee, K. L., van der Veld, W. M. & Margraf, J. Predictors of Incidence, Remission and Relapse of Axis I Mental Disorders in Young Women: A Transdiagnostic Approach. *Clin. Psychol. Psychother.* **24**, 322–331 (2017).

110. Benjet, C. *et al.* Incidence and recurrence of depression from adolescence to early adulthood: A longitudinal follow-up of the Mexican Adolescent Mental Health Survey. *J. Affect. Disord.* **263**, 540–546 (2020).

111. Kendler, K. S., Gatz, M., Gardner, C. O. & Pedersen, N. L. Personality and major depression: A Swedish longitudinal, population-based twin study. *Arch. Gen. Psychiatry* **63**, 1113–1120 (2006).

112. Lorant, V. *et al.* Depression and socio-economic risk factors: 7-Year longitudinal population study. *Br. J. Psychiatry* **190**, 293–398 (2007).

113. De Graaf, R., Bijl, R. V., Ravelli, A., Smit, F. & Vollebergh, W. A. M. Predictors of first incidence of DSM-III-R psychiatric disorders in the general population: Findings from the Netherlands Mental Health Survey and Incidence Study. *Acta Psychiatr. Scand.* **106**, 303–313 (2002).

114. ten Have, M. *et al.* Duration of major and minor depressive episodes and associated risk indicators in a psychiatric epidemiological cohort study of the general population. *Acta Psychiatr. Scand.* **136**, 300–312 (2017).

115. Jorm, A. F. Does old age reduce the risk of anxiety and depression ? A review of epidemiological studies across the adult life span. *Psychol. Med.* **30**, 11–22 (2000).

116. Piccinelli, M. & Wilkinson, G. Gender differences in depression. Critical review. *Br. J. Psychiatry* **177**, 486–492 (2000).

117. Kuehner, C. Why is depression more commen among women than among men? *The Lancet Psychiatry* **4**, 146–58 (2017).

118. Nomaguchi, K. & Milkie, M. A. Parenthood and Well-Being: A Decade in Review. *J. Marriage Fam.* **82**, 198–223 (2020).

119. Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E. & Binau, S. G. Age, Period, and Cohort Trends in Mood Disorder Indicators and Suicide-Related Outcomes in a Nationally Representative Dataset, 2005-2017. **128**, 185–199 (2019).

120. Romeijn, J.-W. & van Loo, H. M. Psychiatric Classification: An A-reductionist Perspective. in *Levels of Analysis in Psychopathology* 349–370 (Cambridge University Press, 2020).

121. Pincus, H. A. DSM-IV: context, concepts and controversies. in *Philosophical Issues in psychiatry II: Nosology* 145–160 (2012).

122. Baumeister, H. & Parker, G. Meta-review of depressive subtyping models. *J. Affect. Disord.* **139**, 126–140 (2012).

123. Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J. & Beckmann, C. F. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 433–447 (2016).

124. Kay, D. W., Garside, R. F., Beamish, P. & Roy, J. R. Endogenous and neurotic syndromes of depression: a factor analytic study of 104 cases. Clinical features. *Br. J. Psychiatry* **115**, 377–388 (1969).

125. Pilowsky, I., Levine, S. & Boulton, D. M. The classification of depression by numerical taxonomy. *Br. J. Psychiatry* **115**, 937–945 (1969).

126. Paykel, E. S. Classification of depressed patients: a cluster analysis derived grouping. *Br. J. Psychiatry* **118**, 275–288 (1971).

127. Andreasen, N. C. & Grove, W. M. The classification of depression: Traditional versus mathematical approaches. *Am. J. Psychiatry* **139**, 45–52 (1982).

128. Merikangas, K. R., Wicki, W. & Angst, J. Heterogeneity of depression. Classification of depressive subtypes by longitudinal course. *Br. J. Psychiatry* **164**, 342–348 (1994).

129. Sullivan, P. F., Kessler, R. C. & Kendler, K. S. Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. *Am. J. Psychiatry* **155**, 1398–1406 (1998).

130. Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G. & Mechelli, A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* **36**, 1140–1152 (2012).

131. Van Loo, H. M., De Jonge, P., Romeijn, J.-W., Kessler, R. C. & Schoevers, R. A. Data-driven subtypes of major depressive disorder: a systematic review. *BMC Med.* **10**, 156 (2012).

132. Wood MacFadyen, H. The classification of depressive disorders. *J. Clin. Psychol.* **31**, 380–394 (1975).

133. Ahmad, A. & Fröhlich, H. Integrating Heterogeneous omics Data via Statistical Inference and Learning Techniques. *Genomics Comput. Biol.* **2**, e32 (2016).

134. Lin, E. & Hsien-Yuan, L. Machine learning and systems genomicsapproaches for multi-omics data. *Biomark. Res.* **5**, (2017).

135. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. *Springer Series in Statistics* (Springer, 2011).

136. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).

137. Silberzahn, R. *et al.* Many analysts, one dataset: Making transparent how variations in analytical choices affect results Authors. *Adv. Methods Pract. Psychol. Sci.* **1**, 337–356 (2018).

138. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).

139. Sampson, R. J., Laub, J. H. & Eggleston, E. P. On the robustness and validity of groups. *J. Quant. Criminol.* **20**, 37–42 (2004).

140. Ulbricht, C. M., Chrysanthopoulou, S. A., Levin, L. & Lapane, K. L. The use of latent class analysis for identifying subtypes of depression: A systematic review. *Psychiatry Res.* **266**, 228–246 (2018).

141. Cicchetti, D. & Rogosch, F. A. Equifinality and multifinality in developmental psychopathology. *Dev. Psychopathol.* **8**, 597–600 (1996).

142. Walter, E. J., Hanna-Jumma, S., Carraretto, M. & Forni, L. The pathophysiological basis and consequences of fever. *Crit. Care* **20**, 200 (2016).

143. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).

144. Haro, J. M. *et al.* ROAMER: roadmap for mental health research in Europe. *Int. J. Methods Psychiatr. Res.* **23**, 1–14 (2014).

145. Insel, T. R. & Cuthbert, B. N. Brain disorders? Precisely. *Science (80-.).* **348**, 499–500 (2015).

146. Lamers, F. *et al.* Identifying depressive subtypes in a large cohort study: Results from the Netherlands Study of Depression and Anxiety (NESDA). *J. Clin. Psychiatry* **71**, 1582–1589 (2010).

147. Milaneschi, Y. *et al.* The effect of FTO rs9939609 on major depression differs across MDD subtypes. *Mol. Psychiatry* **19**, 960–962 (2014).

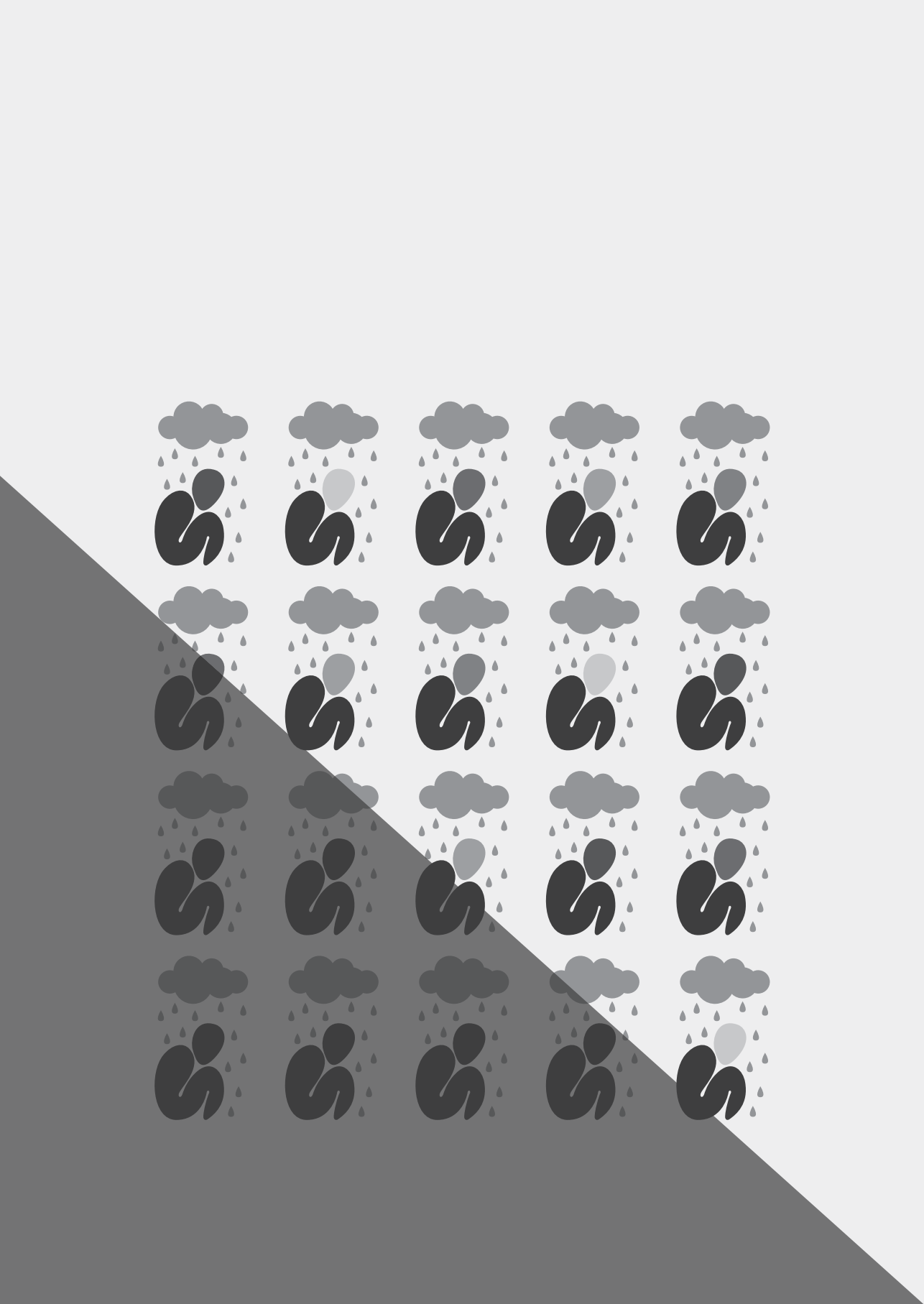
148. Lamers, F. *et al.* Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. *Mol. Psychiatry* **18**, 692–699 (2013).

149. Milaneschi, Y., Lamers, F., Bot, M., Drent, M. L. & Penninx, B. W. J. H. Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol. Psychiatry* **81**, 807–814 (2017).

150. Frayling, T. M. *et al.* A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science* **316**, 889–894 (2007).

151. Lamers, F. *et al.* Serum proteomic profiles of depressive subtypes. *Transl. Psychiatry* **6**, e851 (2016).

152. Orsel, S., Karadag, H., Turkcapar, H. & Karaoglan Kahilogullari, A. Diagnosis and classification subtyping of depressive disorders: Comparison of three methods. *Klin. Psikofarmakol. Bul.* **20**, 57–65 (2010).



Chapter 2

Key risk factors for onset and recurrence of Major Depression: results from Lifelines, a large representative population cohort

Lian Beijers, Klaas J. Wardenaar, Trynke R. de Jong, Robert A. Schoevers, Hanna M. van Loo

Submitted.

Abstract

Background

Major Depression (MD) is a multifactorial disorder with a substantial disease burden, which often has recurrent episodes. Prevention requires intricate knowledge about the key risk factors of MD. This study aims to determine which of the many previously identified risk factors are most important for predicting first onset and recurrence of MD by investigating multivariable models in a longitudinal population study with sufficient sample size.

Methods

The Dutch Lifelines study is a large longitudinal representative population cohort. We selected 21 risk factors for MD, such as socio-demographic variables, neuroticism, family history, stressful life events, childhood trauma, health behaviors, general health status, and metabolic and inflammatory markers. MD onset and recurrence were measured in two follow-up waves (n = 42,724). Relative importance analysis was used to identify key risk factors for MD onset and recurrence.

Results

A family history of anxiety and depression, childhood trauma, higher neuroticism, female sex, younger age, chronic stress, lower physical quality of life and current anxiety disorders were all key risk factors for MD onset. Most key risk factors for MD onset also predicted MD recurrence. Comorbid anxiety and female sex predicted first onset only, whereas lower education levels specifically predicted recurrence.

Conclusion

We identified several key risk factors relevant for onset and recurrence of MD, which could guide primary as well as secondary prevention programs. Our findings suggest that educational inequality plays a role in the course of the disorder, and emphasizes the importance of screening for MD among family members of depressed individuals.

Introduction

Major depression (MD) is among the most prevalent mental disorders worldwide, and is associated with a substantial burden.[1–3] This burden is highest in patients who have a course with chronic or recurrent episodes.[4–9] Since the 1970s, increasing numbers of people in Western countries are receiving psychotherapy or pharmacotherapy for the disorder, yet epidemiological data do not indicate a drop in MD prevalence.[10] The effectiveness of current therapies relative to placebo is modest, and other approaches are necessary to address the public health burden of MD.[11–13] Preventive interventions for both first onsets and recurrent episodes of MD seem like a promising avenue.[14,15] Selective prevention, which targets individuals or subgroups that are at high risk of MD, is thought to be more effective compared to universal interventions, which target the whole population, regardless of risk status.[16,17] Risk factors that cannot be changed with interventions (e.g., gender and age) can still be used to determine which people are at highest risk - interventions to increase resilience in these people specifically may potentially reduce the prevalence of MD.[15] Identifying key risk factors for MD will therefore help us determine which interventions are most likely to succeed in preventing MD episodes.[14,15,17]

A plethora of risk factors for MD have been identified. For example, although the exact mechanisms are unclear, it is well known that demographic factors such as younger age and female sex are risk factors for MD.[18–20] We also know that depression runs in families.[20–22] Based on twin studies, the heritability of MD is estimated to be about 37%[23–26], but rearing experiences are estimated to contribute just as much to trans-generational transmission of MD risk as genetic risk.[27–30] A number of physiological problems such as dysregulations of neuroendocrine[31,32], metabolic[33,34], and inflammatory[34,35] systems and the presence of somatic disorders[36–39] have also been related to higher MD risk, as have several aspects of personality, especially neuroticism.[20,21] Finally, environmental risk factors such as traumatic life experiences[40,41], socioeconomic status[20,42,43] and lifestyle factors like the consumption of alcohol[44,45] and tobacco[46] or the amount of physical movement[47,48] also contribute to MD risk.

Most of this knowledge comes from studies investigating single risk factors, or risk factor domains, and the differences in sample characteristics (e.g., different distributions of sex, age, ethnicity, or socio-economic status) and methodology (e.g., self-report vs. clinician-rated, different time intervals) make it difficult to compare the effects of different risk factors.[49,50] Furthermore, measuring effects of individual risk factors in different samples increases the risks associated with unidentified confounding or mediation, because many of the aforementioned risk factors interact.[51] For example, lower education

levels might predict the onset of MD directly, but this effect could be explained by lower income. Measuring the risks associated with both variables in independent samples might lead to the conclusion that there is a similar effect when in fact it is the same variance that is being explained by both risk factors. Furthermore, to investigate the directionality of the relationship between MD and these factors, longitudinal data are needed from a large number of participants. In summary, in order to identify key risk factors for MD, it is crucial that multiple risk factors are investigated in concert, using a multivariable model in a longitudinal population study with sufficient sample size.

One method to identify key risk factors for MD is relative importance analysis, which calculates the proportion of explained variance of each risk factor, by comparing the statistical fit of possible models including that risk factor to that of all possible models.[52,53] However, this type of analysis has never been applied to risk factors of MD before, since population studies that include sufficient numbers of risk factors are rare, and the computational power required for this type of analysis increases exponentially with each additional risk factor. Most previous studies reported a number of models investigating individual risk factors instead, or opted to specify a single multivariable model including all risk factors that were significant in univariable analyses.[54–56]

In order to investigate the key risk factors of MD onset and recurrence, we performed Relative Importance Analysis using the Lifelines cohort, a large longitudinal population study.[57] From this rich dataset, we selected known risk factors for MD, such as socio-demographic variables, neuroticism, family history, stressful life events, childhood trauma, health behaviors, general health status, and metabolic and inflammatory markers.[21,49,58] Our first aim was to investigate which of these risk factors were most important for predicting onset of the first MD episode. Our second aim was to investigate whether similar or different key risk factors predict recurrence of MD.

Methods

This study was preregistered on the Open Science Framework in February 2020 (https://osf.io/7bptq/).

The Lifelines Cohort Study

The Lifelines Cohort Study is a large population-based cohort study and biobank that is used for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing.[59] Between 2006 and

2013, inhabitants of the northern part of the Netherlands were invited to participate through their general practitioners, in a three-generation design. At wave 1, data were collected for 167,729 participants, aged 6 months to 93 years. Participants visited one of the Lifelines research sites for a physical examination, including lung function, ECG and cognition tests, and completed extensive questionnaires. Fasting blood and 24-h urine samples were processed on the day of collection and stored at -80 ˚C in a fully automated storage facility. The baseline questionnaire consisted of two parts containing questions on, among other topics, demographics, health status, lifestyle, and psychosocial aspects. We made use of the baseline measurement (wave 1, 2007-2013) the first follow-up wave (wave 2, 2014-2017), and the Lifetime Depression Assessment Self-report[60], an add-on online questionnaire administered in 2018 for the Biobanks Netherlands Internet Collaboration project. We included all subjects who participated at wave 1 and the 2018 add-on survey (n = 42,724). The mean intervals between waves 1-2 and waves 2 and the 2018 add-on survey were 3.88 (SD = 1.19) and 2.83 (SD = 1.06) years, respectively.

Baseline predictors

A total of 21 putative risk factors were included in the analyses (see Online Supplement). These risk factors were classified into nine major risk domains: (1) Sex and age, (2) Current social and economic environment (education and income level, unfavourable work status, number of social contacts), (3) Health behaviors (physical movement, current smoking status, drinking alcohol), (4) Somatic health (physical quality of life (QoL), cardiovascular problems, cancer, inflammatory disorders, low-grade inflammation, metabolic syndrome) (5) Anxiety disorders (number of current diagnoses), (6) Family history of anxiety and depression, (7) Personality (neuroticism), (8) Early adverse life events (childhood trauma), (9) Acute and chronic stress. Most of these predictors were assessed at baseline using self-report instruments. The number of anxiety disorders was determined by trained research assistants using the Mini-International Neuropsychiatric Interview (MINI). Low-grade inflammation was measured through serum levels of high-sensitivity C-reactive protein (CRP). The metabolic syndrome diagnosis included measurements of waist circumference and blood pressure, as well as serum levels of glucose, high-density lipoprotein (HDL) cholesterol, and triglycerides. Childhood trauma was not measured at baseline, but in a separate questionnaire that took place an average of 5.6 years after the baseline measurement (SD = 1.3). Family history of anxiety and depression was assessed at the 2018 survey.

Outcomes

For the first two waves the MINI was used to measure MD in the past two weeks.[61] The 2018 survey measured lifetime MD status, age of onset, and the presence of an episode in the past year using the LIfetime Depression Assessment Self-report.[60] Both questionnaires are validated instruments assessing MD according to DSM-IV-TR criteria.[62]

The incidence rate was calculated as the number of new cases per 100 person years.[63] We divided the number of new onsets between wave 1 and the 2018 survey by the cumulative number of years at risk during this period. Among incident cases, we counted the time at risk as the age of onset minus the age at wave 1. We assumed that the average point when a new case emerges lies halfway through the year, so we subtracted half a year from this number.[55,64,65]

To study predictors of the onset of depression, we used an outcome that contrasted all subjects with a first onset of MD between wave 1 and the 2018 survey (i.e., MD not present at wave 1 and age of onset after wave 1) with all subjects who did not qualify for a MD diagnosis at any wave, nor reported lifetime MD. To study predictors of the recurrence of depression we selected all individuals at risk of recurrence, i.e., with at least one episode of MD before wave 1, but not at wave 1. We contrasted subjects with a new episode at wave 2 and/or 3 ('recurrence') with all subjects without episodes at waves 2 or 3 ('non-recurrence').

Statistical analysis

All analyses were performed in *R_3.5.2*.[66]

Missing data handling

Multiple Imputation by Chained Equations was performed on the complete dataset using R-package *mice_3.8.0*.[67] Ten imputed datasets were used and all estimates were pooled across the datasets.

Multicollinearity

We investigated the correlations between the risk factors (Supplementary Figure 1). Correlations higher than 0.2 were observed in 26 out of 253 possible combinations of risk factors, although none were higher than 0.8, so we did not exclude any predictors for reasons of redundancy or multicollinearity.

Relative importance analysis

To gain more insight into the contributions of individual risk factors to the outcomes of interest, we performed relative importance analyses using the R-package *MuMIn_1.43.17*.[68] Each potential risk factor was investigated as an independent variable in a univariable logistic regression analysis with either onset or recurrence as the dependent variable. In the first step of relative importance analysis, we ran multivariable logistic regression models with all possible combinations of significant risk factors from the univariable analyses. In the second step, the importance value for a particular risk factor, which can be interpreted as the probability that a risk factor will be included in the best model, was calculated by summing the Akaike weights for the models in which the risk factor appears, and dividing this number by the sum of the Akaike weights of all models (see Supplementary Methods).[52] The importance value was calculated in every imputed dataset, and the final model combined all risk factors for which the average importance value was over 50%. In order to facilitate future meta-analyses, we also ran a multivariate model with all risk factors that were significant in the univariable analyses.

Results

At baseline, n = 34,694 subjects had never experienced MD. Of this group, 6.9% (n = 2,046) developed at least one episode of MD before the 2018 survey (see Figure 1). This corresponds with 10.5 new cases per 100 person-years. At baseline, there were 6826 individuals at risk of recurrence, 34.1% (n = 2,326) of which developed a new episode at subsequent measurement points. See Table 1 for the baseline characteristics of both samples.

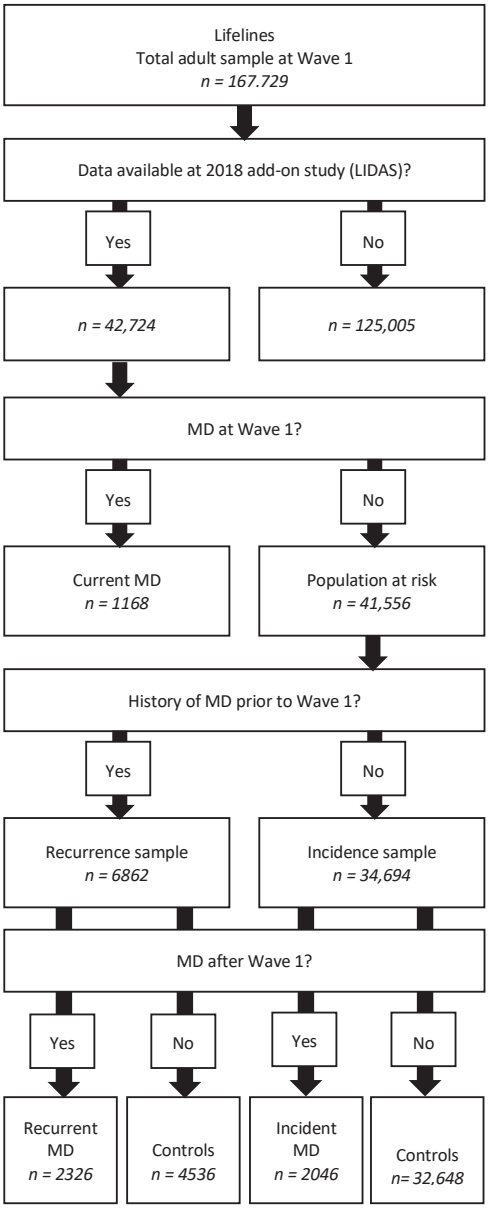


Figure 1. Sample selection

Flowchart of data selection for the incidence and recurrence samples.
LIDAS, LIfetime Depression Assessment Self-report; MD, Major Depression

Table 1. Baseline characteristics

	Onset		Recurrence	
	n	Value	*n*	Value
Sex, % female (SE)	24910	57.74 (0.31)	6659	72.95 (0.54)
Age, mean (SD)	24910	46.77 (12.19)	6659	45.69 (10.92)
Social and economic environment				
Education[1], % (SE)	24431		6533	
Low		25.86 (0.28)		21.54 (0.51)
Intermediate		39.44 (0.31)		41.37 (0.61)
High		34.71 (0.3)		37.09 (0.6)
Unfavourable job status[2], % (SE)	24894	5.87 (0.15)	6653	13.06 (0.41)
Income[3], % (SE)	21648		5937	
Low		16.28 (0.25)		20.53 (0.52)
Intermediate		55.80 (0.34)		54.24 (0.65)
High		27.92 (0.30)		25.23 (0.56)
N contacts past two weeks, mean (SD)	24659	19.22 (18.5)	6578	16.72 (15.96)
Health behaviors				
Physically active[4], % (SE)	23266	41.8 (0.32)	6275	47.41 (0.63)
Smoking status, % (SE)	24252		6504	
Non-smoker		46.61 (0.32)		39.48 (0.61)
Former smoker		37.17 (0.31)		39.45 (0.61)
Current smoker		16.21 (0.24)		21.06 (0.51)
Alcohol consumption, % (SE)				
Heavy drinker (≥ 6 drinks per drinking day)	23450	7.73 (0.17)	6366	7.37 (0.33)
Binge drinker (≥ 2 drinks per day on average)	23453	4.49 (0.14)	6366	3.82 (0.24)
Somatic health				
Physical quality of life (RAND-36), mean (SD)	24268	49.71 (7.57)	6459	49.87 (9.00)
Cancer[5], % (SE)	24881	1.01 (0.06)	6649	1.26 (0.14)
Cardiovascular problems[6], % (SE)	24894	2.02 (0.09)	6653	2.24 (0.18)
Metabolic syndrome (NCEP ATPIII criteria), % (SE)	24586	31.65 (0.30)	6515	33.84 (0.59)
Inflammatory disorders[7], % (SE)	24604	9.91 (0.19)	6606	13.02 (0.41)
Low-grade inflammation (CRP), mean (SD)	10348	2.43 (4.29)	2477	2.6 (4.76)
MD characteristics				
MD at wave 1 (past 2 weeks), % (SE)	24910	0 (N.A.)	6659	0 (N.A.)
MD at wave 2 (past 2 weeks), % (SE)	24355	0.39 (0.04)	4818	7.35 (0.38)
MD at the 2018 survey (past year), % (SE)	24910	4.13 (0.13)	6659	30.86 (0.57)
Lifetime MD (measured at the 2018 survey), % (SE)	24910	7.95 (0.17)	6659	100 (N.A.)
MD age of onset, mean (SD)	2908	41.44 (13.2)	6659	28.46 (11.67)
N anxiety disorders, mean (SD)	24910	0.05 (0.23)	6659	0.18 (0.47)
Family history of depression/anxiety, % (SE)	24795	48.47 (0.32)	6618	79.84 (0.49)
Neuroticism (NEO PI-R), mean (SD)	6550	-0.14 (0.94)	2201	0.62 (1.03)
Childhood trauma (CTQ), % (SE)	19475	21.97 (0.3)	4815	41.43 (0.71)
Acute and chronic stress				
Threatening events (LTE), mean (SD)	24397	1.04 (1.24)	6421	1.47 (1.53)
Chronic stress (LDI), mean (SD)	24395	2.04 (2.06)	6483	3.67 (2.71)

Sample characteristics at wave 1, based on complete data.

[1]Highest completed education: Low = junior general secondary education (mavo/vmbo-t) or lower, or no education; Intermediate = secondary vocational education (mbo), senior general secondary education (havo, vwo, hbs, mms); High = higher vocational education (hbo) or university.

[2]Unfavorable working conditions: being unemployed/looking for work, disabled for work, or on welfare.

[3]Net household income: low (< 1100), intermediate (1100-1899) and high (≥ 1900).

[4]The Dutch Movement Norm classifies physical activity as sufficient when participants report being active for at least half an hour on at least five days per week.

[5]Self-reported life time heart attack, aneurysm in aorta, heart failure, or stroke

[6]Self-reported cancer of any type, current

[7]Self-reported lifetime asthma, ulcerative colitis, rheumatoid arthritis, Crohn's disease, or celiac disease

CRP, C-reactive protein; CTQ, Childhood Trauma Questionnaire; LDI, Long-term Difficulties Index; LTE, List of Threatening Events; MD, Major Depression; NEO PI-R, Revised NEO Personality Inventory; NCEP ATPIII, National Cholesterol Education Programs Adults Treatment Panel III; RAND-36, Research and Development-36 (Dutch version of Medical Outcomes Study 36-Item Short Form Health Survey)

Key risk factors for new onset of MD

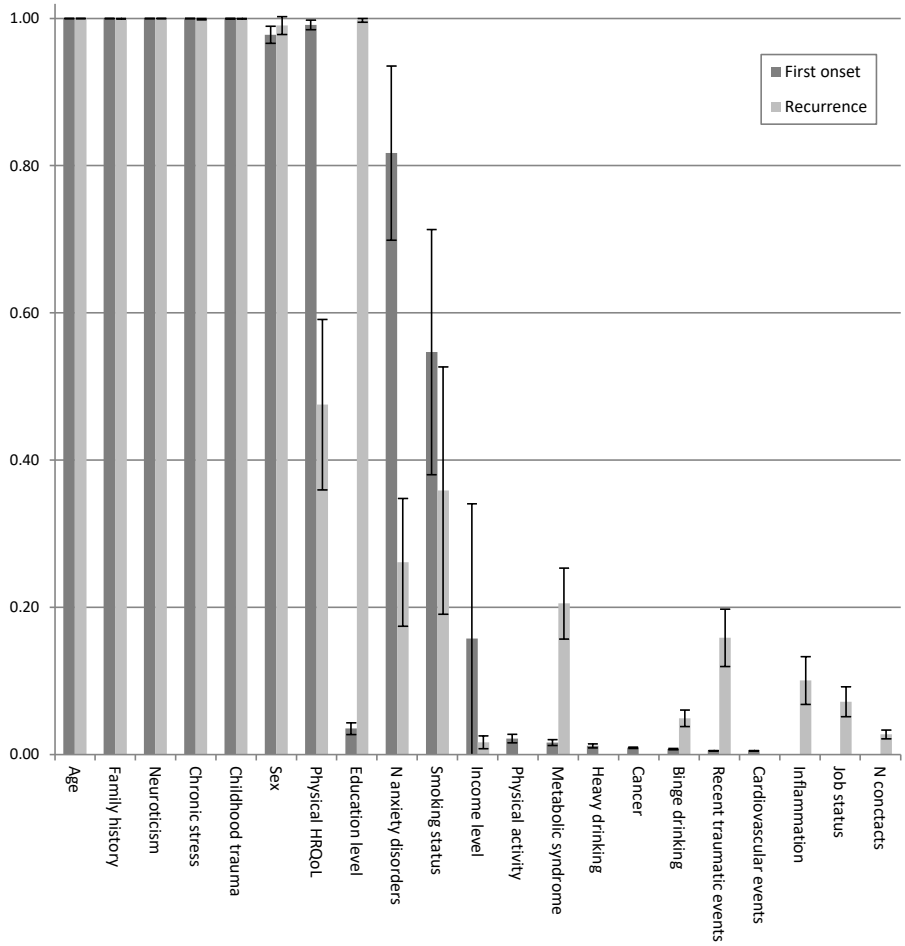
The relative importance of age, sex, family history of depression, neuroticism, childhood trauma, chronic stress, and physical QoL approached 100% (Figure 2). The relative importance of the number of anxiety disorders at wave 1 was also high (82%), whereas the relative importance of smoking behavior was 55%. The relative importance of education level, income, physical activity, the metabolic syndrome, drinking behavior, cancer, cardiovascular events, recent traumatic events, job status, and number of social contacts was low (1-16%). Similar risk factors remained significant in the multivariable model including all the significant risk factors from the univariable analyses (Supplementary Table 2). Family history was one of the strongest dichotomous predictors of incidence (odds ratio (OR) = 1.90) (Table 2). Similar risk (OR) was associated with having two additional anxiety diagnoses at wave 1, having problems with three additional domains of chronic stress (e.g., finances, health situation, or relationships), a 52-point difference on the Neuroticism dimension of the Revised Neuroticism-Extraversion-Openness Personality Inventory[69] (range 48-240), a 14-year age difference (n.b., younger people are at higher risk), or having a more than four standard deviations higher score on physical QoL. Childhood abuse, current smoking, and sex were weaker predictors (OR ~ 1.30).

Key risk factors for recurrent MD

Relative importance analyses showed that risk factors were equally likely to be included in the multivariable models for onset and recurrence, except for the following differences. Education level had a high importance value for MD recurrence, but not for onset (99.8% vs. 3.5%). Physical QoL and the number of anxiety disorders at wave 1 were important risk factors for MD onset, but their importance values for recurrence were only 48% and 26%, respectively. Family history was still one of the strongest dichotomous predictors

(OR = 1.40) in the final multivariable model (see Table 2), but lower education level, which was not included for onset, was the strongest predictor for recurrence of MD (OR = 1.52). Finally, male sex rather than female sex was a significant risk factor for recurrence.

Figure 2. Relative importance of each variable averaged over the imputed datasets



Relative weights for all risk factors that were significant in in univariable analyses for onset and recurrence of MD, averaged over ten imputed datasets, including error bars indicating the standard deviation.

Table 2. Final multivariable binomial regression models for onset and recurrence of MD

	Onset			Recurrence		
Intercept	0.24 (0.20-0.30)	-13.61	< 0.001*	0.72 (0.54-0.96)	-2.22	0.03*
Risk factor	**OR (95% CI)**	**F**	**P**	**OR (95% CI)**	**F**	**P**
Sex (female)	1.25 (1.13-1.39)	4.40	< 0.001*	0.74 (0.66-0.84)	-4.93	< 0.001*
Age (z-transformed)	0.57 (0.54-0.60)	-21.01	< 0.001*	0.76 (0.71-0.81)	-8.64	< 0.001*
Social and economic environment						
Education[1]						
Low	-	-	-	1.56 (1.35-1.80)	6.07	< 0.001*
Intermediate	-	-	-	1.29 (1.15-1.46)	4.20	< 0.001*
Health behaviors						
Smoking status						
Former smoker	1.04 (0.92-1.16)	0.60	0.55	-	-	-
Current smoker	1.31 (1.17-1.48)	4.48	< 0.001*	-	-	-
Somatic health						
Physical quality of life (RAND-36)	1.16 (1.09-1.24)	4.43	< 0.001*	-	-	-
N comorbid anxiety disorders (count 0-4)	1.33 (1.14-1.55)	3.65	< 0.001*	-	-	-
Family history of depression/anxiety	1.90 (1.72-2.09)	12.66	< 0.001*	1.40 (1.22-1.61)	4.76	< 0.001*
Neuroticism (NEO PI-R, z-transformed)	1.29 (1.21-1.37)	8.53	< 0.001*	1.28 (1.20-1.36)	7.71	< 0.001*
Childhood trauma (CTQ)	1.36 (1.21-1.54)	4.99	< 0.001*	1.31 (1.17-1.47)	4.63	< 0.001*
Chronic stress (LDI, z-transformed)	1.26 (1.20-1.33)	9.01	< 0.001*	1.16 (1.11-1.22)	5.78	< 0.001*

Multivariable logistic regression for onset of MD pooled over ten imputed datasets, using all risk factors for which the average importance value was over 50%. Risk factors are dichotomous absent vs. present, unless otherwise specified (i.e., count data, z-transformed continuous risk factors).

*p < 0.05

[1]Highest completed education: Low = junior general secondary education (mavo/vmbo-t) or lower, or no education; Intermediate = secondary vocational education (mbo), senior general secondary education (havo, vwo, hbs, mms); High = higher vocational education (hbo) or university.

CTQ, Childhood Trauma Questionnaire; LDI, Long-term Difficulties Index; NEO PI-R, Revised NEO Personality Inventory; RAND-36, Research and Development-36 (Dutch version of Medical Outcomes Study 36-Item Short Form Health Survey)

Discussion

This study aimed to identify key risk factors for new onsets of MD and compare these with risk factors for recurrence of MD. To this end, we performed relative importance analysis of a comprehensive collection of 21 potential risk factors that are known to be related to MD in a large representative general population sample. We found that age, sex, family history of depression/anxiety, neuroticism, childhood trauma, chronic stress, physical QoL, the number of anxiety disorders at wave 1, and smoking status were key predictors of onset. This means that in the context of these risk factors, other risk factors such as low amounts of physical movement, binge drinking or heavy drinking, and specific issues related to somatic health are of lesser importance. The key risk factors of MD recurrence were similar to those of onset, minus the number of anxiety disorders at wave 1, physical QoL, and smoking. Additionally, they included lower education levels. This risk factor was a stronger predictor for recurrence than family history, which was the strongest predictor for onset. Finally, women had a higher risk of onset, but this was not the case for recurrence. In summary, we have identified a number of possible targets for preventive interventions, including potentially important differences between those for first onset and recurrence of MD.

It is difficult to compare these findings with previous literature directly, because no previous studies have used relative importance analysis. Relative importance analysis is less likely to designate a risk factor as relevant compared to, for example, studies using standard multivariable approaches that present a final model including all risk factors that are significant in univariable analyses. Still, since there are no previous studies that have used the same approach, we will compare our findings with previous general population studies that investigated risk factors for onset and recurrence of MD.

Our findings regarding family history align with the results of the Baltimore Epidemiologic Catchment Area study, in which subjects with a first onset were twice as likely to have reported a parent with depression.[54] Most other general population studies investigating onset and recurrence of MD either excluded family history or used a different definition, precluding direct comparison.[55,65,70–75] Our results also correspond with earlier observations from clinical cohort studies[76–78], including a recent study showing that family history is a much stronger predictor of MD onset than polygenic risk of MD[79], and underline the potential of this risk factor as a target for screening and preventive interventions.[80–83] Since rearing experiences are thought to contribute just as much to trans-generational transmission of MD risk as genetic risk, targeted interventions to improve parenting skills in families with a history of MD are of paramount importance.[15,27–30]

Our findings also confirmed that the well-known gender gap in the prevalence of MD primarily relates to higher incidence in women, as women had a significantly

higher risk of MD onset (7.1% vs 4.4%), but a lower risk of MD recurrence (33.2% vs 35.9%).[54,55,65,75,84–87] This sex difference in MD onset arises in puberty, and is likely due to a combination of factors ranging from genetic and hormonal differences to heightened exposure to severe adversity and structural inequity.[87–89] However, whereas smaller studies often showed no effect of sex on MD recurrence[90–96], in our study the risk of MD recurrence was somewhat higher in men than in women, which calls for awareness of the importance of recurrence prevention in *both* men and women.

In contrast to most other longitudinal population studies, education was not a key risk factor for MD onset in our study, possibly because we included other, potentially confounding, risk factors such as age, income, and health behaviors.[54,55,65,74,75] However, we did find that education level was a strong risk factor for recurrence of MD, which is in line with a meta-analysis showing that education levels were a stronger risk factor for chronic course than for incidence.[97] Thus, strategies for tackling inequality in depression are needed, especially in relation to the course of the disorder.[98]

We also did not confirm a dose-response effect of number of anxiety disorders on MD recurrence[90–92,96], but our finding that this risk factor predicts MD onset is in line with earlier studies.[65,73,74,84] Still, the presence of any anxiety disorder after the MD index episode appears to be a relevant indicator for recurrence[92,96,99], and as such should still be monitored for.

Strengths and limitations

Strengths of the current study included the large general population sample, the longitudinal study design, the inclusion of both men and women and a wide age range, the high number of available risk factors, and the presence of thorough assessments with validated structured questionnaires that enabled us to investigate numerous risk factors for first onsets and recurrence. Furthermore, our use of relative importance analysis enabled us to investigate which risk factors were the most important for predicting onset and recurrence of MD, which is not possible with conventional model selection methods.[54–56,100] However, the results should also be interpreted in the context of several limitations.

First, although the number of included risk factors is large, some potential key risk factors have been excluded. For example, in order to increase the comparability with first onsets of MD, we did not include any risk factors related to the initial MD episodes, such as age of onset or duration, in the recurrence analyses.[96,99] Other potential key risk factors include variables related to the pathophysiology of MD (e.g., monoamine dysregulation[31], increased stress response[32], altered neurocircuitry[101,102]), but at the moment, the Lifelines sample only includes data related to inflammatory and metabolic dysregulations.

Second, there has been considerable discussion in the literature about how to define terms relating to the recurrence of MD.[103,104] Here, we assigned subjects suffering from MD prior to but not at wave 1 to the recurrent group when there was an episode at any of our post-baseline measurements. Unfortunately, due to missing information, we cannot be sure that the subjects assigned to the non-recurrent group were in remission/recovery in the whole period between the assessment waves.

Third, there might have been selective attrition in the sample due to MD or other factors such as higher age or lower education levels.[105–110] This might have led to an underestimation of the incidence and recurrence rates. It is difficult to ascertain the effects of selective attrition on our analyses because we cannot be sure which subjects dropped out due to developing MD after wave 1, which is a common limitation of longitudinal cohort studies.[107–109,111]

Finally, the onset and recurrence analyses were performed in different subsets of the Lifelines sample. Ideally, future waves of the Lifelines study could be used to investigate key risk factors for onset and recurrence in the same subjects to replicate the differences identified here. However, the sample for MD onset was already about three times larger than the recurrence sample, and this difference will only increase when onset and recurrence of MD are studied in the same subjects.[54] Differences in sample size produce differences in statistical power, meaning it is easier for weaker candidate risk factors to reach significance in univariable models for onset. However, this is especially problematic when this is the only criterion for inclusion in the final model, which was not the case here.

Conclusion

We identified a number of key risk factors relevant for population screening to identify subjects at risk of onset or recurrence of MD. For example, the risk of MD recurrence was higher in men than in women in our study, which calls for awareness of the importance of recurrence prevention in *both* men and women. Furthermore, the importance of lower education levels as a predictor for recurrence of MD suggests that strategies for tackling educational inequality in MD are needed, especially in relation to the course of the disorder. Finally, screening for MD among family members of depressed individuals may lead to more timely interventions. Future studies using relative importance analysis in similarly large samples are needed to confirm these results, as well as expand them to include other potential key risk factors.

References

1. Kessler, R. C. The costs of depression. *Psychiatr. Clin. North Am.* **35**, 1–14 (2012).
2. Whiteford, H., Ferrari, A. & Degenhardt, L. Global Burden Of Disease Studies: Implications For Mental And Substance Use Disorders. *Health Aff. (Millwood).* **35**, 1114–20 (2016).
3. World Health Organization. Investing in treatment for depression and anxiety leads to fourfold return. http://www.who.int/mediacentre/news/releases/2016/depression-anxiety-treatment/en/ (2016).
4. Greden, J. F. The burden of disease for treatment-resistant depression. *J. Clin. Psychiatry* **62**, 26–31 (2001).
5. Crown, W. H. *et al.* The impact of treatment-resistant depression on health care utilization and costs. *J. Clin. Psychiatry* **63**, 963–971 (2002).
6. Mauskopf, J. A. *et al.* Nonresponse, partial response, and failure to achieve remission: humanistic and cost burden in major depressive disorder. *Depress. Anxiety* **26**, 83–97 (2009).
7. Ivanova, J. I. *et al.* Direct and indirect costs of employees with treatment-resistant and non-treatment-resistant major depressive disorder. *Curr. Med. Res. Opin.* **26**, 2475–2484 (2010).
8. Lepine, B. A., Moreno, R. A., Campos, R. N. & Couttolenc, B. F. Treatment-Resistant Depression Increases Health Costs and Resource Utilization. *Rev. Bras. Psiquiatr.* **34**, 379–388 (2012).
9. Mrazek, D. A., Hornberger, J. C., Altar, C. A. & Degtiar, I. A review of the clinical, economic, and societal burden of treatment-resistant depression: 1996-2013. *Psychiatr. Serv.* **65**, 977–87 (2014).
10. Jorm, A. F., Patten, S. B., Brugha, T. S. & Mojtabai, R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* **16**, 90–99 (2017).
11. Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A. & Rosenthal, R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* **358**, 252–260 (2008).
12. Coe, R. It's the effect size, stupid. What effect size is and why it is important. in *British Educational Research Association annual conference* 1–18 (2002).
13. Cuijpers, P., Andersson, G., Donker, T. & Van Straten, A. Psychological treatment of depression: Results of a series of meta-analyses. *Nord. J. Psychiatry* **65**, 354–364 (2011).
14. Beshai, S., Dobson, K. S., Bockting, C. L. H. & Quigley, L. Relapse and recurrence prevention in depression: Current research and future prospects. *Clin. Psychol. Rev.* **31**, 1349–1360 (2011).
15. Ormel, J., Cuijpers, P., Jorm, A. F. & Schoevers, R. Prevention of depression will only succeed when it is structurally embedded and targets big determinants. *World Psychiatry* **18**, 111 (2019).
16. Werner-Seidler, A., Perry, Y., Calear, A. L., Newby, J. M. & Christensen, H. School-based depression and anxiety prevention programs for young people: A systematic review and meta-analysis. *Clinical Psychology Review* vol. 51 30–47 (2017).
17. van Zoonen, K. *et al.* Preventing the onset of major depressive disorder: A meta-analytic review of psychological interventions. *Int. J. Epidemiol.* **43**, 318–329 (2014).
18. Ferrari, A. J. *et al.* Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychol. Med.* **43**, 471–481 (2013).
19. Patten, S. B. Incidence of major depression in Canada. *CMAJ* (2000).

20. Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W. A. & Beekman, A. T. F. Prevalence and predictors of recurrence of major depressive disorder in the adult population. *Acta Psychiatr. Scand.* **122**, 184–191 (2010).

21. Burcusa, S. L. & Iacono, W. G. Risk for recurrence in depression. *Clin. Psychol. Rev.* **27**, 959–985 (2007).

22. Weissman, M. M. *et al.* Offspring of depressed parents: 30 years later. *Am. J. Psychiatry* **173**, 1024–1032 (2016).

23. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: Review and meta-analysis. *American Journal of Psychiatry* vol. 157 1552–1562 (2000).

24. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

25. Hyde, C. L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).

26. Howard, D. M. *et al.* Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* **9**, 1470 (2018).

27. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. The Rearing Environment and Risk for Major Depression: A Swedish National High-Risk Home-Reared and Adopted-Away Co-Sibling Control Study. *Am. J. Psychiatry* **177**, 447–453 (2020).

28. Kendler, K. S., Ohlsson, H., Lichtenstein, P., Sundquist, J. & Sundquist, K. The Nature of the Shared Environment. *Behav. Genet.* **49**, 1–10 (2019).

29. Kendler, K. S., Ohlsson, H., Sundquist, K. & Sundquist, J. Sources of parent-offspring resemblance for major depression in a national Swedish extended adoption study. *JAMA Psychiatry* **75**, 194–200 (2018).

30. Weissman, M. M. Is Depression Nature or Nurture? Yes. *American Journal of Psychiatry* vol. 177 376–377 (2020).

31. Albert, P. R., Benkelfat, C. & Descarries, L. The neurobiology of depression-revisiting the serotonin hypothesis. I. cellular and molecular mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2378–2381 (2012).

32. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).

33. Pan, A. *et al.* Bidirectional association between depression and metabolic syndrome: A systematic review and meta-analysis of epidemiological studies. *Diabetes Care* **35**, 1171–1180 (2012).

34. Milaneschi, Y., Lamers, F., Berk, M. & Penninx, B. W. J. H. Depression Heterogeneity and Its Biological Underpinnings: Toward Immunometabolic Depression. *Biological Psychiatry* vol. 88 369–380 (2020).

35. Valkanova, V., Ebmeier, K. P. & Allan, C. L. CRP, IL-6 and depression: A systematic review and meta-analysis of longitudinal studies. *J. Affect. Disord.* **150**, 736–744 (2013).

36. Clarke, D. M. & Currie, K. C. Depression, anxiety and their relationship with chronic diseases: A review of the epidemiology, risk and treatment evidence. *Med. J. Aust.* **190**, S54-60 (2009).

37. Katon, W. J. Epidemiology and treatment of depression in patients with chronic medical illness. *Dialogues Clin. Neurosci.* (2011).

38. Gaynes, B. N., Burns, B. J., Tweed, D. L. & Erickson, P. Depression and health-related quality of life. *J. Nerv. Ment. Dis.* **190**, 799–806 (2002).

39. Penninx, B. W. J. H., Milaneschi, Y., Lamers, F. & Vogelzangs, N. Understanding the somatic consequences of depression: biological mechanisms and the role of depression symptom profile. *BMC Med.* **11**, 129 (2013).

40. Jaworska-Andryszewska, P. & Rybakowski, J. K. Childhood trauma in mood disorders: Neurobiological mechanisms and implications for treatment. *Pharmacol. Reports* **71**, 112–120 (2019).

41. Wiersma, J. E. *et al.* The importance of childhood trauma and childhood life events for chronicity of depression in adults. *J. Clin. Psychiatry* **70**, 983 (2009).

42. Wang, J. L., Schmitz, N. & Dewa, C. S. Socioeconomic status and the risk of major depression: The Canadian national population health survey. *J. Epidemiol. Community Health* **64**, 447–452 (2010).

43. Meng, X. & D'Arcy, C. The projected effect of risk factor reduction on major depression incidence: A 16-year longitudinal Canadian cohort of the National Population Health Survey. *J. Affect. Disord.* **185**, 56–61 (2014).

44. Hartka, E. *et al.* A meta-analysis of depressive symptomatology and alcohol consumption over time. *Br. J. Addict.* **86**, 1283–1298 (1991).

45. Graham, K., Massak, A., Demers, A. & Rehm, J. Does the association between alcohol consumption and depression depend on how they are measured? *Alcohol. Clin. Exp. Res.* **31**, 78–88 (2007).

46. Fluharty, M., Taylor, A. E., Grabski, M. & Munafò, M. R. The Association of Cigarette Smoking With Depression and Anxiety: A Systematic Review. *Nicotine Tob. Res.* **1**, 3–13 (2017).

47. Mammen, G. & Faulkner, G. Physical activity and the prevention of depression: A systematic review of prospective studies. *Am. J. Prev. Med.* **45**, 649–657 (2013).

48. Hiles, S. A., Lamers, F., Milaneschi, Y. & Penninx, B. W. J. H. Sit, step, sweat: Longitudinal associations between physical activity patterns, anxiety and depression. *Psychol. Med.* **47**, 1466–1477 (2017).

49. Kendler, K. S. The dappled nature of causes of psychiatric illness: Replacing the organic-functional/hardware-software dichotomy with empirically based pluralism. *Mol. Psychiatry* **17**, 377–388 (2012).

50. Kendler, K. S. The structure of psychiatric science. *Am. J. Psychiatry* **171**, 931–938 (2014).

51. Babyak, M. A. Understanding confounding and mediation. *Evid. Based. Ment. Health* **12**, 68–71 (2009).

52. Burnham, K. P. & Anderson, D. R. Multimodel Inference: understanding AIC and BIC in Model Selection, Amsterdam Workshop on Model Selection. *Sociol. Methods Res.* **33**, 261–304 (2004).

53. Tonidandel, S. & LeBreton, J. M. Relative Importance Analysis: A Useful Supplement to Regression Analysis. *J. Bus. Psychol.* **26**, 1–9 (2011).

54. Eaton, W. W. *et al.* Population-based study of first onset and chronicity in major depressive disorder. *Arch. Gen. Psychiatry* **122**, 184–191 (2008).

55. De Graaf, R., Bijl, R. V., Ravelli, A., Smit, F. & Vollebergh, W. A. M. Predictors of first incidence of DSM-III-R psychiatric disorders in the general population: Findings from the Netherlands Mental Health Survey and Incidence Study. *Acta Psychiatr. Scand.* **106**, 303–313 (2002).

56. ten Have, M. *et al.* Duration of major and minor depressive episodes and associated risk indicators in a psychiatric epidemiological cohort study of the general population. *Acta Psychiatr. Scand.* **136**, 300–312 (2017).

57. Klijs, B. *et al.* Representativeness of the LifeLines Cohort Study. *PLoS One* **10**, e0137203 (2015).

58. Hölzel, L., Härter, M., Reese, C. & Kriston, L. Risk factors for chronic depression - A systematic review. *J. Affect. Disord.* **129**, 1–13 (2011).

59. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.* **44**, 1172–1180 (2015).

60. Bot, M. *et al.* Validity of LIDAS (LIfetime Depression Assessment Self-report): A self-report online assessment of lifetime major depressive disorder. *Psychol. Med.* **47**, 279–289 (2017).

61. Sheehan, D. V *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59**, 22–57 (1998).

62. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR*. (American Psychiatric Publishing, 2000).

63. Rothman, K. J., Greenland, S. & Associate, T. L. L. Modern Epidemiology, 3rd Edition. *Hastings Cent. Rep.* (2014)

64. Newman, S. C. & Bland, R. C. Incidence of mental disorders in Edmonton: Estimates of rates and methodological issues. *J. Psychiatr. Res.* **32**, 273–282 (1998).

65. Grant, B. F. *et al.* Sociodemographic and psychopathologic predictors of first incidence of DSM-IV substance use, mood and anxiety disorders: Results from the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions. *Mol. Psychiatry* **14**, 1051–66 (2009).

66. R Development Core Team. R: A Language and Environment for Statistical Computing. (2013)

67. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**, 1–67 (2011).

68. Bartón, K. MuMIn: Multi-model inference. (2020).

69. Costa, P. T. & McCrae, R. R. The revised NEO personality inventory (NEO-PI-R). in *The SAGE Handbook of Personality Theory and Assessment: Volume 2 - Personality Measurement and Testing* (2008).

70. Mattisson, C. *et al.* Risk factors for depressive disorders in the Lundby cohort - A 50 year prospective clinical follow-up. *J. Affect. Disord.* **113**, 203–215 (2009).

71. Murphy, J. M., Laird, N. M., Monson, R. R., Sobol, A. M. & Leighton, A. H. Incidence of depression in the Stirling County Study: Historical and comparative perspectives. *Psychol. Med.* **30**, 505–514 (2000).

72. Lukat, J., Becker, E. S., Lavallee, K. L., van der Veld, W. M. & Margraf, J. Predictors of Incidence, Remission and Relapse of Axis I Mental Disorders in Young Women: A Transdiagnostic Approach. *Clin. Psychol. Psychother.* **24**, 322–331 (2017).

73. Klein, D. N. *et al.* Predictors of first lifetime onset of major depressive disorder in young adulthood. *J. Abnorm. Psychol.* **122**, 1–6 (2013).

74. Bromberger, J. T. *et al.* Predictors of first lifetime episodes of major depression in midlife women. *Psychol. Med.* **39**, 55–64 (2009).

75. Beard, J. R., Heathcote, K., Brooks, R., Earnest, A. & Kelly, B. Predictors of mental disorders and their outcome in a community based cohort. *Soc. Psychiatry Psychiatr. Epidemiol.* **42**, 623–630 (2007).

76. Havinga, P. J. *et al.* Doomed for disorder? High incidence of mood and anxiety disorders in offspring of depressed and anxious patients: A prospective cohort study. *J. Clin. Psychiatry* **78**, e8–e17 (2017).

77. Leijdesdorff, S., Van Doesum, K., Popma, A., Klaassen, R. & Van Amelsvoort, T. Prevalence of psychopathology in children of parents with mental illness and/or addiction: An up to date narrative review. *Curr. Opin. Psychiatry* **30**, 312–317 (2017).

78. Van Sprang, E. D. *et al.* Familial risk for depressive and anxiety disorders: Associations with genetic, clinical, and psychosocial vulnerabilities. *Psychol. Med.* 1–11 (2020)

79. Agerbo, E. *et al.* Risk of Early-Onset Depression Associated With Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. *JAMA psychiatry* **75**, 387–397 (2021).

80. National Research Council and Institute of Medicine. *Depression in parents, parenting, and children: Opportunities to improve identification, treatment, and prevention*. (The National Acedemies Press, 2009).

81. Reedtz, C. *et al.* Promotion of Wellbeing for Children of Parents With Mental Illness: A Model Protocol for Research and Intervention. *Front. Psychiatry* **10**, 1–10 (2019).

82. Maciejewski, D., Hillegers, M. & Penninx, B. Offspring of parents with mood disorders: Time for more transgenerational research, screening and preventive intervention for this high-risk population. *Curr. Opin. Psychiatry* **31**, 349–357 (2018).

83. Havinga, P. J. *et al.* Prevention programmes for children of parents with a mood/anxiety disorder: Systematic review of existing programmes and meta-analysis of their efficacy. *Br. J. Clin. Psychol.* **60**, 212–251 (2021).

84. De Graaf, R., Ten Have, M., Tuithof, M. & Van Dorsselaer, S. First-incidence of DSM-IV mood, anxiety and substance use disorders and its determinants: Results from the Netherlands Mental Health Survey and Incidence Study-2. *J. Affect. Disord.* **149**, 100–107 (2013).

85. Wittchen, H. U. *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).

86. Baxter, A. J. *et al.* Challenging the myth of an 'epidemic' of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depress. Anxiety* **31**, 506–516 (2014).

87. Kuehner, C. Why is depression more commen among women than among men? *The Lancet Psychiatry* **4**, 146–58 (2017).

88. Kessler, R. C. Epidemiology of women and depression. *J. Affect. Disord.* **74**, 5–13 (2003).

89. Altemus, M., Sarvaiya, N. & Neill Epperson, C. Sex differences in anxiety and depression clinical perspectives. *Front. Neuroendocrinol.* **35**, 320–330 (2014).

90. Spijker, J. *et al.* Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *J. Affect. Disord.* **81**, 231–240 (2004).

91. Hardeveld, F. *et al.* Recurrence of major depressive disorder across different treatment settings: Results from the NESDA study. *J. Affect. Disord.* **147**, 225–231 (2013).

92. ten Have, M. *et al.* Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. *Acta Psychiatr. Scand.* **137**, 503–515 (2018).

93. Mattisson, C., Bogren, M., Horstmann, V., Munk-Jörgensen, P. & Nettelbladt, P. The long-term course of depressive disorders in the Lundby Study. *Psychol. Med.* **37**, 883–891 (2007).

94. Nöbbelin, L., Bogren, M., Mattisson, C. & Brådvik, L. Risk factors for recurrence in depression in the Lundby population, 1947–1997. *J. Affect. Disord.* **228**, 125–131 (2018).

95. Murphy, J. M., Olivier, D. C., Sobol, A. M., Monson, R. R. & Leighton, A. H. Diagnosis and outcome: depression and anxiety in a general population. *Psychol. Med.* **16**, 117–126 (1986).

96. Hoertel, N. *et al.* A comprehensive model of predictors of persistence and recurrence in adults with major depression: Results from a national 3-year prospective study. *J. Psychiatr. Res.* **95**, 19–27 (2017).

97. Lorant, V. *et al.* Socioeconomic inequalities in depression: A meta-analysis. *Am. J. Epidemiol.* **157**, 98–112 (2003).

98. Patel, V. *et al.* Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry* **17**, 76–89 (2018).

99. Buckman, J. E. J. *et al.* Risk factors for relapse and recurrence of depression in adults and how they operate: A four-phase systematic review and meta-synthesis. *Clin. Psychol. Rev.* **64**, 13–38 (2018).

100. Harrell, F. E., Lee, K. L., Matchar, D. B. & Reichert, T. A. Regression models for prognostic prediction: Advantages, problems, and suggested solutions. *Cancer Treat. Rep.* **69**, 1071–1077 (1985).

101. Ferrari, F. & Villa, R. F. The Neurobiology of Depression: an Integrated Overview from Biological Theories to Clinical Evidence. *Mol. Neurobiol.* **54**, 4847–4865 (2017).

102. Nestler, E. J. *et al.* Neurobiology of depression. *Neuron* **34**, 13–25 (2002).

103. Frank, E. *et al.* Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Arch. Gen. Psychiatry* **48**, 851–855 (1991).

104. Rush, A. J. *et al.* Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology* **31**, 1841–1853 (2006).

105. Marcellus, L. Are we missing anything? Pursuing research on attrition. *Can. J. Nurs. Res.* **36**, 82–98 (2004).

106. Lamers, F. *et al.* Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). *Compr. Psychiatry* **53**, 63–70 (2012).

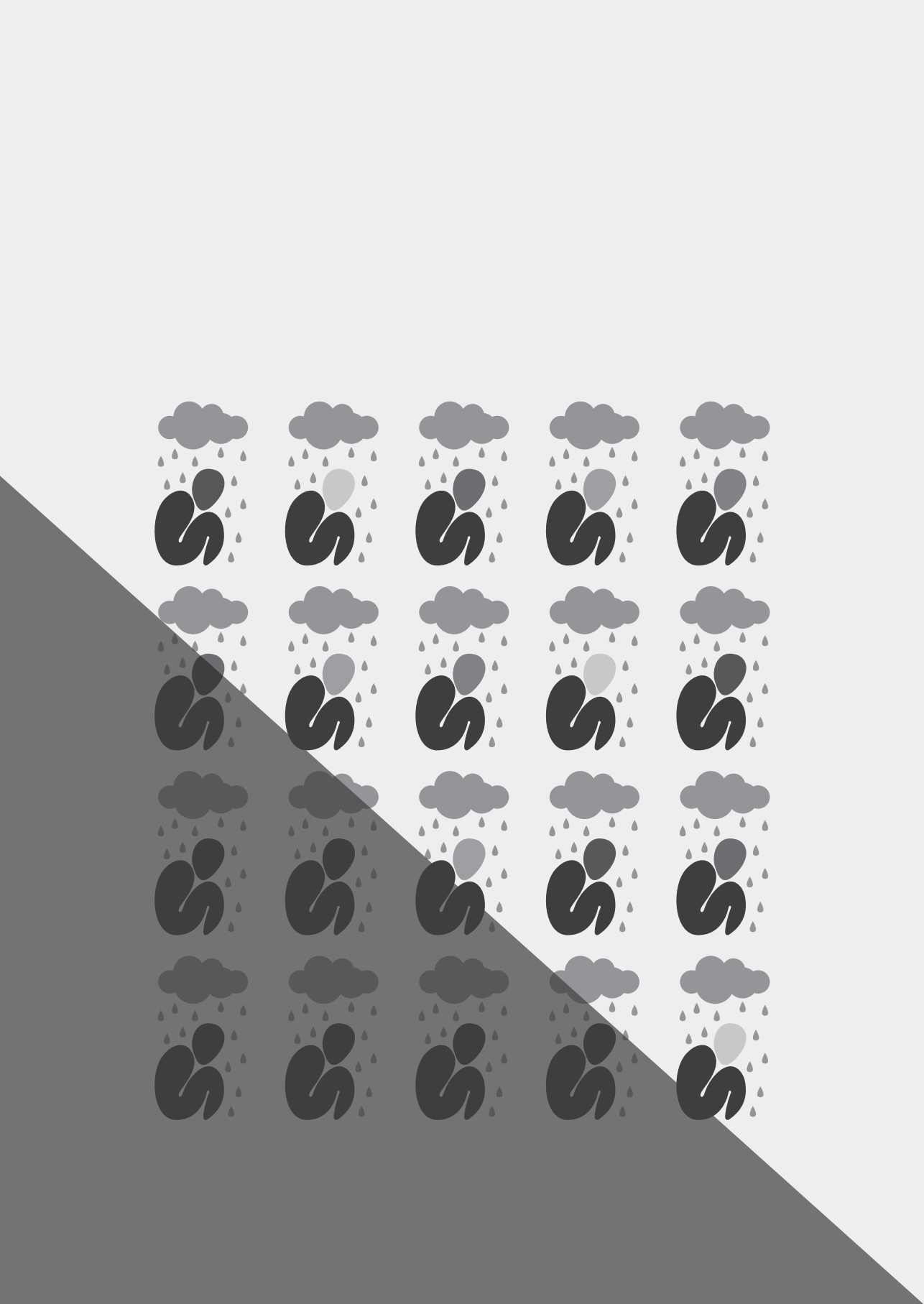
107. Eaton, W. W., Anthony, J. C., Tepper, S. & Dryman, A. Psychopathology and attirition in the epidemiologic catchment area surveys. *Am. J. Epidemiol.* **135**, 1051–1059 (1992).

108. De Graaf, R., Bijl, R. V., Smit, F., Ravelli, A. & Vollebergh, W. A. M. Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Am. J. Epidemiol.* **152**, 1039–1047 (2000).

109. De Graaf, R., Van Dorsselaer, S., Tuithof, M. & Ten Have, M. Sociodemographic and psychiatric predictors of attrition in a prospective psychiatric epidemiological study among the general population. Result of the Netherlands Mental Health Survey and Incidence Study-2. *Compr. Psychiatry* **54**, 1131–1139 (2013).

110. Chatfield, M. D., Brayne, C. E. & Matthews, F. E. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J. Clin. Epidemiol.* **58**, 13–19 (2005).

111. Eerola, M., Huurre, T. & Aro, H. The problem of attrition in a Finnish longitudinal survey on depression. *Eur. Jouranl Epidemiol.* **20**, 113–120 (2005).



Chapter 3

Prevalence of internalizing disorders, symptoms, and traits across age using advanced nonlinear models

Hanna M. van Loo, MD*, Lian Beijers*, Martijn Wieling, Trynke R. de Jong, Robert A. Schoevers, Kenneth S. Kendler

* These authors contributed equally to this work.

Psychological Medicine 1-10 (2021).

The supplementary material for this article can be found at https://doi.org/10.1017/S0033291721001148.

Abstract

Background

Most epidemiological studies show a decrease of internalizing disorders at older ages, but it is unclear how the prevalence exactly changes over the life course, and whether there are different patterns for internalizing traits, and for men and women. This study investigates the impact of age and sex on the point prevalence across different mood and anxiety disorders, internalizing symptoms, and neuroticism.

Methods

We used cross-sectional data on 146,315 subjects, aged 18-80 years, from the Lifelines Cohort Study, a Dutch general population sample. Between 2012-2016, five current internalizing disorders – major depression, dysthymia, generalized anxiety disorder, social phobia and panic disorder – were assessed according to DSM-IV criteria. Depressive symptoms, anxiety symptoms, neuroticism, and negative affect were also measured. Generalized additive models were used to identify nonlinear patterns of internalizing disorders and traits over lifetime, and to investigate sex differences.

Results

The point prevalence of internalizing disorders generally increased between the ages of 18-30 years, stabilized between 30-50, and decreased after age 50. The patterns of internalizing symptoms and traits were different. NA and neuroticism gradually decreased after age 18. Women reported more internalizing disorders than men, but the relative difference remained stable across age (relative risk ~ 1.7).

Conclusions

The point prevalence of internalizing disorders was typically highest between age 30-50, but there were differences between the disorders, which could indicate differences in etiology. The relative gender gap remained similar across age, suggesting that changes in sex hormones around the menopause do not significantly influence women's risk of internalizing disorders.

Introduction

Depressive and anxiety disorders occur across all age ranges and are associated with significant disability.[1,2] Yet, how exactly internalizing disorders differ across age and sex is a subject of debate and few studies have been able to study their patterns over lifetime in detail. More insight into these patterns can be used to identify target populations for public health interventions.[3] Furthermore, this insight could inform hypotheses on specific risk factors for internalizing disorders over the course of life. For example, it has been suggested that changes in women's reproductive hormones during the menopause increase their risk for internalizing disorders, but results are inconclusive.[4–8] Different developments in prevalence in men and women around the age of menopause could support this hypothesis.

The first question concerns the exact development of different internalizing disorders over lifetime. Most studies in the general population find a decrease of internalizing disorders in older age.[9–14] However, it remains unclear whether this decrease in prevalence is linear or nonlinear[12], and whether there are significant differences in trajectories across these various highly comorbid internalizing disorders[13].

Second, there is a clear gender gap in the prevalence of depression and anxiety disorders, with women being affected roughly twice as often as men.[4,15–17] However, is this true over the entire lifespan? Some studies suggest that the gender gap remains the same across the lifespan.[1,12,16,18,19] but other studies found a decreased[12,20,21] as well as an increased gap[22] in older ages.

Lastly, it is unclear whether there are significant differences in trajectories across these various highly comorbid internalizing disorders, and how these trajectories of internalizing disorders compare with the trajectories of internalizing symptoms and traits, such as depressive symptoms, anxiety symptoms, negative affect (NA), and neuroticism.[3,12,23] Insight in the difference between the trajectories of internalizing disorders, symptoms, and traits can inform discussions on classification, such as whether internalizing disorders and traits are sufficiently similar constructs so that traits could serve as measures of internalizing disorders for research and clinical purposes.[24,25]

The study of these questions requires large general population samples with well-measured phenotypes, and statistical methods that are able to identify potentially nonlinear developments. To date, no studies have used advanced nonlinear statistical methods to investigate the point prevalence of different internalizing disorders, symptoms and traits over the lifetime and compared these across sex.

Our aim is to investigate the prevalence of different internalizing disorders across age and sex, and compare the results of internalizing disorders with internalizing symptoms and traits. We investigate the point prevalence of major depression (MD), dysthymia (DYS), generalized anxiety disorder (GAD), panic disorder (PD), and social phobia (SPH) diagnosed at interview by DSM-IV criteria in a sample of 146,315 participants aged 18-80 years from Lifelines, a study in the Dutch general population. We also study the rates of depressive and anxiety symptoms, NA, and neuroticism. Generalized additive models (GAMs) allow us to model nonlinear patterns and test for significant differences in the development of the different internalizing disorders, symptoms and traits, and compare results for men and women.

Methods

Sample

The Lifelines Cohort Study is a multidisciplinary prospective population-based cohort study of 167,729 subjects in the north of the Netherlands. It was established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioral, physical, and psychological factors contributing to health and disease, with a special focus on multimorbidity and complex genetics.[26,27] Between 2006 and 2013, an index population aged 25–49 years was recruited via participating general practitioners. Subsequently, older and younger family members were invited to participate in Lifelines. In addition, adults could self-register via the Lifelines website. In total, 49% of the included participants were invited through their GP, 38% were recruited via participating family members, and 13% self-registered [27]. Most participants (57%) were included in 2012-2013.[28] Baseline data were collected for 167,729 participants.

The Lifelines adult study population is broadly representative for the adult population of the north of the Netherlands. Demographic, socioeconomic, and general health characteristics of the Lifelines cohort are described elsewhere.[28] All participants provided written informed consent. The Lifelines Cohort Study was approved by the Medical Ethics Committee of the University Medical Center Groningen, The Netherlands. In the current study, we included all baseline participants aged 18-80 years (n = 146,315) who had available data on one or more of the internalizing disorders or symptoms. We

excluded 299 participants over 80 years because of the low sample size for the statistical analyses.

Measurements

Internalizing disorders

Current MD, DYS, SPH, and GAD were assessed according to DSM-IV-TR criteria with a standardized diagnostic interview based on the Mini-International Neuropsychiatric Interview (MINI).[29] Trained medical assistants administered sections of the MINI to all participants during their visit to the research facilities and entered the responses into the computer. Conform DSM-IV-TR duration criteria, MD, DYS, GAD, and PD were rated as present if the subject reported the required symptoms in the past 2 weeks, 2 years, 6 months, and 1 month, respectively.[30] SPH was assessed during the past month. For further details, see Supplementary Methods.

Internalizing traits

Depressive and anxiety symptoms

Using the symptoms of MD and GAD assessed with the MINI, we created two sum scores for depressive (range 0-9) and anxiety symptoms (range 0-7). As above, MD symptoms were assessed in the past 2 weeks, and GAD symptoms in the past 6 months. Due to changes in the design of the interview, only part of the sample (n = 73,805) had data on additional symptoms of MD and GAD if the core criteria were absent. This subsample with complete data was used for the analyses of MD and GAD symptoms (Supplementary Methods).

Negative affect

Negative affect (NA) was assessed with the Positive and Negative Affect Schedule (PANAS) using 10 items including feeling irritable, ashamed, upset, nervous, guilty, scared, hostile, jittery, afraid, and distressed.[31,32] Subjects were asked to rate how often they experienced each item in the past 4 weeks on a 5-point Likert scale resulting in a score ranging from 10-50.

Neuroticism

Current neuroticism was assessed with the Revised NEO Personality Inventory.[33,34] The NEO PI-R Neuroticism subscale consists of 48 items covering the facets of anxiety, angry/hostility, depression, self-consciousness, impulsiveness, and vulnerability. Items were answered on a 5-point Likert scale resulting in a sum score ranging from 48 to 240.

The initial questionnaire excluded the depression and anxiety facets to limit the total length of the questionnaires, but these were added later. Here we only studied participants for whom complete data on all subscales on the NEO were available (n = 42,658).

Statistical analysis

Weighted point prevalence

Because women and certain age groups were overrepresented in Lifelines (Supplementary Figure 1), we used a person weighting factor based on age and sex to estimate the point prevalence of internalizing disorders and traits for the Dutch general population. Data on the sex and age distribution of the Dutch population in 2011 were derived from the CBS Statline data (Supplementary Methods).[35]

Generalized additive models

Generalized additive models (GAM) were used to assess the prevalence of internalizing disorders and traits over the lifetime. GAM are regression models that can identify and characterize nonlinear regression effects, by automatically determining the optimal combination of nonlinear basis functions (e.g. linear terms, polynomial terms, cubic terms, etc.).[36–38] Overfitting is prevented by minimizing a combination of the error and a non-linearity penalty.[37] All analyses were performed in R using the packages *mgcv_1.9.29* and *itsadug_2.3*.[38–40] We modeled the prevalence of each internalizing disorder, and the means of the symptom scores and neuroticism score as a (potentially) nonlinear function of age, and tested if there was a significant interaction effect between sex and age, i.e. if the patterns across age varied depending on sex. Subsequently, we modelled the patterns of the five internalizing disorders to investigate if the intercept and the pattern across age varied depending on the disorder type. For these models, the prevalence of any disorder served as the dependent variable, and the type of disorder was used as the independent variable. The reference classes were varied to make sure the results were robust.

Sensitivity analysis

Internalizing disorders are highly comorbid.[41,42] Therefore, we performed a sensitivity analysis by including a random intercept for each subject in the GAM. This random intercept accounted for individual variation in vulnerability for internalizing disorders, irrespective of age, so that the fixed effect of age on internalizing disorders on a group level could be estimated. As the current software was not able to run a generalized additive model with random effects for the full sample, we divided the sample in 10 random subsamples of 14,624 individuals each. These subsamples were matched to the full sample

based on age and sex distributions. Then, we performed the GAM *without* and *with* random intercepts for these 10 subsamples, and compared the results.

Because family history is an important risk for developing internalizing disorders, we also performed a sensitivity analysis by including a random intercept in the GAMs for individual disorders in the full sample. This random intercept accounted for family variation in vulnerability for internalizing disorders.

Results

Point prevalence

The included 146,315 participants had a mean age of 44.2 years (SD 12.7) and 58.6% were women (Table 1). The age and sex weighted point prevalence rates showed that current GAD was reported most frequently (3.7%), followed by MD (2.0%), DYS (1.0%), and SPH (0.8%). PD in the past month was rare (0.21%). The point prevalence rates differed significantly between all disorders as indicated by the significant intercepts (Supplementary Table 2). The unweighted prevalence rates were somewhat higher for all disorders because of the sex and age composition of Lifelines participants, including a higher percentage of women than the general Dutch population (Supplementary Table 1).

Lifetime patterns of internalizing disorders

All internalizing disorders showed significant nonlinear patterns over the lifetime (Figure 1, Supplementary Table 2). The general trend was that their prevalence increased from the age of 18 until the age of 30, stabilized until the age of 50, and then decreased. However, there were also differences between the disorders, as indicated by their significantly different curves. The prevalence of SPH and PD decreased relatively early in life, whereas the prevalence of MD peaked at two ages, around 30 years and 50 years, a pattern not seen with other disorders. Additionally, the prevalence of GAD and DYS dropped more steeply after the age of 50 than did the other disorders. The curves for GAD-DYS and for PD-SPH were not significantly different when changing the reference class, indicating no robust difference in their curves.

Table 1. Baseline characteristics

	n	Total	Men	Women
Demographics				
Sex	146315		41.42%	58.58%
Age, mean (SD)	146315	44.21 (12.74)	44.84 (12.78)	43.77 (12.69)
Education level, % (SE)[x]				
Low	142735	29.61 (0.12)	29.68 (0.19)	29.56 (0.16)
Intermediate	142735	40.12 (0.13)	38.67 (0.20)	41.15 (0.17)
High	142735	30.27 (0.12)	31.65 (0.19)	29.29 (0.16)
Internalizing disorders, % (SE)*				
MD (2 weeks)	146314	1.98 (0.04)	1.42 (0.05)	2.53 (0.06)
Dysthymia (2 years)	142549	1.04 (0.03)	0.77 (0.04)	1.30 (0.04)
GAD (6 months)	146315	3.71 (0.05)	2.79 (0.07)	4.62 (0.08)
PD (1 month)	146315	0.21 (0.01)	0.15 (0.02)	0.27 (0.02)
SPH (1 month)	146313	0.84 (0.03)	0.75 (0.04)	0.93 (0.04)
Any mood disorder	145793	3.00 (0.05)	2.19 (0.07)	3.81 (0.08)
Any anxiety disorder	146313	4.32 (0.06)	3.33 (0.08)	5.30 (0.08)
Any internalizing disorder	145956	5.82 (0.07)	4.41 (0.09)	7.22 (0.10)
Internalizing traits, mean (SD)*				
MD symptoms (range: 0-9)	73805	0.53 (1.16)	0.40 (1.02)	0.65 (1.28)
GAD symptoms (range: 0-7)	73805	1.04 (1.76)	0.80 (1.55)	1.27 (1.92)
Neuroticism (range: 48-240)	42658	119.65 (21.14)	115.35 (20.26)	124.15 (21.11)
Negative affect (range: 10-50)	138859	20.54 (5.24)	19.61 (5.02)	21.47 (5.28)

[x]Highest completed education: "Low" is completed junior general secondary education (mavo/vmbo-t) or lower, or no education; "Intermediate" is completed secondary vocational education (mbo), senior general secondary education (havo, vwo, hbs, mms); "High" is completed higher vocational education (hbo) or university.

*Age and sex weighted estimates to the average Dutch population in 2011. For unweighted estimates, see Supplementary Table 1.

DYS, dysthymia; GAD, generalized anxiety disorder; MD, major depression; PD, panic disorder; SD, standard deviation; SE, standard error; SPH, social phobia.

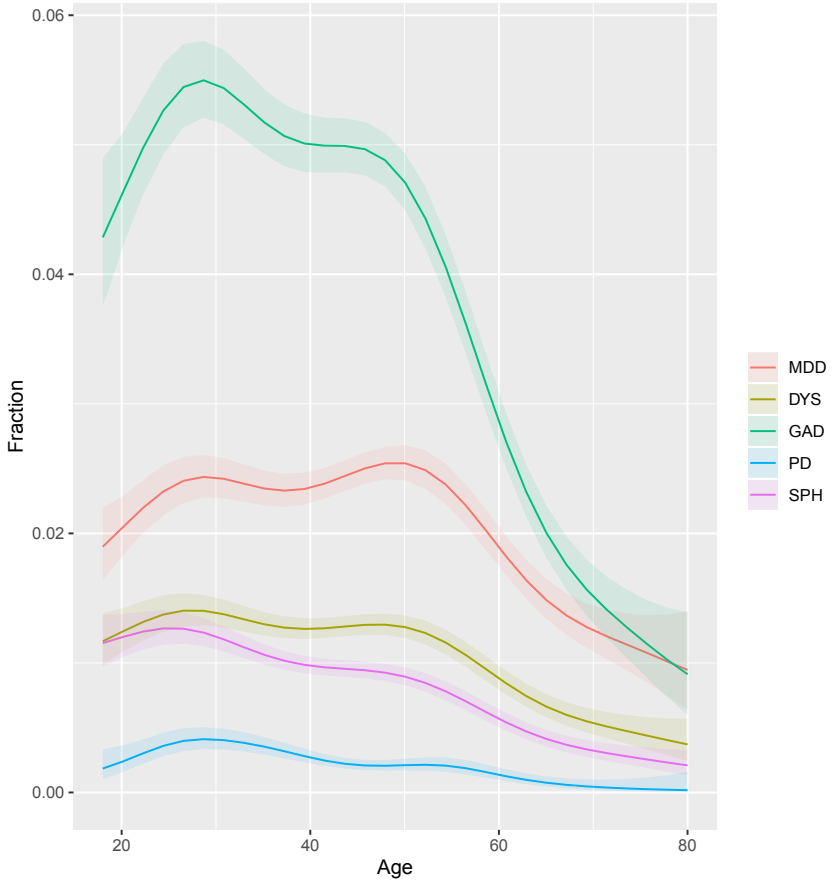


Figure 1. Estimated point prevalence for each internalizing disorder by age

Point prevalence for each internalizing disorder by age, as estimated by a generalized additive model. All patterns were nonlinear as indicated by the smoothing curves with effective degrees of freedom larger than 1 with P-values < 0.05 (Supplementary Table 2). The smoothing curves were all significantly different from each other except for SPH-PD and for DYS-GAD.

DYS, dysthymia; GAD, generalized anxiety disorder; MD, major depression; PD, panic disorder; SPH, social phobia.

Sex differences and similarities

As expected, women reported more internalizing disorders than men across the entire age range. The intercepts for each disorder were all significantly different for each disorder (Figure 2, Supplementary Table 3). However, the curves showing the increase and decrease of prevalence over age were not significantly different between the sexes, and this was true for each internalizing disorder. This implied that the odds ratio and the relative risk (i.e., prevalence women/prevalence men) were stable across the different age groups: about 1.7 for MD, DYS, GAD and PD, and 1.2 for SPH (Supplementary Table 4).

Comparison with internalizing symptoms and neuroticism

Internalizing symptoms and traits showed different patterns across age than did internalizing disorders (Figure 3, Supplementary Table 3). Depressive symptoms decreased slightly from age 18 until the age of 35, increased until the age of 50, and then decreased again until the age of 65, after which symptoms increased again. Anxiety symptoms increased until the age of 40, and then decreased, with a stabilization after age 70. Neuroticism and NA decreased largely linearly from the age of 18 years. NA diminished linearly except from an increase from the age of 45 until the age of 55, but this increase was minor (< 0.5 point on a scale from 10-50), and neuroticism stabilized from the age of 50.

Comparable to the internalizing disorders, women scored higher on all internalizing traits than men, especially in depressive and anxiety symptoms (ratio W/M ~ 1.6) and less for NA and neuroticism (ratio W/M ~ 1.1) (Supplementary Table 4). The curves for depressive symptoms were similar in men and women, meaning that the absolute difference in the number of depressive symptoms remained constant over lifetime. The curves for generalized anxiety symptoms, neuroticism, and NA were significantly different across sex, although Figure 3 shows that these differences were modest.

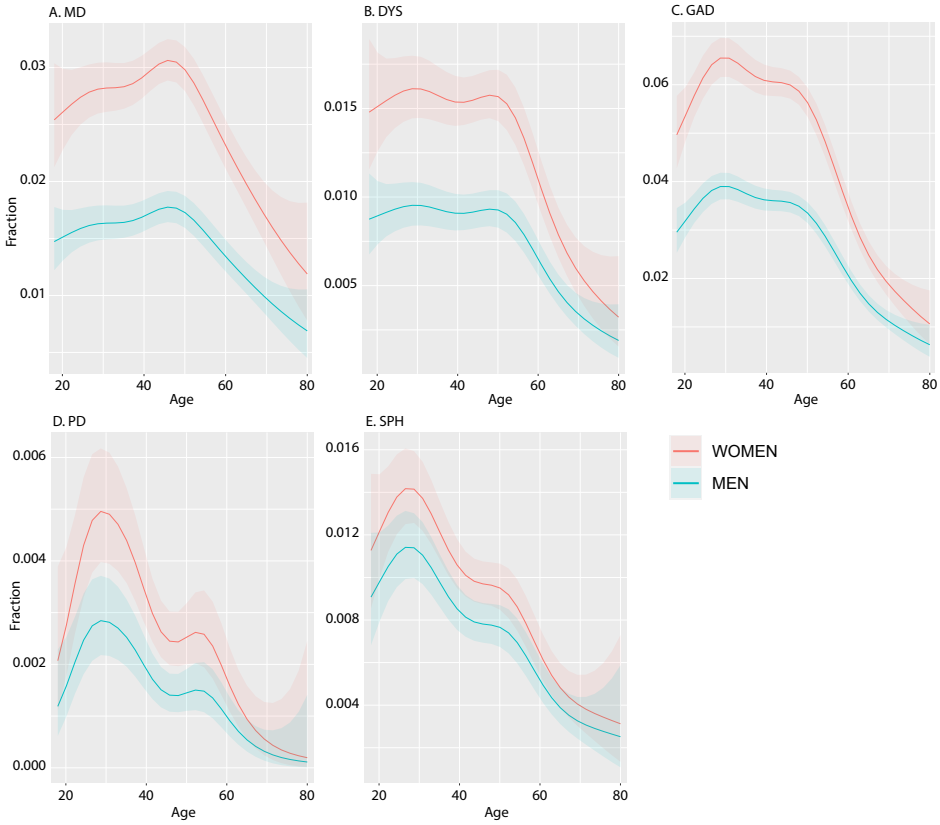


Figure 2. Estimated point prevalence for internalizing disorders in men and women

Point prevalence for both genders for each internalizing disorder by age, as estimated by generalized additive models for each disorder separately. For all five disorders there were differences in intercepts between men and women but smoothing curves were not significantly different (see Supplementary Table 3). Therefore, this figure is based on the models without interaction term.

DYS, dysthymia; GAD, generalized anxiety disorder; MD, major depression; PD, panic disorder; SPH, social phobia.

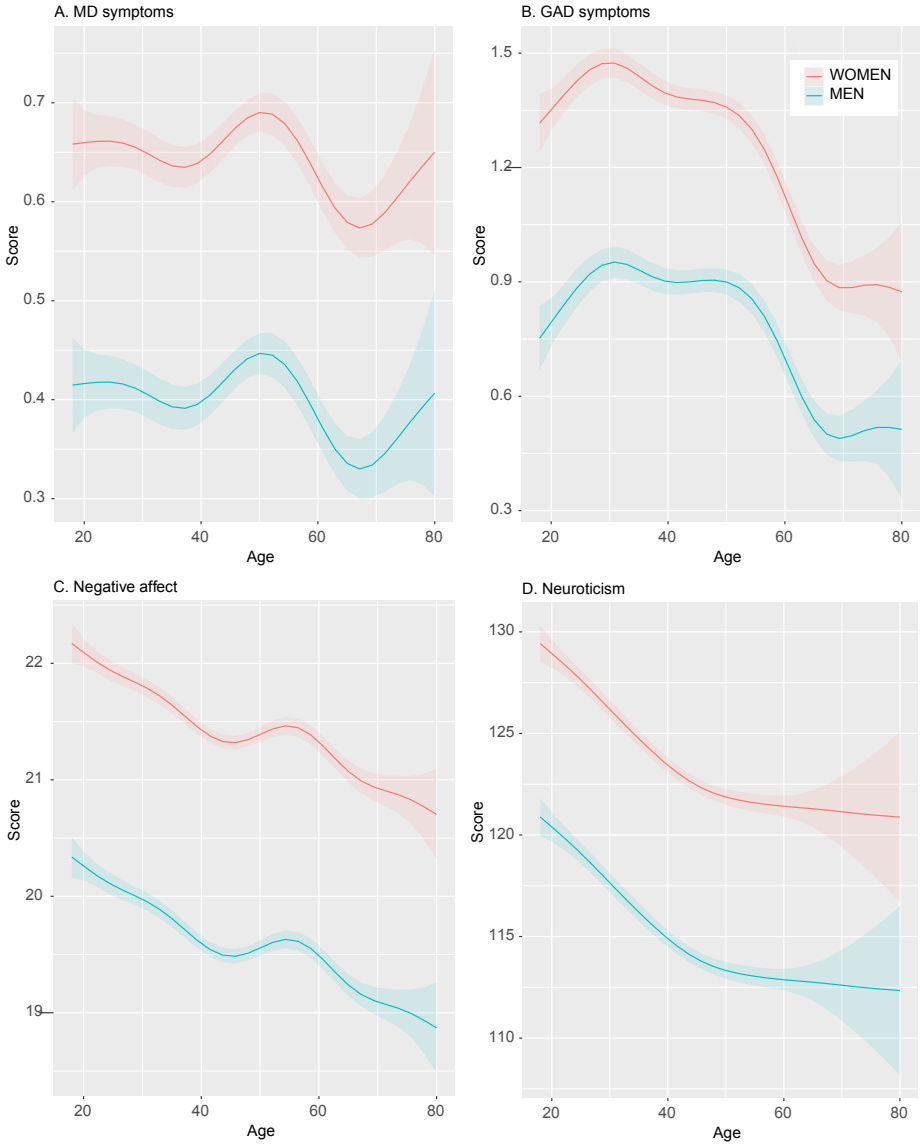


Figure 3. Estimated curves for internalizing traits in men and women

Average internalizing trait scores for both genders by age, as estimated by generalized additive models for each trait separately. As can be seen in Supplementary Table 3, there were differences in intercepts between men and women for each trait, and smoothing curves were also significantly different except for MD symptoms. Therefore, 3A is based on a model without interaction terms.

GAD, generalized anxiety disorder; MD, major depression.

Sensitivity analysis

To investigate the potential impact of comorbidity on the different trajectories of the internalizing disorders over lifetime, we performed a sensitivity analysis comparing models excluding and including random intercepts for each subject in 10 random subsamples each including about 10% of the sample. The estimated trajectories of the prevalence of internalizing disorders over lifetime were similar in all models including and excluding random intercepts (online Supplementary Table S5). To investigate the potential impact of the family structure of the Lifelines sample on our results, we performed another sensitivity analysis comparing models excluding and including random intercept for family structure. The estimated trajectories were again similar in all models including and excluding random intercepts (online Supplementary Table S6).

Discussion

In this study of 146,315 subjects from the Dutch general population aged 18-80 years, we investigated the patterns of the point prevalences of MD, DYS, GAD, SPH, and PD across different ages and sex. In general, our modeling indicated an increase in the prevalence of internalizing disorders from the age of 18 years, a plateau phase between 30-50 years of age, and a decrease after age 50. There were differences in the nonlinear patterns over lifetime between most disorders. Internalizing symptoms and neuroticism showed a distinctly different pattern over the lifetime compared with internalizing disorders. Although women reported more internalizing disorders and symptoms and higher neuroticism than men, the relative risk over the life course was remarkably similar.

To our knowledge, no previous studies used GAM to investigate the development of different internalizing disorders and traits over lifetime and across sex. Thus, we cannot directly compare the nonlinear patterns and statistical differences between the internalizing disorders, traits, and the sexes with results of previous studies. However, we can compare some key findings with previous findings.

First, our estimated point prevalences of the internalizing disorders are close to estimates of point prevalence in previous studies.[16,43–45] The relative differences in point prevalence for men and women are also as expected.[4,15–17,46]

Second, similar to this study, two reviews found that the point prevalence of internalizing disorders followed a nonlinear development over lifetime following an inversed U-shape.[16,45] Anxiety disorders manifested an initial rise in prevalence until age 30, followed by a decrease which was more pronounced after age 50, similar to our

findings.[16] The pattern for MD was slightly different –a rise in the prevalence of MD until age 50, followed by a decrease, and a second rise after age 75. This review also suggested similar curves for men and women across the lifespan.[16] Another review described an increase in the prevalence of DYS at early ages with a peak around 50 years.[45] Unlike our study, these reviews included studies with substantial heterogeneity, used relatively few data points (e.g., 141[45]), and did not formally test for nonlinearity or sex differences in their results.

Since this is the first study that used advanced nonlinear models to investigate the prevalence of internalizing disorders and traits across age and sex, we should be careful in drawing definitive conclusions. But if the results prove to be robust, they may have several implications.

First, the fact that the relative gender gap remains stable over the lifetime has implications for hypotheses about risk factors for internalizing disorders. Women clearly report more internalizing disorders than men. Previous studies showed that the gender gap in MD prevalence arises in puberty, due to higher incidence rates in women.[4,47,48] One of the hypotheses for this gender gap are changes in female sex hormones during lifetime, for instance around the menopause. There are suggestions that estrogens are neuroprotective, and a decrease in estrogens in menopause would increase women's risk of MD.[49] Several cross-sectional and longitudinal studies have studied the prevalence of MD and anxiety disorders around the menopause in women, but results are inconsistent.[4–8] Our study shows that around the age of the menopause, women indeed report more MD and depressive symptoms. However, there is a similar rise in MD and depressive symptoms in men in this age group. This implies that perimenopausal changes in female sex hormones probably do not significantly influence women's risk of depression, unless male hormonal changes or other male-specific risk factors exist that explain the similar increase in depression prevalence in middle-aged men. It is more likely that shared risk factors -e.g., psychosocial distress- explain the similar rise in depression prevalence in both sexes during midlife.[5] A similar argument can be made for anxiety disorders, in which the relative gender gap is also stable across age.

Second, the prevalence of most internalizing disorders showed different patterns over lifetime, which suggests that these disorders are not entirely identical constructs, but may have meaningful differences in etiology. At the same time, the similarity of the general pattern among the internalizing disorders suggests that there are likely shared risk factors.[50–52]

Third, the lifetime patterns of internalizing disorders differed from those of the internalizing traits, suggesting that the relationship between internalizing disorders and

traits is complex, or at least not stable across the lifespan. For instance, the prevalence of depressive symptoms, but not MD, was rising after the age of 65. This may be due to the fact that older subjects report somatic symptoms of depression more often without having episodes of MD.[53,54] In any case, the difference implies that we should be cautious in reducing internalizing disorders to high scores on symptom dimensions.[25,52]

In any case, the fact that internalizing disorders show different patterns across age and sex than internalizing symptoms and neuroticism is relevant for the debate on the nature and classification of internalizing disorders. In this debate, psychopathology is assumed to exist on a continuum instead of there being clear boundaries between health and disease.[25,55] Although we only investigated differences in prevalence rates, our data show that there may be important differences between internalizing disorders and symptoms and traits. This difference implies that we should be cautious in reducing internalizing disorders to high scores on symptom dimensions.[25,52] This concern is supported by genetic studies showing that depressive symptoms are not always good proxies for MD.[24,56]

Strengths and limitations

This is the first study that used advanced nonlinear models to investigate the development of internalizing disorders over lifetime in a large sample from the general population. The disorders were assessed with structured interviews by trained research assistants, and focused on current psychopathology to minimize recall bias.[57,58]

We also note a number of limitations. Our study uses cross-sectional data, and therefore cannot exclude period or cohort effects as an explanation for the change in point prevalence estimates across different ages. It is unlikely, however, that our findings are exclusively based on period and cohort effects. A recent study in 611,880 subjects from the US population controlling for period and cohort effects showed that prevalence of depressive episodes followed an inverse U-shaped curve with increasing prevalence from the age of 18 and decreased after age 32, and that psychological distress declined with age.[3] Also population studies that were performed two decades apart indicate that the reduction of internalizing disorders is associated with older age.[11,16,44] Future assessment waves of Lifelines would allow an investigation of age, period, and cohort effects.

Similar to these population studies, we observed a reduction in the prevalence of internalizing disorders at older age. There are two types of explanations for the decline of internalizing disorders; (1) age is protective against internalizing disorders, (2) age is not protective, but internalizing disorders are less frequently measured in older participants due to biases. Selection bias occurs when older individuals with MD are relatively less often participating in population studies than younger individuals with MD due to increased

risk of death, difficulty in establishing contact or increased refusals.[59–61] Reporting bias might be a result of older people being less likely to report symptoms of depression.[62,63] It is also possible that the prevalence of depression at older age is lower because individuals suffering from depression are more likely to have died earlier due to related causes such as heart problems (i.e. survivor bias).[64,65]. However, in Lifelines, we found no interaction effect between age and the presence of an internalizing disorder at baseline when predicting participation at follow up (2014–2017) (data not shown). This means that the impact of having a disorder on attrition for any reason was not different for older as compared to younger subjects, which makes selection bias a less likely explanation for the reduction in prevalence after age 50. Follow-up studies are needed to investigate explanations for the decline of internalizing disorders, symptoms, and traits in older participants.

Third, we assessed current internalizing disorders based on structured interviews with trained research assistants, which can be considered a strength. However, there were two limitations in the assessments. Disability was not assessed for MD and GAD, and DYS was not assessed in subjects who satisfied criteria for MD, which could have biased prevalence rates upwards and downwards, respectively. It is most likely that these biases have been minor given that our estimates of MD, GAD, and DYS are comparable to previous estimates.[11,16,44,45]

Fourth, the presence of internalizing symptoms may influence subjects' reports on internalizing traits like neuroticism, which could complicate disentangling between states and traits. Previous studies showed that subjects with depressive symptoms may temporarily score higher on neuroticism.[66,67] If internalizing symptoms indeed have a strong effect on neuroticism, then we would have expected to see a similarity between the patterns of internalizing symptoms and neuroticism across age. However, in our study, neuroticism scores were not showing the same patterns as depressive symptoms, generalized anxiety symptoms, or NA. For example, neuroticism scores were lower in subjects aged 30–50 years than in younger subjects, whereas depressive symptoms were higher in this age group. Although these findings do not fully exclude that internalizing symptoms may have influenced neuroticism scores, it shows that the influence in our study is probably modest.

Conclusion

This study identified different patterns in point prevalence for most internalizing disorders, symptoms and traits over lifetime. The overall prevalence of internalizing disorders and traits in women was higher than in men, but the patterns across age were remarkably similar in both sexes. These results indicate that certain hypotheses for the gender gap, e.g., the changes in female sex hormones during menopause, are unlikely explanations. Future studies are needed to investigate the causes for the initial rise in internalizing disorders and their decline at older age, taking into account the sex similarities.

References

1. Ferrari, A. J. *et al.* Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* **10**, e1001547 (2013).
2. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).
3. Twenge, J. M., Cooper, A. B., Joiner, T. E., Duffy, M. E. & Binau, S. G. Age, Period, and Cohort Trends in Mood Disorder Indicators and Suicide-Related Outcomes in a Nationally Representative Dataset, 2005-2017. **128**, 185–199 (2019).
4. Kuehner, C. Why is depression more commen among women than among men? *The Lancet Psychiatry* **4**, 146–58 (2017).
5. Rössler, W., Ajdacic-Gross, V., Riecher-Rössler, A., Angst, J. & Hengartner, M. P. Does menopausal transition really influence mental health? Findings from the prospective long-term Zurich study. *World Psychiatry* **15**, 146–154 (2016).
6. Judd, F. K., Hickey, M. & Bryant, C. Depression and midlife: Are we overpathologising the menopause? *J. Affect. Disord.* **136**, 199–211 (2012).
7. Vivian-Taylor, J. & Hickey, M. Menopause and depression: Is there a link? *Maturitas* **79**, 142–146 (2014).
8. Bryant, C., Judd, F. K. & Hickey, M. Anxiety during the menopausal transition: A systematic review. *J. Affect. Disord.* **139**, 141–148 (2012).
9. Scott, K. M. *et al.* Age patterns in the prevalence of DSM-IV depressive/anxiety disorders with and without physical co-morbidity. *Psychol. Med.* **38**, 1659 (2008).
10. Kessler, R. C. *et al.* Age differences in the prevalence and co-morbidity of DSM-IV major depressive episodes: Results from the WHO world mental health survey initiative. *Depress. Anxiety* **27**, 351–364 (2010).
11. de Graaf, R., ten Have, M., van Gool, C. & van Dorsselaer, S. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc. Psychiatry Psychiatr. Epidemiol.* **47**, 203–213 (2012).
12. Jorm, A. F. Does old age reduce the risk of anxiety and depression ? A review of epidemiological studies across the adult life span. *Psychol. Med.* **30**, 11–22 (2000).
13. Wells, J. E. *et al.* Prevalence, interference with life and severity of 12 month DSM-IV disorders in Te Rau Hinengaro: The New Zealand Mental Health Survey. *Aust. N. Z. J. Psychiatry* **40**, 845–854 (2006).
14. Trollor, J. N., Sachdev, P. S., Anderson, T. M., Andrews, G. & Brodaty, H. Age shall not weary them: Mental health in the middle-aged and the elderly. *Aust. N. Z. J. Psychiatry* **41**, 581–589 (2007).
15. Wittchen, H. U. *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).
16. Baxter, A. J. *et al.* Challenging the myth of an 'epidemic' of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depress. Anxiety* **31**, 506–516 (2014).

17. McLean, C. P., Asnaani, A., Litz, B. T. & Hofmann, S. G. Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* **45**, 1027–1035 (2011).

18. Cairney, J. & Wade, T. J. The influence of age on gender differences in depression. *Soc. Psychiatry Psychiatr. Epidemiol.* **37**, 401–408 (2002).

19. Kessler, R. C., McGonagle, K. A., Swartz, M., Blazer, D. G. & Nelson, C. B. Sex and depression in the National Comorbidity Survey. I: Lifetime prevalence, chronicity and recurrence. *J. Affect. Disord.* **29**, 85–96 (1993).

20. Bebbington, P. E. *et al.* The influence of age and sex on the prevalence of depressive conditions: report from the National Survey of Psychiatric Morbidity. *Psychol. Med.* **28**, 9–19 (1998).

21. Wittchen, H.-U. & Jacobi, F. Size and burden of mental disorders in Europe-a critical review and appraisal of 27 studies B. *Eur. Neuropsychopharmacol.* **15**, 357–376 (2005).

22. Kessler, R. C. *et al.* Age differences in major depression: Results from the national comorbidity survey replication (NCS-R). *Psychol. Med.* **40**, 225–237 (2010).

23. Keyes, K. M. *et al.* Age, period, and cohort effects in psychological distress in the United States and Canada. *Am. J. Epidemiol.* **179**, 1216–1227 (2014).

24. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat. Genet.* **52**, 437–447 (2020).

25. Kotov, R. *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).

26. Stolk, R. P. *et al.* Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur. J. Epidemiol.* **23**, 67–74 (2008).

27. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int. J. Epidemiol.* **44**, 1172–1180 (2015).

28. Klijs, B. *et al.* Representativeness of the LifeLines Cohort Study. *PLoS One* **10**, e0137203 (2015).

29. Sheehan, D. V *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59**, 22–57 (1998).

30. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. (American Psychiatric Association, 2000).

31. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *J. Pers. Soc. Psychol.* **54**, 1063–1070 (1988).

32. Crawford, J. R. & Henry, J. D. The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *Br. J. Clin. Psychol.* **43**, 245–265 (2004).

33. Costa, P. T. & McCrae, R. R. *Revised NEO Personality Inventory and NEO Five-Factor Inventory: Professional Manual, PAR.* (Psychological Assessment Resources, Inc., 1992).

34. Hoekstra, H., Ormel, J. & Fruyt, F. De. NEO-PI-R / NEO-FFI: Big Five Personality Inventory Manual. *Lisse: Swets & Zeitlinger* (2007).

35. Centraal Bureau voor de Statistiek (CBS). StatLine - Nederlandse bevolking 2011 naar geslacht en leeftijd. *Statline* (2020).

36. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. *Springer Series in Statistics* (Springer, 2011).

37. Wieling, M. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *J. Phon.* **70**, 86–116 (2018).

38. Wood, S. N. *Generalized additive models: An introduction with R, second edition*. *Generalized Additive Models: An Introduction with R, Second Edition* (CRC Press, 2017).

39. Marra, G. & Wood, S. N. Practical variable selection for generalized additive models. *Comput. Stat. Data Anal.* **55**, 2372–2387 (2011).

40. van Rij, J., Wieling, M., Baayen, R. H. & van Rijn, H. itsadug: Interpreting time series and autocorrelated data using gamms. (2016).

41. Kessler, R. C., Chiu, W. T., Demler, O. & Walters, E. E. Prevalence, Severity, and Comorbidity of 12-Month DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* **62**, 617–627 (2005).

42. Bijl, R. V, van Zessen, G., Ravelli, A., de Rijk, C. & Langendoen, Y. The Netherlands Mental Health Survey and Incidence Study (NEMESIS): objectives and design. *Soc. Psychiatry Psychiatr. Epidemiol.* **33**, 581–586 (1998).

43. Baxter, A. J., Scott, K. M., Vos, T. & Whiteford, H. A. Global prevalence of anxiety disorders: A systematic review and meta-regression. *Psychol. Med.* **43**, 897–910 (2013).

44. Bijl, R. V, Ravelli, A. & van Zessen, G. Prevalence of psychiatric disorder in the general population: Results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* **33**, 587–595 (1998).

45. Charlson, F. J., Ferrari, A. J., Flaxman, A. D. & Whiteford, H. A. The epidemiological modelling of dysthymia: Application for the Global Burden of Disease Study 2010. *J. Affect. Disord.* **151**, 111–120 (2013).

46. Kuehner, C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. *Acta Psychiatr. Scand.* **108**, 163–174 (2003).

47. Kessler, R. C. Epidemiology of women and depression. *J. Affect. Disord.* **74**, 5–13 (2003).

48. Altemus, M., Sarvaiya, N. & Neill Epperson, C. Sex differences in anxiety and depression clinical perspectives. *Front. Neuroendocrinol.* **35**, 320–330 (2014).

49. Georgakis, M. K. *et al.* Association of Age at Menopause and Duration of Reproductive Period With Depression After Menopause. *JAMA Psychiatry* **73**, 139 (2016).

50. Kendler, K. S. *et al.* The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. *Am. J. Psychiatry* **168**, 29–39 (2011).

51. Vink, D., Aartsen, M. J. & Schoevers, R. A. Risk factors for anxiety and depression in the elderly: A review. *J. Affect. Disord.* **106**, 29–44 (2008).

52. Schoevers, R. A., Beekman, A. T. F., Deeg, D. J. H., Jonker, C. & Van Tilburg, W. Comorbidity and risk-patterns of depression, generalised anxiety disorder and mixed anxiety-depression in later life: Results from the AMSTEL study. *Int. J. Geriatr. Psychiatry* **18**, 994–1001 (2003).

53. Balsis, S. & Cully, J. A. Comparing depression diagnostic symptoms across younger and older adults. *Aging Ment. Health* **12**, 800–806 (2008).

54. Hegeman, A. J. M., Kok, R. M., Van Der Mast, R. C. & Giltay, E. J. Phenomenology of depression in older compared with younger adults: Meta-analysis. *Br. J. Psychiatry* **200**, 275–281 (2012).

55. Kendell, R. & Jablensky, A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am. J. Psychiatry* **160**, 4–12 (2003).

56. Kendler, K. S. *et al.* Shared and specific genetic risk factors for lifetime major depression, depressive symptoms and neuroticism in three population-based twin samples. *Psychol. Med.* **49**, 2745–2753 (2019).

57. Kruijshaar, M. E. *et al.* Lifetime prevalence estimates of major depression: An indirect estimation method and a quantification of recall bias. *Eur. J. Epidemiol.* **20**, 103–111 (2005).

58. Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M. & Wittchen, H. U. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int. J. Methods Psychiatr. Res.* **21**, 169–184 (2012).

59. Holwerda, T. J. *et al.* The relationship between generalized anxiety disorder, depression and mortality in old age. *Int. J. Geriatr. Psychiatry* **22**, 241–249 (2007).

60. Schoevers, R. A. *et al.* Association of depression and gender with mortality in old age: Results from the Amsterdam Study of the Elderly (AMSTEL). *Br. J. Psychiatry* **177**, 366–342 (2000).

61. Beekman, A. T. F. *et al.* The natural history of late-life depression: A 6-year prospective study in the community. *Arch. Gen. Psychiatry* **59**, 605–611 (2002).

62. Knäuper, B. & Wittchen, H. U. Diagnosing major depression in the elderly: Evidence for response bias in standardized diagnostic interviews? *J. Psychiatr. Res.* **28**, 147–164 (1994).

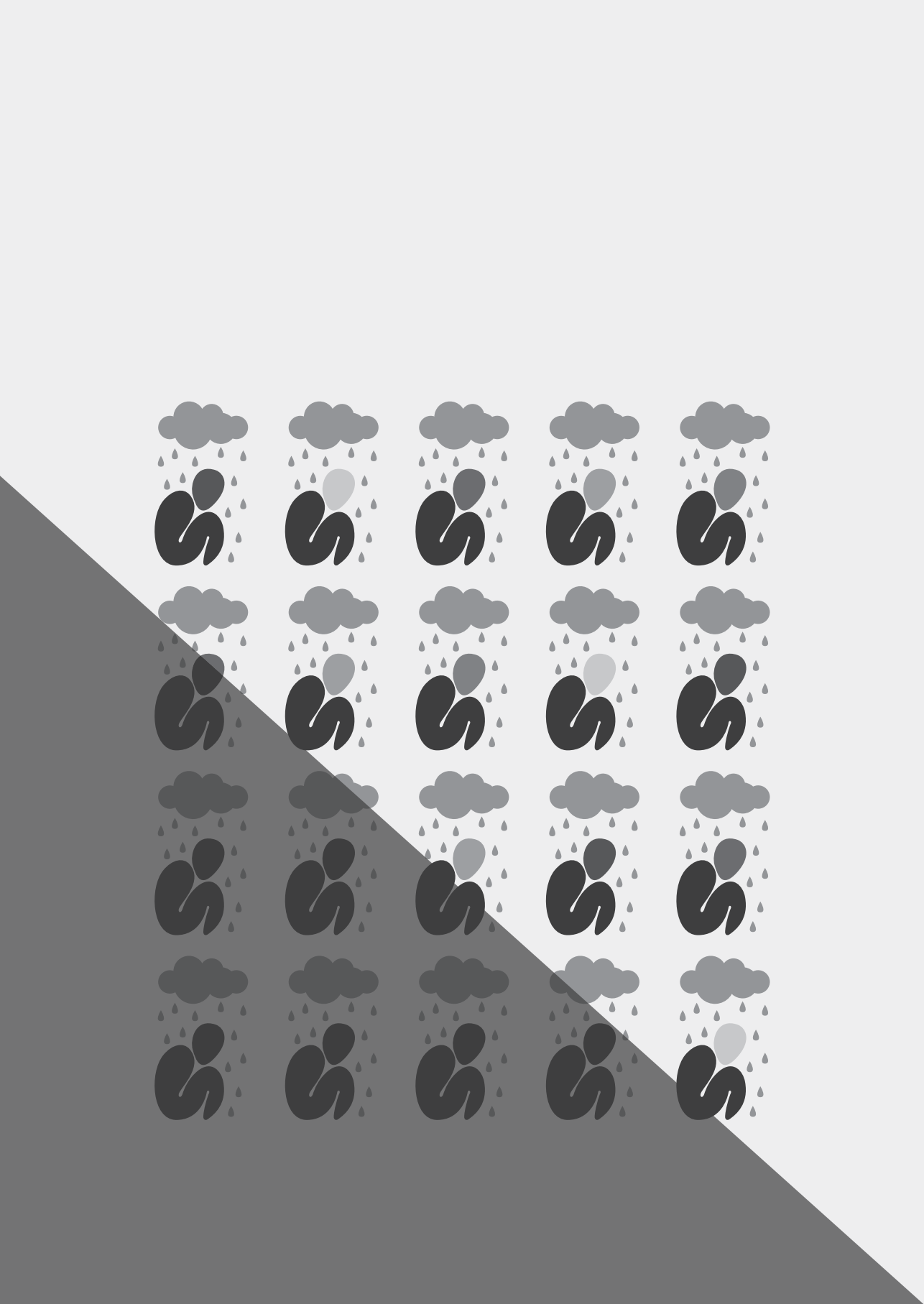
63. Lyness, J. M. *et al.* Older Age and the Underreporting of Depressive Symptoms. *J. Am. Geriatr. Soc.* **43**, 216–221 (1995).

64. White, L. O., Schulz, C., Klein, A. M. & von Klitzing, K. Declining effects of child maltreatment on mental health in the elderly: Another nail in the coffin of the deficit model of aging or a healthy survivor bias? *J. Affect. Disord.* **255**, 180–181 (2019).

65. Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).

66. Ormel, J. *et al.* Neuroticism and common mental disorders: Meaning and utility of a complex relationship. *Clin. Psychol. Rev.* **33**, 656–697 (2013).

67. Kotov, R., Gamez, W., Schmidt, F. & Watson, D. Linking 'Big' personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychol. Bull.* **136**, 768–821 (2010).



Chapter 4

Data-driven biological subtypes of depression: a systematic review of biological approaches to depression subtyping

Lian Beijers, Klaas J. Wardenaar, Hanna M. van Loo and Robert A. Schoevers

Molecular Psychiatry **24**, 888-900 (2019).

The supplementary material for this article can be found at https://doi.org/10.1038/s41380-019-0385-5.

Abstract

Background

Research into major depression (MD) is complicated by population heterogeneity, which has motivated the search for more homogeneous subtypes through data-driven computational methods to identify patterns in data. In addition, data on biological differences could play an important role in identifying clinically useful subtypes. This systematic review aimed to summarize evidence for biological subtypes of MD from data-driven studies.

Methods

We undertook a systematic literature search of PubMed, PsycINFO, and Embase (December 2018). We included studies that identified (1) data-driven subtypes of MD based on biological variables, or (2) data-driven subtypes based on clinical features (e.g., symptom patterns) and validated these with biological variables post-hoc.

Results

Twenty-nine publications including 24 separate analyses in 20 unique samples were identified, including a total of ~4000 subjects. Five out of six biochemical studies indicated that there might be depression subtypes with and without disturbed neurotransmitter levels, and one indicated there might be an inflammatory subtype. Seven symptom-based studies identified subtypes, which were mainly determined by severity and by weight gain vs. loss. Two studies compared subtypes based on medication response. These symptom-based subtypes were associated with differences in biomarker profiles and functional connectivity, but results have not sufficiently been replicated. Four out of five neuroimaging studies found evidence for groups with structural and connectivity differences, but results were inconsistent. The single genetic study found a subtype with a distinct pattern of SNPs, but this subtype has not been replicated in an independent test sample. One study combining all aforementioned types of data discovered subtypes with different levels of functional connectivity, childhood abuse, and treatment response, but the sample size was small.

Conclusion

Although the reviewed work provides many leads for future research, the methodological differences across studies and lack of replication preclude definitive conclusions about the existence of clinically useful and generalizable biological subtypes.

Introduction

Major depression (MD) is an important contributor to the global burden of disease.[1–6] Unfortunately, research into the disorder mechanisms is hampered by the heterogeneity of subjects with the same diagnosis[7–12], who can vary with regard to their symptom patterns, course trajectories, and treatment responses, and who may have different biological dysregulations or disturbances[13–17]. Identification of more homogeneous patient subgroups or subtypes is expected to improve our understanding of patient-specific etiological mechanisms and thus the development of more biologically informed, patient-specific diagnoses and treatments.[15,18–22] Data-driven approaches address this issue by using computational methods to identify patterns in data, which might otherwise be missed.

Although data-driven approaches to psychiatric diagnostics have long been used in psychiatric research[23–28], they have recently gained more popularity[29–31]. This is likely due to the growing realization that better-specified phenotypes are needed, but also due to the increasing availability of suitable datasets and ongoing advances in statistics and machine learning.[31–33]

Ideally, data-driven subtypes should capture subjects with similar etiological backgrounds. However, research into diagnostic subtypes of depression has thus far predominantly focused on subtyping based on symptom patterns or clinical features.[30,34] However, there is little evidence showing that heterogeneity in etiology and treatment response are best explained by variations at this phenomenological level. In different patients, similar clinical features or symptom patterns could be caused by different underlying mechanisms, limiting these features' usefulness to differentiate between etiologically distinct patients.[9,14,35] Therefore, it might be necessary to include other sources of heterogeneity, including biochemical markers, genetic variations, and brain region activity/connectivity. Indeed, research initiatives such as Research Domain Criteria and large-scale projects, such as the Roadmap for Mental Health Research in Europe have emphasized the need to incorporate multiple levels when investigating psychiatric disorder mechanisms.[36–38]

Given the above, it is important to gain insight into existing knowledge about the role of biological factors in MD heterogeneity. Therefore, this systematic review aimed to evaluate if consistent biological distinctions can be made between subtypes of MD. This systematic review encompassed all published studies that either used a data driven method to identify biological depression subtypes or identified psychometric symptom-based subtypes (e.g., symptoms or course) and investigated these subtypes' associations with biological factors.

Methods

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[39,40]. The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) in November 2017 (number: CRD42017079287).

Eligibility criteria

We included all studies (published up to 31 December 2018) that met the following criteria:

(1) Original studies (including pre-print e-publications).
(2) Subjects aged 18–65 years.
(3) At least 60% of patients satisfied the criteria for current or lifetime MD according to the Diagnostic and Statistical Manual of Mental Disorders (DSM III or later versions), the International Classification of Diseases (ICD-9 or ICD-10), or as indicated by a high score on a structured questionnaire (see Supplementary Methods for used cut-offs). Inclusion of controls up to 50% of the sample was allowed.
(4) Describing (a) data-driven depression subtypes based on biomarker input or (b) comparison of biological measurements between data-driven depression subtypes based on psychometric data.

Data-driven subtypes were defined as (sub)groupings of subjects that are identified using any kind of statistical clustering technique (e.g., latent variable analysis, hierarchical clustering). Psychometric data were defined as clinician- or self-report measurements of symptoms, functioning, or personality. Biomarkers were divided into biochemical, genetic, and neuroimaging markers. Biochemical biomarkers were defined as measurements of small molecules, proteins, or peptides in tissues such as cerebrospinal fluid (CSF), saliva, or blood. Genetic markers were defined as short DNA sequences (i.e., single-nucleotide polymorphisms, SNPs), longer sequences (e.g., mini-satellites), or other variables that require genotyping analysis (e.g., polygenic risk scores). Neuroimaging biomarkers were defined as some feature of the brain (e.g., structural features, connectivity, or activation patterns) measured by neuroimaging (e.g., magnetic resonance imaging, MRI).

Search strategy

The literature search was conducted in PubMed, Embase, and PsycINFO, using filters to select only original research papers. Nine known relevant papers were used to test the

sensitivity of different search terms. The retrieved articles were supplemented with studies cited in the reference lists and studies found by hand search.

Study selection

The literature search resulted in 128 articles. Analysis of the reference lists of included articles yielded 10 articles and a hand search yielded three articles, giving a total of 141 unique titles. Two independent raters reviewed these articles' titles and abstracts to identify studies that should be excluded (LB and HMvL, $\kappa = 0.69$). If at least one reviewer assessed a study as possibly relevant, it was included in the further review process. Of the initial 141 articles, 47 were selected for full-text review. Twenty-nine articles satisfied inclusion criteria (see Figure 1, for excluded papers see Supplementary Table S1).

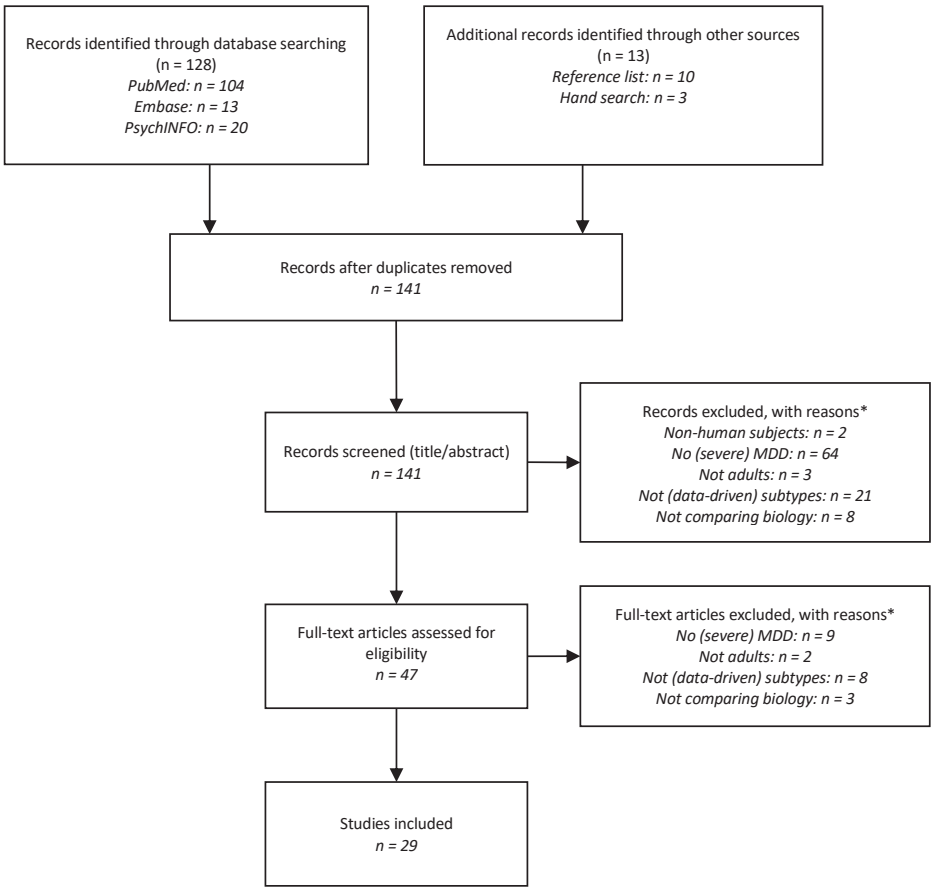


Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram

*All in/exclusion criteria were assessed for all papers, so the sum of all criteria might be larger than the total numbers of studies excluded based on full text. For specific reasons for exclusion per paper, see Supplementary Table S1.

Data collection

Using standardized Excel sheets, the first author extracted the following information for each article: the study aim(s), sample information (*n*, mean age, % female), methodology (biological or psychometric clustering, measures and statistical methods used), and results (subtype descriptions/ differences). When subtypes from earlier studies were used, all relevant information about the cluster identification was extracted from the paper described the original analysis.

Results

Study characteristics

Data-driven biological depression subtypes have received increasing attention recently (Figure 3). The 29 identified publications (Table 1) presented 25 separate analyses performed in 20 unique samples (n ~ 4000 for biological comparisons). Fourteen analyses used a data-driven method to identify biological subtypes among depressed patients. Nine analyses used a data-driven method to identify psychometric symptom-based subtypes and compared biological measures between them. One study combined psychometric and biological data.

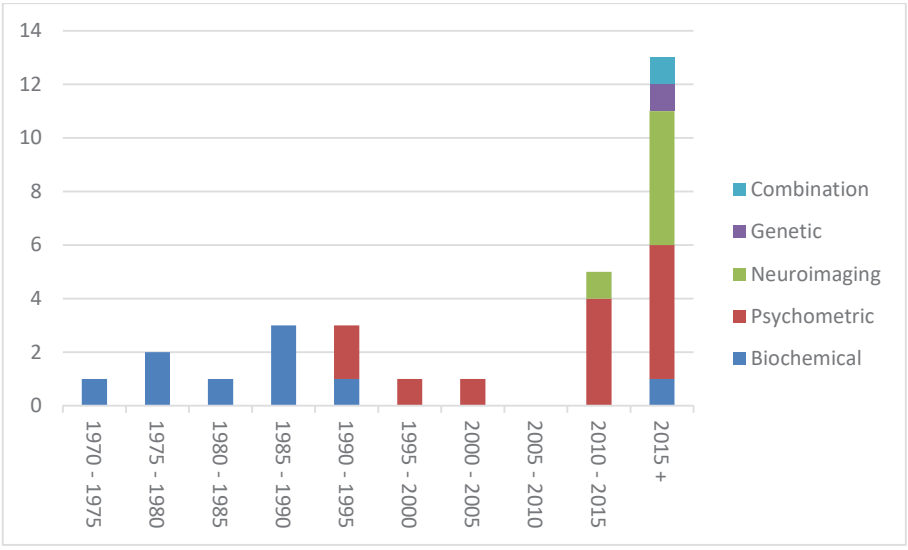


Figure 2. Types of study per 5-year period

Genetic: studies that use DNA sequences (i.e., single-nucleotide polymorphism), or other variables that require genotyping analysis (e.g., polygenic risk scores). Neuroimaging: studies that use some feature (e.g., structural features, or activation patterns) of the brain as indexed by neuroimaging techniques (e.g., electroencephalography, magnetic resonance imaging). Psychometric: studies that use clinician- or subject-scored measurements of symptoms, functioning, or personality. Biochemical: studies that use measurements of small molecules, proteins, or peptides in tissues such as cerebrospinal fluid, saliva, or blood. Combination: studies that use more than one type of data.

Table 1. Studies using data-driven methods to identify biologically distinct subtypes of depression

Study[a]	Sample size[b] (*n* = MD / total psycho-pathology)	Subtyping based on	Post-hoc tests	Statistical method	Nr. of sub-types	Short subtype descriptions (significant post-hoc differences in italics)
Biochemical subtyping studies						
Åsberg et al., 1973	36/43	CSF: 5-HIAA, IAA	Sex, age, response to nortriptyline treatment	Chi-square test for normality	2	1. High 5-HIAA *Significant response to treatment (n = 13)* 2. Low 5-HIAA *No significant response to treatment (n = 7)*
Åsberg et al., 1976a[c] **Åsberg et al., 1976a** **Åsberg et al., 1976b**	61/68	CSF: 5-HIAA	Sex, age, DRS (Cronholm & Ottoson), CPRS Nr. suicide attempts	Chi-square test for normality	2	1. Normal *15 % chance of suicide attempt* 2. Low 5-HIAA *40% chance of suicide attempt*
Maas et al., 1982	n. a./75	CSF: 5-HIAA, HVA, MHPG Urine: MHPG	-	-	1	n. a.
Gibbons et al., 1986	49/83	CSF: 5-HIAA, HVA, MHPG	Sex, age, height Previous medication use Polarity (unipolar/bipolar)	Gaussian Mixture Model	2	1. Normal. 2. Low 5-HIAA, HVA; normal MHPG *Average 4.9 cm taller*
Westenberg et al., 1988	n. a./38	CSF: 5-HIAA, HVA	-	Gaussian Mixture Model	2	1. Normal 2. Low 5-HIAA, low HVA
Davis et al., 1988	85/132	CSF: 5-HIAA, HVA, MHPG Urine: MHPG, epinephrine, norepinephrine, metanephrine, normetanephrine, vanillymandelic acid	Sex, age Polarity (unipolar/bipolar) Medication responders (imipramine/amitriptyline)	Gaussian Mixture Model	2	1. Normal levels 2. Higher catecholamine levels, low HVA
Maes et al., 1990[b,d]	96/100 females	Plasma: ACTH, TSH, T4, L-TRP, cortisol, DST	SCID	1. Forgy's centroid 2. Density search method 3. K-means	2[e]	1. Normal levels 2. Higher FT4, cortisol, ACTH, lower basal TSH, L-TRP Highest frequency of psychomotor symptoms[f], distinct quality of mood, early morning awakening, nonreactivity, decreased weight

Haroon et al., 2018	39/42	Left basal ganglia glutamate (MRS measurement) Plasma: CRP	Sex, age, race, education, smoking status, BMI Comorbid disorders, depression variables, multiple behavioral and cognitive variables Local- and network-level measures of functional integrity (based on BOLD-oscillatory activity)	Hierarchical clustering	2	1. Low glutamate/CRP 2. High glutamate/CRP More females; BMI higher; more smokers, lower AOO, more severe depression; lower functional integrity in left basal ganglia (based on 1/3 measures); lower network connectivity in network of 41 ROIs displaying lower functional integrity
Psychometric subtyping studies						
Maes et al., 1990a	n. a./100 females	SCID	Plasma: ACTH, TSH, T4, L-TRP, cortisol, DST	1. Forgy's centroid 2. Density search method 3. K-means	2	1. Highest frequency of psychomotor disturbances[f], distinct quality of mood, early morning awakening, nonreactivity, cognitive disturbances, decreased weight *Melancholic specifier more likely; worse scores on all markers* 2. Lower frequency of same symptoms *Less severe; no melancholic specifier; all minor depression*
Maes et al., 1992	61/80 males	SCID	Age, HAMD-17 Plasma: ACTH, TSH, T4, L-TRP, cortisol, DST	1. Forgy's centroid 2. Density search method	2	1. Highest frequency of psychomotor disturbances[f], distinct quality of mood, early morning awakening, nonreactivity, cognitive disturbances, loss of energy *Older; melancholic specifier more likely; more DST non-suppressors, lower TSH* 2. Lower frequency of same symptoms *Less severe; no melancholic specifier; all minor depression*
Schotte et al., 1997	165/220	SCID	Sex, age, marital status, employment status, hospitalization history, parasuicide, HAMD-17, BDI Plasma: DST	1. Forgy's centroid 2. K-means 3. Ward's error sum of squares	2	1. Highest frequency of psychomotor disturbances[f], distinct quality of mood, early morning awakening, nonreactivity, diurnal variation *More severe; more hospitalizations; melancholic specifier more likely; more DST non-suppressors* 2. Lower frequency of same symptoms *Less severe; no melancholic specifier*

Feighner et al., 2000		49/49	% change at 7 and 14 days in HAMD, MADRS, CSRS, CGIS	Sex, age, ethnic origin platelet 5-HT uptake	FASTCLUS; k-means with non-random initialization	4	1. Moderate non-responders *Lower 5-HT uptake* 2. Responders 3. High responders 4. Non-responders *Worst scores, lowest 5-HT uptake*
Orsel et al., 2010		65/78	SCID	Sex, age, age of onset, marital and employment status, personality disorder Plasma: TSH, T3, DST	1. K-means 2. Between groups linkage 3. Ward's methods	2	1. Highest frequency of psychomotor disturbances[f], distinct quality of mood, early morning awakening, nonreactivity, suicidal ideation, feelings of guilt *Melancholic specifier more likely; more DST non-suppressors; lower T3* 2. Lower frequency of same symptoms *Less severe; melancholic specifier less likely; all minor depression*
Lamers et al., 2010	**Lamers et al., 2012**	n. a./233	CIDI, IDS-SR	BMI, MbS components Plasma: CRP, IL-6, TNF-α Saliva: Cortisol measures (AUCi, AUCg, diurnal slope)	Latent Class Analysis	3	1. Highest frequency of decreased weight/appetite, insomnia, early morning awakening, diurnal variation *Higher AUCg, higher diurnal slope, more smokers* 2. Highest frequency of increased weight/ appetite, leaden paralysis *More females; lower alcohol intake; higher MbS scores, BMI, blood pressure, CRP, IL-6, TNF-α, diurnal cortisol slope, leptin, FABPa, complement C3, insulin and B2M; lower IGFBP1, IGFBP2, MSLN.* 3. Low frequency of all symptoms *Not included in biological analysis* BDNF did not differ between subtypes.
	Bus et al., 2014	743/818		Plasma: BDNF			
	Lamers et al., 2016	n. a./359		Sex, age, alcohol intake, smoking history, comorbid illness, medication use Plasma: A panel of 171 analytes involved in various hormonal, immunological, metabolic and neurotrophic pathways.			

Milaneschi et al., 2014[g]	**Milaneschi et al., 2014** **Milaneschi et al., 2015** **Milaneschi et al., 2016**	1178/1178 1733/1733 1176/1176	CIDI, IDS-SR	Sex, age rs9939609 A allele Plasma: Leptin GPRS for psychiatric and metabolic traits + SNP heritability	Latent Class Analysis	3	1. Highest frequency of decreased appetite/weight[h] *Younger; associated with schizophrenia GPRS* 2. Highest frequency of increased appetite/weight[h] *More females; highest percentage of rs9939609 A allele; higher leptin; associated with BMI and triglycerides GPRS* 3. Low frequency of all symptoms *Not included in biological analysis.*
Ballard et al., 2018		86/128	Composite SI score (BDI: Hopelessness and Suicidal Thoughts + HAM-D: Suicidal Thoughts)	Age, sex, BMI, smoking status, education levels, multiple variables about psychiatric (family) history Plasma: change and baseline levels of BDNF, VEGF, S100B, IDO, kynurenine, kynurenine acid, quinolinic acid, TNF-α, sTNFR1, IFN-γ, IL-1, IL-2, IL-5, IL6, IL-8, IL-10, cortisol	Growth Mixture Modelling	3	1. Non-responders *Higher chance of history of self-injury, SI at intake* 2. Responders 3. Remitters *Older; longer episode; higher chance of sexual abuse; higher baseline levels of IL-5*
Maglanoc et al., 2019		178/178	BAI-II, BDI	Sex, age, history of depression/anxiety/other Axis-I disorder, medication use Lagged and non-lagged BOLD correlations	High-dimensional data clustering		1. High unable to relax, symptom network more influenced by feelings of dislike, worthlessness, loss of interest *30-edge non-lagged functional connectivity subnetwork identifying this cluster* 2. High fear of worst happening, high symptom network more influenced by sadness, guilt, fatigue 3. High guilt, symptom network more influenced by fatigue, loss of energy and pleasure *24-edge non-lagged functional connectivity subnetwork identifying this cluster* 4. Lowest overall symptoms, high change in sleeping pattern 5. Low symptoms, some higher than in subgroup 3 *All subgroups: 22-edge non-lagged functional connectivity subnetwork with significant effect of subgroup*

Neuroimaging subtyping studies						
Cheng et al., 2014	61/61	Fractional anisotropy in ROI (different between subjects/controls)	Sex, age, age of onset, years of education, disease duration, HAMD-17, HAMA-14	Classification Analysis	2	Stronger WM connectivity *Age of onset 18 - 29* 2: WM microstructural deficits *Age of onset 30 - 45, older* *No differences in depression or anxiety severity*
Drysdale et al., 2017	n. a./220	CCA based on HAMD-17 items and BOLD-correlations (significantly correlated with ≥ 1 HAMD-17 items)	Age, medication use, HAMD-17 Response to rTMS	Hierarchical clustering	4	1. Low anhedonia-related connectivity, high anxiety related connectivity *Reduced connectivity in fronto-amygdala, high anxiety; highest response to rTMS* 2. Low anhedonia-related connectivity, low anxiety-related connectivity *Lowest total HAMD score.* 3. High anhedonia-related connectivity, low anxiety related connectivity *Hyperconnectivity in thalamic and frontostriatal networks, increased anhedonia/psychomotor retardation* 4. High anhedonia- and high anxiety-related connectivity
Feder et al., 2017	159/180	BOLD correlations between ROI (based on meta-analyses)	Sex, age, comorbid disorder/disease, disease duration (months), medication use, HAMD-17, CES-D	1. Hierarchical clustering 2. K-means	2	1. Higher connectivity *Longer duration, more severe* 2. Lower connectivity *Shorter duration, less severe*
Price et al., 2017a	68/68	Lagged BOLD correlations in ROI spanning VAN, DMN, CCN – at rest	Sex, age, diagnoses	S-GIMME making use of Walktrap	2	1. Subtype A. Fewer specific connections. Increased DMNA-VAN, DMN-DMN 2. Subtype B. greater overall connectivity, specifically in VAN-VAN and VAN-DMN and in CCN-driven paths *87.0% female; more comorbid anxiety*
Price et al., 2017b[i]	80/80	Lagged BOLD correlations in ROI spanning VAN, DMN, CCN – during positive emotion induction	Sex, age, diagnoses, symptom severity (from factor score)	S-GIMME making use of Walktrap	2	1. Subtype A. Directed pathway linking two DMN hubs 2. Subtype B. Lacking within-DMN connectivity, dACC-driven paths more prominent. *More severe affective symptoms*

Genetic subtyping studies						
Yu et al., 2017	203/203	Hamming distance based on: 1. 898 SNPs 2. 19 SNPs significantly associated with MD at P < 0.05.	-	1. Hierarchical clustering 2. Multi-dimensional scaling	2	1. Subtype clustered together based on the MD SNPs in both techniques. 2. All others.
Combination studies						
Tokuda et al., 2017	67/134	Sex, age BDI, MINI, HAMD, CATS, other psychiatric variables blood BDNF/ cortisol levels BOLD correlations between ROI (based on ICA) SNPs and methy-lation of BDNF/ serotonin genes	-	Multiple co-clustering based on non-parametric mixture models	5	1. High FC scores; low initial depression, unfavorable treatment response, high childhood abuse scores 2. Moderate FC scores; low initial depression, low childhood abuse scores 3. Low FC scores; high initial depression, favorable treatment response, high childhood abuse scores *Comparison of 3 depressed clusters, 2 healthy control clusters not included. Fourteen other models with 3-9 clusters not evaluated.*

5-HIAA, 5-hydroxyindoleacetic acid; 5-HT, 5-hydroxytryptamine; ACTH, adrenocorticotropic hormone; AUCi, area under the curve with respect to the increase; AUCg, area under the curve with respect to the ground; BDI, Beck depression index; BDNF, brain-derived neurotrophic factor; BMI, body mass index; BOLD, blood oxygen level dependent; CATS, childhood abuse trauma scale; CCN, cognitive control network; CES-D, center for epidemiological studies depression scale; CGIS, clinical global impression scale; CIDI, composite international diagnostic interview; CPRS, comprehensive psychopathological rating scale; CRP, C-reactive protein; CSRS, Carroll self-rating scale; DHEA-S, dehydroepiandrosteronsulfate; DMN, default mode network; DRS, depression rating scale; DST, dexamethasone suppression test; GPRS, genomic profile risk score; HAMA, Hamilton anxiety scale; HAMD, Hamilton depression rating scale; HVA, homovanillic acid; IAA, indole-3-acetetic acid; ICA, independent component analysis; IDO, indoleamine 2;3-dioxygenase; IDS-SR, inventory of depressive symptomatology self-report; IL, interleukin; IFN, interferon; LH, luteinizing hormone; L-TRP, tryptophan; MADRS, Montgomery-Åsberg depression rating scale; MbS, metabolic syndrome; MD, major depression; MHPG, 3-methoxy-4-hydroxyphenylglycol; MINI, mini international neuropsychiatric interview; MRS, magnetic resonance spectroscopy; ROI, region of interest; rTMS, repetitive transcranial magnetic stimulation; S100B, S100 calcium-binding protein B; SCID, structured clinical interview for DSM disorders; S-GIMME, subgroup group iterative multiple model estimation; SI, suicidal ideation; SNP, single-nucleotide polymorphism; sTNFR1, soluble tumor necrosis factor receptor 1; T3, triiodothyronine; T4, thyroxine; TNF-α, tumor necrosis factor alpha; TSH, thyroid-stimulating hormone; VAN, ventral affective network; VEGF, vascular endothelial growth factor; WM, white matter

[a]The second column lists studies making use of the same subtyping model, which is listed in the first column

[b]Sample size for biological comparison; sometimes the clustering was done on a larger sample. Other diagnoses: anxiety disorders, minor depression, and bipolar disorder

[c]Sample from Åsberg et al.[41], +25 new subjects. The DRS was used in the first sample. In the second sample, the CPRS was used

[d]Sample from Maes et al.[51]

[e]Number of classes was defined a priori. There was no comparison with models with different numbers of classes

[f]Both psychomotor agitation and retardation are contained in the 'psychomotor disorders' item from the SCID

[g]Sample from Lamers et al.[55], +lifetime depression subjects

[h]The frequencies of other symptoms were not compared statistically

[i]Sample from Price et al.[71]

Subtyping based on neurotransmitter distributions

Six biochemical subtyping studies investigated the existence of depression subtypes with different monoamine and catecholamine levels (Table 1). Most (5/6) studies assessed these levels in CSF in relatively few patients (mean n = 73) that were (periodically) off antidepressant medication. Two studies found one patient cluster with 5-hydroxyindoleacetic acid (5-HIAA, principal serotonin metabolite) levels that were similar to healthy controls, and another cluster with lower levels.[41–44]. One study found a unimodal distributions for 5-HIAA in a depressive sample.[45] Two studies identified a cluster with low levels of 5-HIAA and homovanillic acid (HVA, a major catecholamine metabolite).[43,44] Four studies also found a unimodal distribution for both CSF and urinary 4-hydroxy-3-methoxyphenyl glycol (HMPG, principal norepinephrine metabolite).[43,45–48] A study that measured 5-HIAA, HVA, and HMPG in CSF and epinephrine, norepinephrine, metanephrine, normetanephrine, and vanillymandelic acid in urine found two clusters. One cluster had levels similar to healthy controls. The other cluster had higher catecholamine levels and lower HVA.[48]

Overall, five out of six studies indicate that there might be depression subtypes with and without disturbed neurotransmitter levels. Unfortunately, little is known about the subtypes' associated clinical characteristics and outcomes. For example, because there was only a single study that contained information on previous medication use, we cannot be sure that there is no effect of tricyclic antidepressant use. None of the four studies investigating sex and age found any differences between clusters. A small number post-hoc comparisons were done (see Table 1), but they differed between the studies. For example, only one study (n = 68) compared severity scores (not significant).[42]

One biochemical subtyping study investigated depression subtypes with hierarchical clustering based on levels of glutamate in the basal ganglia (measured with magnetic resonance spectroscopy) and plasma C-reactive protein in 42 subjects with MD.[49] The first cluster had low levels of both, whereas the second had high levels. The first cluster showed higher functional integrity (i.e., the capability of the brain to respond flexibly to the environment, operationalized as the variability in blood oxygen level dependent (BOLD) signals[50]), but only for one out of three measures. Although this study applied a promising approach to investigate heterogeneity on a biochemical level, the sample was small.

Biological correlates of psychometric subtypes

Symptom-based subtyping

Four early psychometric symptom-based subtyping studies focused on validating the melancholic subtype through cluster analysis using depressive symptom scores.[51–54] Two

subtypes with different severity levels were identified in each study. The subtype with the highest symptom scores was labelled 'melancholic' in all studies, despite incomplete overlap with the DSM-specifier criteria. The 'melancholic' subtype in these studies showed more psychomotor disturbances, a distinct quality of mood, early morning awakening, and nonreactivity in all studies, but other symptoms (e.g., weight loss, loss of energy) differed only in part of the studies. The 'non-melancholic' subtype rarely showed melancholic features, contained most cases of minor depression, and showed lower general symptom scores. Comparisons of blood biomarkers showed more disturbances in the hypothalamus–pituitary–thyroid system (increased thyroxine[51] and reduced thyroid-stimulating hormone[51,52]), the hypothalamus–pituitary–adrenal axis (more non-suppressors in the dexamethasone suppression test[51–53]) and a decreased availability of L-tryptophan[51] in the 'melancholic' subtype compared with the 'nonmelancholic' subtype.

Two later studies used latent class analysis (LCA) using depressive to cluster subjects with a current or lifetime depression in the Netherlands Study of Depression and Anxiety (NESDA).[55,56] Both studies found one moderate and two severe subtypes. The moderate subtype had relatively low probabilities on all symptoms and contained a large proportion of lifetime depression cases. The first severe subtype showed higher probabilities of hypersomnia and increased appetite and weight (A+/W+; named 'severe atypical' by the authors). The second subtype showed higher probabilities of insomnia, early morning awakening, diurnal variation, and decreased appetite and weight (A–/ W–; named 'severe typical' by the authors) and showed the highest overall severity. The subtypes from the original analysis were found to be quite stable over 2 years (76%) and they were used in several subsequent NESDA studies to compare biological measures (usually excluding the moderate subtype).[55,57] These studies showed that the A+/W+subtype had a higher weight, higher leptin scores, higher metabolic syndrome risk[58,59], higher inflammation marker levels (e.g., C-reactive protein, interleukin-6)[54], lower cortisol-related score[58], and a higher probability of carrying a genetic variant of obesity-associated protein (FTO; rs9939609)[56,60] than the W–/A– subtype. Polygenic risk analyses showed that both the A–/W– subtype had a stronger association with polygenic risk for schizophrenia, whereas the A+/W+ subtype was associated with risk for high body mass index (BMI) and high blood triglyceride levels.[17] Proteomic analyses showed two markers that were lower (insulin-like growth factor-binding protein 1 and 2) and six that were higher (leptin, insulin, fatty-acid-binding protein, mesothelin, complement C3, beta-2-microglobulin) in A+/W+ subjects compared with A–/W– subjects.[61] Plasma brain-derived neurotrophic factor (BDNF) levels were similar in both subtypes.[62]

The most recent psychometric study used high dimensional data-driven clustering based on a combination of 21 depressive and 21 anxiety items.[63,64] This study identified five clusters with different symptom profiles, which mainly differed by severity and the number of subjects with and without history of depression. Subsequently, resting-state functional connectivity (see below) with and without time lag was compared between clusters. There was no difference in lagged functional connectivity, but the authors reported some differences between clusters in non-lagged connectivity, especially in the frontotemporal network and both the default mode network (DMN) component and praecuneus (parietal lobe). However, the exact details and relevance of these differences remained unclear.

In summary, four studies found two subtypes with different symptom levels, with the severe subtype showing some symptoms related to DSM-defined depression with a melancholic specifier and several specific biological correlates. The two studies that included disaggregated appetite/ weight symptoms found one low-severity subtype and two severe subtypes that mainly differed in terms of appetite/ weight change. The A+/ W+ subtype was associated with several weight and metabolism-related biomarkers and the A–/W– subtype was associated with higher cortisol dysregulation. The final study found subgroups that mainly differed in symptom severity but might have some different functional connectivity patterns. Together, these results indicate a possible connection between biological dysregulation and specific depressive symptom profiles. Still, they provide no conclusive evidence for the actual existence of DSM-defined melancholic or atypical subtypes, mainly because the analyses did not always include all criterion symptoms and the number of independent replications has been limited.

Medication response

Studies of plasma drug concentrations were not included in this review because they were not considered biochemical biomarkers. Two medication response studies met the inclusion criteria.

The first study clustered MD subjects based on improvement after treatment (at 7- and 14-day follow-up) with netamiftide (a synthetic pentapeptide antidepressant) and subsequently measured platelet 5-hydroxytryptamine (5-HT) uptake levels.[65] This study found four subgroups with increasing levels of symptom improvement. Subgroups with higher levels of improvement contained more patients with medication blood levels above the minimum projected therapeutic concentration with corresponding platelet 5-HT uptake levels. Although interesting, this study was small (n = 49) and has not been replicated.

The second study clustered subjects based on improvements in suicidal ideation after ketamine treatment.[66] Growth mixture modelling identified three clusters of non-responders, remitters, and responders. Logistic regression showed some modest differences in clinical characteristics. The non-responder cluster reported more often self-harm behavior and suicidal ideation. Many biomarkers were compared (see Table 1), but there was only one significant difference (higher baseline IL-5 was associated with higher odds of being in the responder group vs the remitter group). Nevertheless, this study applied a sophisticated approach, and future medication response studies could benefit from a similar approach to study heterogeneity in MD treatment outcomes and their (biological) correlates.

Subtyping based on neuroimaging

Three functional connectivity studies used resting-state functional magnetic resonance imaging to gain insight into interactions between brain regions that occur when a subject is at rest. This technique makes use of the BOLD signal, which exhibits low-frequency spontaneous fluctuations in the resting brain. These fluctuations show a high degree of temporal correlation across widely separated brain regions with demonstrable structural connections, suggesting potential functional brain networks.[67]

Each of the studies identified two subtypes, but the identified subtypes differed considerably. One study used hierarchical clustering on BOLD correlations between 38 regions of interest (ROIs) that were selected based on a meta-analysis.[68,69] The two resulting subtypes mainly differed in symptom duration and severity, and connectivity differences were not reported. Another study used hierarchical clustering on BOLD correlations between 15 ROIs from the ventral affective network, DMN, and cognitive control network.[70,71] The authors found two subtypes that they labelled 'typical' and 'atypical'. The typical subtype (71%) exhibited a distinct pattern of connectivity across DMN nodes, comparable to the patterns previously reported in depressed subjects.[70] The atypical subtype showed a divergent connectivity profile lacking DMN connectivity, but with increased dorsal anterior cingulate driven connectivity. In another study, the same authors investigated similar BOLD correlations during positive mood induction and again found a typical and atypical connectivity subtype.[72] However, these subtypes showed no significant overlap with those from the previous results.

The fourth functional connectivity study identified four clusters based on a different analytical approach.[73] Canonical correlation analysis, mapping correlations between linear combinations of functional connectivity features (BOLD correlations significantly correlated with ≥1 HAMD-17 items) and linear combinations of symptoms, resulted

in two components. The first component (labelled 'ANH') was defined by a pattern of predominantly frontostriatal and orbitofrontal contemporary correlations and was related to anhedonia and psychomotor retardation. The second (labelled 'ANX') was defined by a pattern of predominantly limbic correlations and was correlated with anxiety and insomnia. Next, hierarchical cluster analysis identified four subtypes with different combinations of high/ low component scores: +ANH/+ANX, +ANH/–ANX, –ANH/+ANX, –ANH/–ANX. These subtypes showed different levels of 5-week improvement (> 25% HAMD reduction) after treatment with repetitive high-frequency transcranial magnetic stimulation of the dorsomedial prefrontal cortex. The –ANH/+ANX subtype showed the highest response (33/40), followed by +ANH/–ANX (25/41). Lower responses were observed in –ANH/–ANX (4/16) and +ANH/+ANX (8/27).

One study used diffusion tensor imaging to cluster subjects based on fractional anisotropy (FA) scores of white matter, a measure of structural integrity.[74] The authors defined their ROIs *a posteriori* based on the differences between depressed subjects and paired healthy controls. The patterns of these differences were a combination of patterns from two distinct subtypes with a different age, which contributed different FA characteristics. The first subtype was aged 18–29 years and characterized by increased FA, especially in the corpus callosum, corticospinal midbrain, and inferior fronto-occipital fasciculus. FA near the midbrain was positively correlated with depression severity. The second subtype was aged 30–45 years and characterized by decreased FA, especially in the fronto-occipital fasciculus and posterior limb of internal capsule, which was negatively correlated with depression severity. The authors linked the subtype clustering to age of onset. However, the effects of age vs. age of onset could not be distinguished well because there were many first-episode subjects.

In summary, the results described above indicate that it might be possible to identify subgroups of patients based on neuroimaging data related to functional connectivity, as well as structural patterns. However, methodological differences between studies were considerable, making it hard to draw overall conclusions. In the area of structural imaging, larger studies incorporating relevant covariates (e.g., number of previous episodes, medication use) are needed.

Subtyping based on genetics

A single genetic subtyping study was conducted in a mixed cohort of depressed (n = 203) and healthy (n = 196) unrelated Mexican-American subjects.[75] Analyses were run with 83,898 SNPs and next with a limited set of 19 SNPs that were associated with MD (at $p < 0.05$). Hierarchical clustering with genetic distance measures using all SNPs yielded

no subtypes, but two distinct subtypes were found using only the 19 MD-related SNPs. The smallest of these subtypes (n = 41) included only subjects with MD but did not differ from the rest in terms of age, gender distribution, and overall severity. This study does not allow for strong conclusions about the existence of genetic MD subtypes, especially given its small sample size and the lack of a replication sample. In addition, preferably future studies also assess the association with additional clinically relevant correlates (e.g., course of illness, antidepressant use).

Subtyping based on a combination of psychometric and biological data

One study in 67 MD patients and 67 controls combined resting-state functional connectivity measures with clinical questionnaire scores and various biomarkers (plasma BDNF and cortisol, SNPs and DNA methylation for BDNF and serotonin genes) in a high-dimensional co-clustering model.[76] Co-clustering starts by clustering features, and then clusters subjects based on different sets of features.[77] The model that best separated between depressed subjects and healthy controls described five clusters, three of which contained depressed subjects. These three clusters were characterized by different mean levels of functional connectivity (between the angular gyrus and the DMN) and different levels of childhood abuse trauma, as well as treatment response. This study is unique in that it combined multiple types of data in the clustering process, but the sample was rather small, and included healthy controls, which might hamper the identification of subtypes of MD.

Discussion

This systematic review identified 29 publications presenting results on data-driven biological depression subtypes.

The biochemical studies indicated that some depressive subjects have higher levels of catecholamine levels and lower monoamine levels and some have levels that are comparable to healthy controls. These findings are interesting as they align with the monoamine hypothesis of depression, but also show that this theory is unlikely to apply to all depressed patients. Considering the importance of this hypothesis in modern psychiatry, it is remarkable to see that the reviewed biochemical studies - which all predate the introduction of selective serotonin reuptake inhibitors - already showed this, but no follow-up research has been performed.[78,79] These findings need to be replicated in larger samples of medication-free patients with sufficient post-hoc validation.

The symptom-based studies all found subtypes that, to some extent, differed on biological measures. Some studies identified a severe subtype that showed some overlap with the DSM melancholic specifier and had lower L-TRP scores and more dysregulation of the stress and metabolic systems than the non-severe subtype.[51–54,80] Other studies that incorporated more symptoms (i.e., both appetite/weight loss and gain) identified two severe subtypes, characterized either by appetite/weight loss or appetite/weight gain and that differed mainly on biomarkers that are related to weight, metabolism, inflammation, and stress.[17,56,58,59,62,81,82] The most recent study found subgroups mainly characterized by different levels of severity of depressive and anxiety symptoms, with potential differences in non-lagged functional connectivity. However, the clinical relevance of these differences remained unclear and replication is needed.

The neuroimaging studies found a range of different subtypes. The single structural study found two subtypes with different FA patterns, which the authors attributed to an age of onset difference, although other explanations cannot be ruled out.[83] The connectivity studies did provide rather inconsistent results with one study finding four subtypes with specific connectivity patterns related to anxiety and anhedonia[73] and another study finding two weakly separated subtypes[69]. These divergent findings are likely to have resulted from methodological differences. Fortunately, MD subtyping by neuroimaging is a rather new and active field, which holds promise for improvements in future research.[84]

The single genetic study identified two subtypes, one containing only patients and another containing patients and controls.[75] Although interesting, the presented results were limited because of a relatively small sample, few tests of clinical or other differences between the subtypes, and a lack of replication in independent samples. Genetic subtyping is a promising research direction, but to perform person-centered clustering based on large numbers of SNPs sample sizes should be large, and results should be replicated in independent test samples.

Limitations

This review has a number of limitations. First, relatively few studies have been performed, even though we considered a broad range of methods. Second, selective publication may have influenced the presented results (i.e., null results may have been less likely to be published). Third, although our inclusion criteria were quite broad, the specific focus on biological subtyping led to exclusion of some studies with clear relevance to the broader topic of depression heterogeneity, such as purely symptom-based clustering studies (see van Loo et al. (2012) and Marquand et al. (2016) for reviews).

Across the reviewed studies, both the quality and used methodology was highly variable. Studies differed in analysis techniques, symptom data, sample composition, and sample size. Most biochemical studies used similar techniques, but generally had small sample sizes, increasing the potential influence of random error on the results.[41-48] Of the psychometric studies, the older studies used nonparametric clustering methods and generally had smaller sample sizes[51–54,80] compared to the newer studies, which used parametric mixture methods[17,56,58,59,62,81,82]. Although larger, one of the newer studies combined symptom reports from current and lifetime patients, which is suboptimal.[56] The selected symptoms in psychometric studies may also have influenced the results. The older studies did not include weight gain, precluding identification of an atypical subtype. The newer studies did include both variants of weight change, but this might also be problematic because, based on its assumption of local independence, LCA will classify subjects with opposed symptom variants into different classes by default. This has led some to question whether the classes should actually be seen as appetite/weight-change classes, rather than clinically meaningful subtypes of depression.[85] The connectivity studies' methods also varied considerably in measurement setting (e.g., rest vs. task), ROIs, analytical steps (e.g., data reduction prior to clustering or not) and the kind of calculated BOLD correlations (e.g., lagged vs. non-lagged). The differing results across studies indicate that these variations may have strongly influenced findings. The influence of methodological choices on results also became evident in the genetic study; a subtype was only identified when using a heavily restricted set of selected SNPs. Finally, the study that combined multiple types of data included many controls, as did some other studies.[64,75,76] This might lead to a focus on the difference between patients and controls, rather than subtypes of MD, especially when separation between these groups is used as a selection criterion.

Another issue is the inconsistent use of terminology, particularly in the psychometric studies. Many of these studies use labels like 'atypical' and 'melancholic' to describe their clusters, even though the clusters do not match the DSM criteria of these specifiers, which might lead to misinterpretations.[86]

Finally, there are some general issues with cluster validation. Importantly, none of the studies performed out-of-sample validation. Instead, if any validation was performed, the same data were used for both model development and validation (i.e., double dipping[87]). One study used out-of-sample data, but compared the subtypes with healthy controls, rather than subtypes with each other.[73] Another major issue is the lack of validation of subtyping models against the null hypothesis that clusters are not present in the data, which was only done in a single study.[69] Indeed, only few techniques are capable of this

(e.g., SigClust, which tests if data can be modelled as coming from a single multivariate Gaussian distribution[88]), but some post-hoc tests exist.[89,90] Finally, evaluating the value of the model vs. the null hypothesis was sometimes hampered because cluster separation was artificially enhanced, for example, by the removal of edge cases or intermediate clusters.[59,61,73]

Implications

More homogeneous diagnostic (sub)categories could help the development of more tailored diagnostics and treatments.[15,18–22] However, the process of establishing such categories is complicated. Amendments to the DSM-5 are required to be backed up by evidence of improvements in validity, reliability, and clinical utility.[86] Although the reviewed studies provide interesting insights into depression heterogeneity, the resulting clusters should not be seen as equivalent or directly translatable to DSM-style subtype classifications for several reasons. Most importantly, cluster analyses are not meant to identify broadly applicable and clinically useful categories, but to identify structures in datasets. As such, cluster models will almost always provide a solution (especially in large samples), without any guarantee that the results are meaningful.[31] Also, the results of cluster analyses can vary considerably, depending on the used algorithms, sample populations, and (pre-processing of) input variables. Thus, results from a single cluster analysis should preferably not be interpreted as new clinical depression subtypes. Rather, if sufficiently replicated and robust, cluster analysis results could be useful contributions to the total evidence base for new clinical subtypes.

Future directions

The most recent reviewed studies are from the fields of neuroscience and genetics, and it seems likely that subtyping research in these fields will continue. In contrast, researchers seem to have largely lost interest in biochemical subtyping since 1990, leaving much to be investigated in this interesting area. Future studies could also use hybrid clustering models that can incorporate both dimensional and discrete variations among patients, or try to combine different types of data.[76,91,92] From a methodological perspective, it would be interesting to learn more about the quantitative effects of used clustering methods, sample, and data choices. Finally, general aspects that could contribute to more consistent and replicable results are the use of sufficient sample sizes, replication samples, using appropriate statistical methods, and investigation of clinically relevant correlates. Preferably, studies should use standardized sets of outcomes so results can be compared across studies.[93]

Conclusion

Although the reviewed literature provides ample starting points for future research, the methodological differences across studies and lack of replication preclude definitive conclusions about the existence of clinically useful and generalizable biological subtypes.

References

1. Alonso, J. *et al.* Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr. Scand. Suppl.* **109**, 38–46 (2004).
2. Sobocki, P., Jönsson, B., Angst, J. & Rehnberg, C. Cost of depression in Europe. *J. Ment. Health Policy Econ.* **9**, 87–98 (2006).
3. Whiteford, H. A. *et al.* Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010. *Lancet* **382**, 1575–1586 (2013).
4. Greenberg, P. E., Fournier, A.-A., Sisitsky, T., Pike, C. T. & Kessler, R. C. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J. Clin. Psychiatry* **76**, 155–62 (2015).
5. Chisholm, D. *et al.* Scaling-up treatment of depression and anxiety: A global return on investment analysis. *Lancet Psychiatry* **3**, 415–424 (2016).
6. Kessler, R. C., Broment, E. J., de Jonge, P., Shahly, V. & Wilcox, M. The Burden of Depressive Illness. in *Public Health Perspectives on Depressive Disorders* (ed. Cohen, N. L.) 40 (Johns Hopkins University Press, 2017).
7. Rush, A. J. The varied clinical presentations of major depressive disorder. *J. Clin. Psychiatry* **68**, 4–10 (2007).
8. Nandi, A., Beard, J. R. & Galea, S. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. *BMC Psychiatry* **9**, 31 (2009).
9. Goldberg, D. The heterogeneity of 'major depression'. *World Psychiatry* **10**, 226–228 (2011).
10. Wardenaar, K. J. & de Jonge, P. Diagnostic heterogeneity in psychiatry: towards an empirical solution. *BMC Med.* **11**, 1 (2013).
11. Fried, E. I. & Nesse, R. M. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR???D study. *J. Affect. Disord.* **172**, 96–102 (2015).
12. Monroe, S. M. & Anderson, S. F. Depression: The Shroud of Heterogeneity. *Curr. Dir. Psychol. Sci.* **24**, 227–231 (2015).
13. Krishnan, V. & Nestler, E. J. The molecular neurobiology of depression. *Nature* **455**, 894–902 (2008).
14. Hasler, G. Pathophysiology of depression: do we have any solid evidence of interest to clinicians? *World Psychiatry* **9**, 155–61 (2010).
15. Simon, G. E. & Perlis, R. H. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry* **167**, 1445–1455 (2010).
16. Ozomaro, U., Wahlestedt, C. & Nemeroff, C. B. Personalized medicine in psychiatry: problems and promises. *BMC Med.* **11**, 132 (2013).
17. Milaneschi, Y. *et al.* Polygenic dissection of major depression clinical heterogeneity. *Mol. Psychiatry* **21**, 516–22 (2016).
18. Korte, S. M. *et al.* The many different faces of major depression: It is time for personalized medicine. *Eur. J. Pharmacol.* **753**, 88–104 (2015).

19. Bartova, L., Berger, A. & Pezawas, L. Is there a personalized medicine for mood disorders? *Eur. Arch. Psychiatry Clin. Neurosci.* **260**, 121–126 (2010).

20. Nierenberg, A. A. Advancing the treatment of depression with personalized medicine. *J. Clin. Psychiatry* **73**, e17 (2012).

21. Miller, D. B. & O'Callaghan, J. P. Personalized medicine in major depressive disorder - Opportunities and pitfalls. *Metabolism* **62**, S34–S39 (2013).

22. Trivedi, M. H. Right patient, right treatment, right time: biosignatures and precision medicine in depression. *World Psychiatry* **15**, 237–238 (2016).

23. Kay, D. W., Garside, R. F., Beamish, P. & Roy, J. R. Endogenous and neurotic syndromes of depression: a factor analytic study of 104 cases. Clinical features. *Br. J. Psychiatry* **115**, 377–388 (1969).

24. Pilowsky, I., Levine, S. & Boulton, D. M. The classification of depression by numerical taxonomy. *Br. J. Psychiatry* **115**, 937–945 (1969).

25. Paykel, E. S. Classification of depressed patients: a cluster analysis derived grouping. *Br. J. Psychiatry* **118**, 275–288 (1971).

26. Andreasen, N. C. & Grove, W. M. The classification of depression: Traditional versus mathematical approaches. *Am. J. Psychiatry* **139**, 45–52 (1982).

27. Merikangas, K. R., Wicki, W. & Angst, J. Heterogeneity of depression. Classification of depressive subtypes by longitudinal course. *Br. J. Psychiatry* **164**, 342–348 (1994).

28. Sullivan, P. F., Kessler, R. C. & Kendler, K. S. Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. *Am. J. Psychiatry* **155**, 1398–1406 (1998).

29. Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G. & Mechelli, A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* **36**, 1140–1152 (2012).

30. Van Loo, H. M., De Jonge, P., Romeijn, J.-W., Kessler, R. C. & Schoevers, R. A. Data-driven subtypes of major depressive disorder: a systematic review. *BMC Med.* **10**, 156 (2012).

31. Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J. & Beckmann, C. F. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 433–447 (2016).

32. Ahmad, A. & Fröhlich, H. Integrating Heterogeneous omics Data via Statistical Inference and Learning Techniques. *Genomics Comput. Biol.* **2**, e32 (2016).

33. Lin, E. & Hsien-Yuan, L. Machine learning and systems genomicsapproaches for multi-omics data. *Biomark. Res.* **5**, (2017).

34. Baumeister, H. & Parker, G. Meta-review of depressive subtyping models. *J. Affect. Disord.* **139**, 126–140 (2012).

35. Kendell, R. & Jablensky, A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am. J. Psychiatry* **160**, 4–12 (2003).

36. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).

37. Haro, J. M. *et al.* ROAMER: roadmap for mental health research in Europe. *Int. J. Methods Psychiatr. Res.* **23**, 1–14 (2014).

38. Insel, T. R. & Cuthbert, B. N. Brain disorders? Precisely. *Science* **348**, 499–500 (2015).

39. The Prisma Group from Moher D, Liberati A, Tetzlaff J, A. D. Preferred reporting items for systematic reviews and meta analyses: The Prisma Statement. *PLoS Med* **6**, 1–15 (2009).

40. Moher, D. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* **4**, 1 (2015).

41. Asberg, M., Bertilsson, L., Tuck, D., Cronholm, B. & Sjöqvist, F. Indoleamine metabolites in the cerebrospinal fluid of depressed patients before and during treatment with nortriptyline. *Clin. Pharmacol. Ther.* **14**, 277–86 (1973).

42. Asberg, M., Thoren, P., Traskman, L., Bertilsson, L. & Ringberger, V. 'Serotonin depression'--a biochemical subgroup within the affective disorders? *Science* **191**, 478–480 (1976).

43. Gibbons, R. D. & Davis, J. M. Consistent evidence for a biological subtype of depression characterized by low CSF monoamine levels. *Acta Psychiatr. Scand.* **74**, 8–12 (1986).

44. Westenberg, H. G. M. & Verhoeven, W. M. A. CSF monoamine metabolites in patients and controls: support for a bimodal distribution in major affective disorders. *Acta Psychiatr. Scand.* **78**, 541–549 (1988).

45. Maas, J. W. *et al.* Pre-treatment neurotransmitter metabolites and response to imipramine or amitriptyline treatment. *Psychol Med* **12**, 37–43 (1982).

46. Åsberg, M. *et al.* CSF monoamine metabolites in melancholia. *Acta Psychiatr. Scand.* **69**, 201–219 (1984).

47. Azorin, J. M. *et al.* Plasma levels of 3-methoxy-4-hydroxyphenylglycol in depressed patients compared with normal controls. *Neuropsychobiology* **23**, 18–24 (1990).

48. Davis, J. M. *et al.* Cerebrospinal fluid and urinary biogenic amines in depressed patients and healthy controls. *Arch. Gen. Psychiatry* **45**, 705–717 (1988).

49. Haroon, E. *et al.* Increased inflammation and brain glutamate define a subtype of depression with decreased regional homogeneity, impaired network integrity, and anhedonia. *Transl. Psychiatry* **8**, 189 (2018).

50. Protzner, A. B., Kovacevic, N., Cohn, M. & McAndrews, M. P. Characterizing Functional Integrity: Intraindividual Brain Signal Variability Predicts Memory Performance in Patients with Medial Temporal Lobe Epilepsy. *J. Neurosci.* **33**, 9855–9865 (2013).

51. Maes, M., Cosyns, P., Maes, L., D'Hondt, P. & Schotte, C. Clinical subtypes of unipolar depression: Part I. A validation of the vital and nonvital clusters. *Psychiatry Res.* **34**, 29–41 (1990).

52. Maes, M., Maes, L., Schotte, C. & Cosyns, P. A clinical and biological validation of the DSM-III melancholia diagnosis in men: Results of pattern recognition methods. *J. Psychiatr. Res.* **26**, 183–196 (1992).

53. Schotte, C. K. W., Maes, M., Cluydts, R. & Cosyns, P. Cluster analytic validation of the DSM melancholic depression. The threshold model: Integration of quantitative and qualitative distinctions between unipolar depressive subtypes. *Psychiatry Res.* **71**, 181–195 (1997).

54. Orsel, S., Karadag, H., Turkcapar, H. & Karaoglan Kahilogullari, A. Diagnosis and classification subtyping of depressive disorders: Comparison of three methods. *Klin. Psikofarmakol. Bul.* **20**, 57–65 (2010).

55. Lamers, F. *et al.* Identifying depressive subtypes in a large cohort study: Results from the Netherlands Study of Depression and Anxiety (NESDA). *J. Clin. Psychiatry* **71**, 1582–1589 (2010).

56. Milaneschi, Y. *et al.* The effect of FTO rs9939609 on major depression differs across MDD subtypes. *Molecular psychiatry* **19**, 960–962 (2014).

57. Lamers, F. *et al.* Stability and transitions of depressive subtypes over a 2-year follow-up. *Psychol. Med.* **42**, 2083–2093 (2012).

58. Lamers, F. *et al.* Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. *Mol. Psychiatry* **18**, 692–699 (2013).

59. Milaneschi, Y., Lamers, F., Bot, M., Drent, M. L. & Penninx, B. W. J. H. Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol. Psychiatry* **81**, 807–814 (2017).

60. Frayling, T. M. *et al.* A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science* **316**, 889–894 (2007).

61. Lamers, F. *et al.* Serum proteomic profiles of depressive subtypes. *Transl. Psychiatry* **6**, e851 (2016).

62. Bus, B. A. A. *et al.* Low serum BDNF levels in depressed patients cannot be attributed to individual depressive symptoms or symptom cluster. *World J. Biol. Psychiatry* **15**, 561–569 (2014).

63. Bouveyron, C., Girard, S. & Schmid, C. High-dimensional data clustering. *Comput. Stat. Data Anal.* **52**, 502–519 (2007).

64. Maglanoc, L. A. *et al.* Data-Driven Clustering Reveals a Link Between Symptoms and Functional Brain Connectivity in Depression. *Biol. psychiatry. Cogn. Neurosci. neuroimaging* **4**, 16–26 (2019).

65. Feighner, J. P., Sverdlov, L., Nicolau, G. & Noble, J. F. Cluster analysis of clinical data to identify subtypes within a study population following treatment with a new pentapeptide antidepressant. *Int. J. Neuropsychopharmacol.* **3**, 237–242 (2000).

66. Ballard, E. D. *et al.* Characterizing the course of suicidal ideation response to ketamine. *J. Affect. Disord.* **241**, 86–93 (2018).

67. Greicius, M. Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.* **24**, 424–430 (2008).

68. Sundermann, B., Olde lütke Beverborg, M. & Pfleiderer, B. Toward literature-based feature selection for diagnostic classification: a meta-analysis of resting-state fMRI in depression. *Front. Hum. Neurosci.* **8**, (2014).

69. Feder, S. *et al.* Sample heterogeneity in unipolar depression as assessed by functional connectivity analyses is dominated by general disease effects. *J. Affect. Disord.* **222**, 79–87 (2017).

70. Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D. & Pizzagalli, D. A. Large-Scale Network Dysfunction in Major Depressive Disorder A Meta-analysis of Resting-State Functional Connectivity. *JAMA Psychiatry* **72**, 603–611 (2015).

71. Price, R. B. *et al.* Parsing Heterogeneity in the Brain Connectivity of Depressed and Healthy Adults During Positive Mood. *Biol. Psychiatry* **81**, 347–357 (2017).

72. Price, R. B., Gates, K., Kraynak, T. E., Thase, M. E. & Siegle, G. J. Data-driven subgroups in depression derived from directed functional connectivity paths at rest. *Neuropsychopharmacology* **42**, 2623–2632 (2017).

73. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2017).

74. Cheng, Y. *et al.* Delineation of early and later adult onset depression by diffusion tensor imaging. *PLoS One* **9**, (2014).

75. Yu, C., Arcos-Burgos, M., Licinio, J. & Wong, M. L. A latent genetic subtype of major depression identified by whole-exome genotyping data in a Mexican-American cohort. *Transl. Psychiatry* **7**, (2017).

76. Tokuda, T. *et al.* Identification of depression subtypes and relevant brain regions using a data-driven approach. *Sci. Rep.* **8**, 14082 (2018).

77. Tokuda, T. *et al.* Multiple co-clustering based on nonparametric mixture models with heterogeneous marginal distributions. *PLoS One* **12**, e0186566 (2017).

78. Albert, P. R., Benkelfat, C. & Descarries, L. The neurobiology of depression-revisiting the serotonin hypothesis. I. cellular and molecular mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2378–2381 (2012).

79. Lacasse, J. R. & Leo, J. Serotonin and depression: A disconnect between the advertisements and the scientific literature. *PLoS Medicine* **2**, 1211–1216 (2005).

80. Maes, M. *et al.* Clinical subtypes of unipolar depression: Part III. Quantitative differences in various biological markers between the cluster-analytically generated nonvital and vital depression classes. *Psychiatry Res.* **34**, 59–75 (1990).

81. Rodgers, S. *et al.* Serum testosterone levels and symptom-based depression subtypes in men. *Front. Psychiatry* **6**, (2015).

82. Bot, M. *et al.* Serum proteomic profiling of major depressive disorder. *Transl. Psychiatry* **5**, e599 (2015).

83. Zhu, X. *et al.* Altered white matter integrity in first-episode, treatment-naive young adults with major depressive disorder: A tract-based spatial statistics study. *Brain Res.* **1369**, 223–229 (2011).

84. Koenigs, M., Baskin-Sommers, A., Zeier, J. & Newman, J. P. Investigating the neural correlates of psychopathy: A critical review. *Mol. Psychiatry* **16**, 792–799 (2011).

85. van Loo, H. M., Wanders, R. B. K., Wardenaar, K. J. & Fried, E. I. Problems with latent class analysis to detect data-driven subtypes of depression. *Mol. Psych.* **23**, 495-496 (2016).

86. American Psychiatric Association. Submit Proposals for Making Changes to DSM–5. (2017) Accessed from: https://www.psychiatry.org/psychiatrists/practice/dsm/submit-proposals.

87. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. & Baker, C. I. Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci.* **12**, 535–540 (2009).

88. Liu, Y., Hayes, D. N., Nobel, A. & Marron, J. S. Statistical significance of clustering for high-dimension, low-sample size data. *J. Am. Stat. Assoc.* **103**, 1281–1293 (2008).

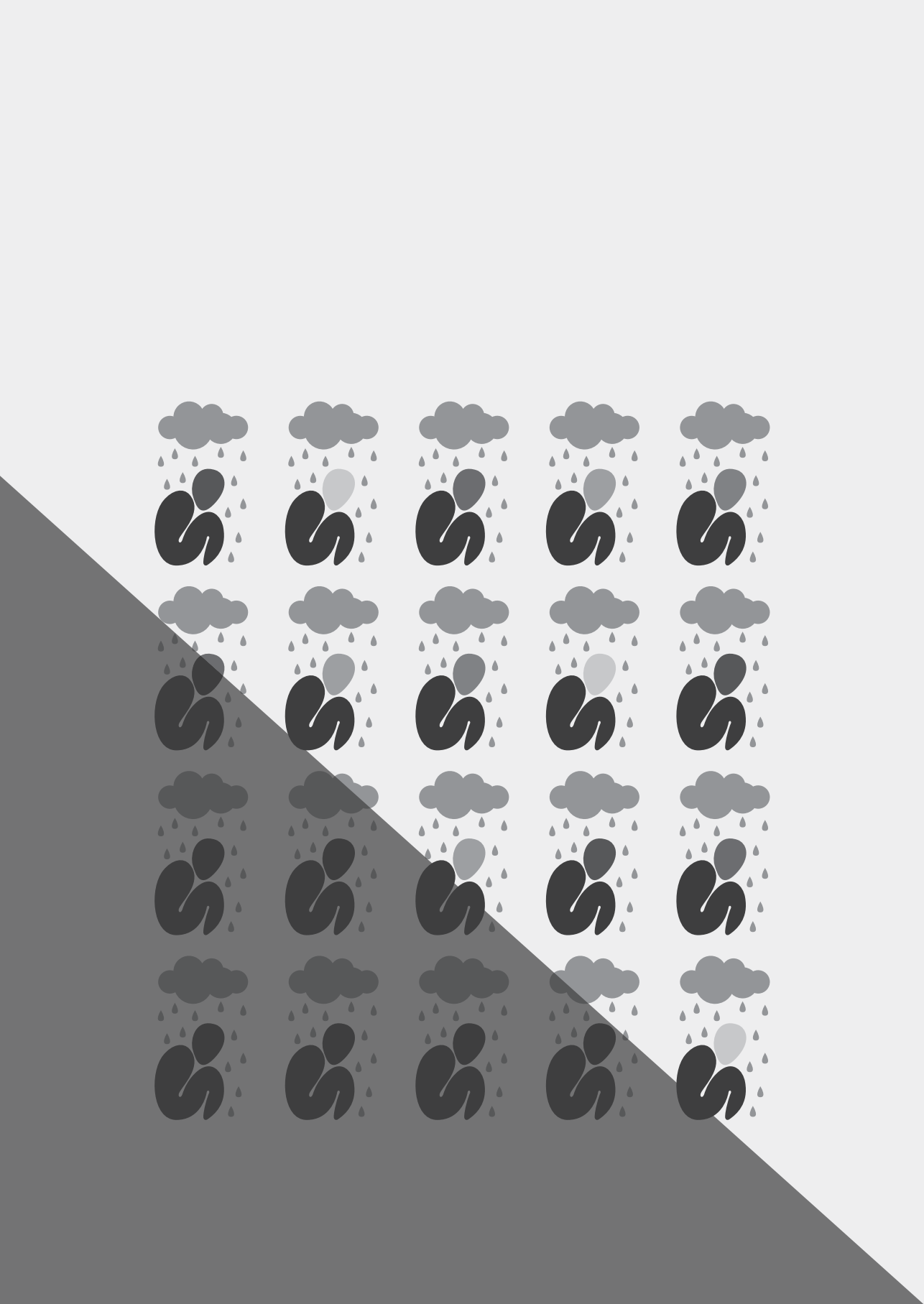
89. Tibshirani, R., Walther, G. & Hastie, T. Estimating the Number of Data Clusters via the Gap Statistic. *J. R. Stat. Soc.* **63**, 411–423 (2001).

90. Kou, J. Estimating the number of clusters via the GUD statistic. *J. Comput. Graph. Stat.* **23**, 403–417 (2014).

91. Lubke, G. H. & Muthén, B. Investigating Population Heterogeneity With Factor Mixture Models. *Psychol. Methods* **10**, 21–39 (2005).

92. Wardenaar, K. J., Wanders, R. B. K., ten Have, M., de Graaf, R. & de Jonge, P. Using a hybrid subtyping model to capture patterns and dimensionality of depressive and anxiety symptomatology in the general population. *J. Affect. Disord.* **215**, 125–134 (2017).

93. Kapur, S., Phillips, a G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).



Chapter 5

Biomarker-based subtyping of depression and anxiety disorders using Latent Class Analysis: a NESDA study

Lian Beijers, Klaas J. Wardenaar, Fokko J. Bosker, Femke Lamers, Gerard van Grootheest, Marrit K. de Boer, Brenda W.J.H. Penninx and Robert A. Schoevers

Psychological Medicine **49**, 617-627 (2018).

The supplementary material for this article can be found at https://doi.org/10.1017/S0033291718001307.

Abstract

Background

Etiological research of depression and anxiety disorders has been hampered by diagnostic heterogeneity. In order to address this, researchers have tried to identify more homogeneous patient subgroups. This work has predominantly focused on explaining interpersonal heterogeneity based on clinical features (i.e. symptom profiles). However, to explain interpersonal variations in underlying pathophysiological mechanisms, it might be more effective to take biological heterogeneity as the point of departure when trying to identify subgroups. Therefore, this study aimed to identify data-driven subgroups of patients based on biomarker profiles.

Methods

Data of patients with a current depressive and/or anxiety disorder came from the Netherlands Study of Depression and Anxiety, a large, multi-site naturalistic cohort study (n = 1460). Thirty-six biomarkers (e.g. leptin, brain-derived neurotrophic factor, tryptophan) were measured, as well as sociodemographic and clinical characteristics. Latent class analysis of the discretized (lower 10%, middle 80%, upper 10%) biomarkers were used to identify different patient clusters.

Results

The analyses resulted in three classes, which were primarily characterized by different levels of metabolic health: 'lean' (21.6%), 'average' (62.2%) and 'overweight' (16.2%). Inspection of the classes' clinical features showed the highest levels of psychopathology, severity, and medication use in the overweight class.

Conclusion

The identified classes were strongly tied to general (metabolic) health and did not reflect any natural cut-offs along the lines of the traditional diagnostic classifications. Our analyses suggested that especially poor metabolic health could be seen as a distal marker for depression and anxiety, suggesting a relationship between the 'overweight' subtype and internalizing psychopathology.

Introduction

Depression and anxiety are highly prevalent and associated with a substantial burden on both patients, their caregivers and society as a whole.[1–3] The overlap between depressive and anxiety symptoms and disorders is the rule rather than the exception, and much of the same treatments are currently used for patients in both diagnostics categories.[4–6] Despite the many research efforts that have been made over the past decades, the etiological mechanisms underlying depressive and anxiety disorders are still poorly understood, which is partially due to the heterogeneous nature of the patient population[7,8], resulting in small to moderate observed treatments effects.[9–12] Scientific progress is more likely if the problem of diagnostic heterogeneity is effectively addressed. Unfortunately, more homogeneous clinical subtypes have been shown to be weakly associated with etiology, the course of illness, and treatment response.[13] Alternative data-driven subtype classifications based on cluster analyses, have also been shown to have limited associations with specific etiological mechanisms.[14-17] This limited association with etiology might be explained by the fact that these subtypes are primarily optimized to differentiate between symptom patterns and not between biological mechanisms.[18] Subtypes might be more suitable to investigate etiological heterogeneity when biological profiles are used as the point of departure.

It has been acknowledged that theories assuming that there is a single biological disturbance underlying depression in all patients (e.g., monoamine deficiency) have limited validity[19–21]. Rather, the specific disturbances are likely to differ between patients, even those with similar symptomatology and/or diagnoses.[22] If we consider this, the low observed effect sizes in treatment trials could be explained by the fact that only in part of the patients treated with a certain compound (e.g., selective serotonin reuptake inhibitors), the treatment actually affects their individual biological disturbances.[23] Identification of more homogeneous biomarker-based subtypes in the patient population could help to gain more insight into patient-specific etiological mechanisms and to better target treatments to those that are likely to benefit.[24–33]

To our knowledge, the existence of depression/anxiety subtypes with different biological disturbance profiles and their manifest clinical characteristics have not been extensively studied before. Therefore, this study aimed to identify biomarker-based subtypes using latent class analysis (LCA) a large and well phenotyped sample of depressive and/or anxiety patients (n = 1460). In this sample, 36 biomarkers were measured, representing different underlying mechanisms that have previously been found to be relevant to depression and anxiety disorders (e.g. hypothalamus-pituitary-adrenal

axis function, inflammation). Next, in order to investigate the clinical relevance and utility of the identified biomarker-based subtypes, clinical characteristics and symptomatology were compared across the identified classes.

Methods and materials

Participants and procedures

The Netherlands Study of Depression and Anxiety (NESDA) is a multisite naturalistic cohort study that aims to examine the long-term course of depressive and anxiety disorders. A detailed description of the NESDA study design and sampling procedures can be found elsewhere.[34] In brief, the NESDA cohort consists of 2981 subjects, aged 18—65 years, with (a history of) anxiety and/or depressive disorder and healthy controls. The research protocol was approved by the Medical Ethical Committees of participating institutes, and after complete description of the study, all respondents provided written informed consent. For the present study, all 1460 subjects with a current (last month) diagnosis of a depressive or anxiety disorder according to the Composite International Diagnostic Interview (CIDI; WHO version 2.1) were selected. For additional analyses (see below), 634 healthy control subjects were added to the dataset.

Measurements

During a 4-hour baseline assessment including interviews, a medical examination, a cognitive computer task and collection of blood samples, extensive information was gathered about key (mental) health outcomes and demographic, psychosocial, clinical and biological determinants. Additional measures (written questionnaires and saliva samples) were carried out at home by the participants. Diagnostic and statistical manual of mental disorders-fourth edition (DSM-IV) diagnoses of depressive (minor depression, dysthymia and major depression (MD)) and anxiety disorders (generalized anxiety disorder, social phobia, agoraphobia and panic disorder) were established using the Composite International Diagnostic Interview (CIDI) 2.1. Depression symptom severity was measured with the 30-item Inventory of Depressive Symptomatology Self Report (IDS-SR).[35,36] Anxiety symptom severity was measured with the Beck Anxiety Inventory (BAI).[37] Sociodemographic information included age, sex, number of chronic diseases, drinking and smoking behavior. The biological data consisted of waist circumference, body mass index (BMI), blood pressure (systolic/diastolic and the ankle-brachial index), heart

rate variability and markers from blood and saliva samples. A summary of the biological methods is provided below. For more details, see Supplementary Table S1.

Blood markers

Venous blood was drawn prior to the interview session (between 8:00 and 9:00 am) after an overnight fast. Venous blood samples were transferred to a local lab for routine assessments; serum and plasma were spun down within an hour and stored at −80 ˚C for later analyses. The routine assays included hematological variables (hemoglobin, hematocrit, and erythrocyte count), liver function (γ-glutamyltransferase, aspartate aminotransferase, alanine aminotransferase) and kidney function (creatinine), as well as markers related to the metabolic syndrome (MetS) and obesity (glucose, cholesterol, triglycerides, HDL- and LDL-cholesterol, thyroid-stimulating hormone, free thyroxine). Additional biomarkers are described below. Most of these markers' associations with mental-health outcomes have previously been established.[38-40]

Inflammation

C-reactive protein (hsCRP) was measured in duplicate by an in-house enzyme-linked immunosorbent assay (ELISA), based on purified protein and polyclonal anti-hsCRP antibodies (Dako, Glostrup, Denmark). Plasma IL-6 levels were measured in duplicate by a high sensitivity ELISA (PeliKine Compact TM ELISA, Sanquin, Amsterdam, The Netherlands). Plasma tumor necrosis factor alpha (TNF-α) levels were assayed in duplicate using a high-sensitivity solid phase ELISA (Quantikine® HS Human TNF-α Immunoassay, R&D systems Inc, Minneapolis, MN, USA). Tryptophan and (OH-) kynurenine concentrations were assayed by an automated online solid-phase extraction-liquid chromatographic-tandem mass spectrometric (XLC-MS/MS) method. Both inflammation and degradation of tryptophan along the kynurenine pathway have been found to be associated with depression in this sample.[41,42]

Neuroplasticity

Brain-derived neurotrophic factor (BDNF) protein levels were measured in serum samples using the Emax Immuno Assay system from Promega (Madison, WI, USA). IGF-I (nmol/l) was assayed centrally by chemiluminescence immunoassay of EDTA plasma on the Liaison autoanalyzer (DiaSorin, S.p.A., Italy). Both measures have previously been found to be associated with depression with melancholic features in this sample.[43]

Mineral balance

Measurements of parathyroid hormone (PTH) were performed at the Endocrine Laboratory of the VU University Medical Center. PTH levels were measured in EDTA plasma using an intact PTH assay. Intact PTH levels were measured using an immunometric assay (Architect, Abbott Diagnostics, Abbott Park, IL). Vitamin D was measured based on circulating levels of 25(OH)D, extracted and analysed by XLC-MS/MSa (Spark Holland, Emmen, the Netherlands) and coupled to a Quattro Premier XE tandem mass spectrometer (Waters Corp., Milford, MA, USA). Vitamin D has previously been found to be associated with depression in this sample.[44]

Steroid hormones

Dehydroepiandrosterone and its sulphate conjugate (DHEA/-S) were determined using a delayed one-step immunoassay with a chemiflex assay protocol. Sex Hormone Binding Globulin (SHBG) was determined using a two-step immunoassay with a chemiflex assay protocol. Total estradiol (E2) was determined using a delayed one-step immunoassay with a chemiflex assay protocol. Previously, steroid hormones were found to be associated with anxiety and depression in this sample.[45,46]

Leptin

Plasma leptin concentrations were measured in EDTA plasma using an ELISA kit (Human Leptin ELISA Kit; Linco Research, Inc, St. Charles, Missouri). Leptin has previously been found to be associated with depression (with atypical features) in this sample.[47]

Saliva markers

On a single day, prior to the first face-to-face assessment session, participants themselves collected saliva samples at home using Salivettes (Sarstedt AG and Co, Nürmbrecht, Germany). Measurements were taken at awakening (T_1), 30 (T_2), 45 (T_3) and 60 (T_4) minutes later, and in the evening (22:00 (T_5) and 23:00 (T_6)). Additionally, the dexamethasone suppression test (Carroll, 1982) was carried out by oral administration of a 0.5 mg dexamethasone pill directly after T_6 and a final cortisol measurement the next morning at awakening (T_7). The saliva samples were used to assess levels of cortisol, amylase, and testosterone. These measures have previously been observed to be associated with depression and anxiety in this sample.[45,48,49]

Heart rate variability

A heart rate recording was performed with the Vrije Universiteit Ambulatory Monitoring System (VU-AMS).[50] Subjects were wearing the VU-AMS device during a large part of the NESDA baseline assessment, while participating in different assessment parts (i.e., medical examination, interviewing, and a computer task). The start of the various assessment parts was marked with an event marker to divide the total recording into fixed periods. Movement registration through vertical accelerometry was used to excise periods where subjects were non-stationary. For this project, VU-AMS heart rate variability (HRV) during resting baseline, and HRV change from baseline to two stress conditions (interview and stressful computer task) were used. HRV has previously been found to be associated with depression in this sample.[51]

Statistical analyses

LCA was used to identify data-driven subgroups with distinct biological profiles. LCA assumes that an unobserved, latent categorical variable explains the association among a set of observed variables. The input variables are listed in Supplementary Table S1. Models with increasing numbers of classes were estimated and compared. The optimal model selection was based on the highest entropy and the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Relative entropy is a measure of classification accuracy (range: 0–1), with values closer to 1 indicating greater accuracy. Lower AIC and BIC values indicate that the model provides a better description of the data. Because these measures do not always agree, parsimony and interpretability were also taken into account. To avoid convergence on a solution at a local maximum LCA was run using up to 1000 random starting values and 250 final stage optimizations. LCA was conducted using Mplus version 5.[52,53] To investigate the influence of the different variables on the model solution Cramer's V was calculated for each biomarker variable.[54]

After the optimal model was identified, subjects were assigned to their most likely class based on their highest posterior class probability. Differences between classes (in biomarkers, sociodemographics, DSM-IV diagnoses and clinical characteristics) were investigated by using two-tailed χ^2 statistics for categorical variables and one-way analysis of variance statistics for continuous variables, or by using the Kruskall–Wallis test if the outcomes were not normally distributed. The False Discovery Rate controlling procedure was used to counteract the problem of multiple comparisons.[55] All comparisons were conducted using SPSS for Windows, version 23 (IBM Inc., USA). To evaluate if the identified class structure was specifically informative about the biological heterogeneity in depressed/anxious patients or was more broadly reflective of normal biological variations

in the population as a whole, all analyses were rerun in a dataset including both patients and healthy controls (n = 2094).

Data pre-processing

Clinically determined cut-off values to categorize the biomarker variables were available for some but not all biomarkers. Therefore, an alternative approach was taken, categorizing every variable based on percentiles, with the lowest and highest scoring 10th percentile of all subjects being coded as −1 and 1, respectively, and the middle 80% being coded 0. The 10th percentile cut-off was chosen to make sure that especially the more extreme, potentially clinically relevant variations in biomarker levels would be represented in the eventual LCA model. Setting the cut-offs at higher percentile values would lead to biomarker variables with more subjects in both the lower and upper category, but with a more within-category variation of biomarker levels. This makes the categorization potentially less useful for differentiation between subjects with distinct biomarker patterns as relevant interpersonal differences are eliminated by pooling patients with different biomarker levels in the same category. Indeed, exploratory analyses using 25th and 50th percentiles as cut-offs led models with many classes (i.e., the AIC and BIC kept decreasing with each class addition) that could not be easily distinguished from each other in terms of their specific biomarker patterns. In the final coding scheme, the number of subjects per decile varied between 8% and 12%, because in some cases a large number of subjects had the same score, and we chose to use the percentile closest to 10. The above-described coding was done separately for different sex/age (< 30, 30-50, > 50) strata, because the distributions of some biological variables are known to differ across sex and age (e.g., testosterone). Differences between classes on other potentially relevant covariates were investigated after identification of the optimal latent class model.

To ensure that the model solution would not be driven by the fact that some variables were essentially measuring the same thing, all variables with a correlation above 0.75 before recoding were summed after categorization (i.e., the hematological markers, the heart rate variability reactivity in both test conditions, systolic and diastolic blood pressure, aspartate and alanine aminotransferase). Subjects with a score of $\leqslant -1$ got a value of −1, those with a score of 0 got a value of 0 and subjects with a score of $\geqslant 1$ got a value of 1 on a newly constructed compound variable. Recoding was done using SPSS for Windows, version 23 (IBM Inc., USA). The final dataset included 36 biological variables.

Results

The LCA results are shown in Table 1. The lowest BIC combined with adequate entropy indicated that the 3-class model best described the data. Although the AIC decreased for the more complex models up to six classes, these models were less optimal in terms of parsimony and showed only marginally higher entropy. For each class, Figure 1 shows the probabilities of biomarker scores in the top or bottom 10th percentile.

Table 1. Statistics for LCA models with different numbers of classes, based on the sample of subjects with current psychopathology (n = 1406)

Classes	DF	AIC	BIC	H
2	149	61 930.889	62 718.531	0.725
3	224	61 360.457	62 544.564[a]	0.765
4	299	61 079.887	62 660.458	0.756
5	374	60 934.908	62 911.944	0.765
6	449	60 807.111[a]	63 180.612	0.797[a]

[a]Most favorable score.

DF, Degrees of Freedom; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; H, Entropy

Table 2 shows the distribution of biomarkers across the three latent classes. Most biomarkers differed among classes, with the notable exception of the cortisol-related biomarkers. Waist circumference, BMI and leptin had the largest effect (Cramer's V > 0.5) on the model solution and were used to inform class labelling. Other variables (e.g., measures on inflammation and steroid hormones) showed smaller, but still meaningful, variation between classes (Cramer's V = 0.335–0.209, see Table 3).

The first class was labeled 'lean' (n = 311) because it was associated with a healthy BMI and a comparatively high probability of being in the bottom 10th percentile for the obesity/MetS markers. The second class was labeled 'average' (n = 910) because it showed a low probability for extreme scores on any of the biomarkers. The third class was labeled 'overweight' (n = 239) and was characterized by a pattern of probabilities almost perpendicular to the lean class, with comparatively high probabilities to be in the upper 10% on obesity and MetS-related markers.

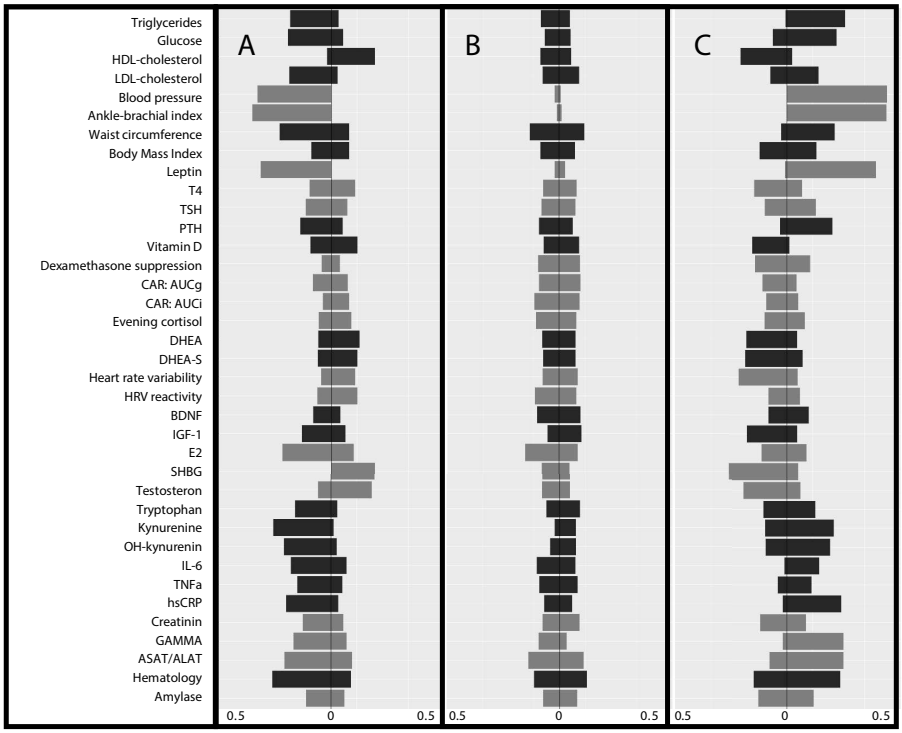


Figure 1. Probabilities to score in the lowest (left) or highest (right) 10-th percentile of each variable for the different latent classes in the sample including subjects with current psychopathology

A = Lean (21.6%), B = Average (62.2%), C = Overweight (16.2%). Color groups consist of biomarkers with similar themes, as indicated in Supplementary Table S1. Abbreviations: Blood pressure, combination of systolic and diastolic pressure; T4, free thyroxine; TSH, thyroid stimulating hormone; PTH, parathyroid hormone; CAR, cortisol awakening response; AUCg, area under the curve with respect to the ground, AUCi, area under the curve with respect to increase; DHEA(-S), dehydroepiandrosterone(-sulphate); HRV, heart rate variability; HRV reactivity, combination of HRV reactivity in both stress situations; BDNF, brain-derived neurotrophic factor, IGF-1 insulin-like growth factor 1; E2, estradiol; SHBG, sex hormone binding globulin; IL-6, interleukin 6; TNFa, tumor necrosis factor alpha; hsCRP, C-reactive protein; GAMMA, gamma-glutamyltransferase; ASAT/ALAT, combination of aspartate aminotransferase and, alanine aminotransferase; hematology, combination of hemoglobin, hematocrit, and erythrocyte values

Table 2. Distribution of biomarkers across identified classes in the sample of subjects with current psychopathology

Hypothesis	Variable (abbreviation, unit)	Lean (*n* = 311)	Average (*n* = 910)	Overweight (*n* = 239)	*P**
Neuroplasticity	Brain-derived neurotrophic factor (BDNF, ng/ml)	8.66 (3.13)	9.28 (3.66)	9.19 (3.67)	0.033
	Insulin-like Growth Factor 1 (IGF-1, nmol/l)	25.1 (8.7)	26.88 (8.42)	23.15 (7.87)	<0.001[a,b]
Mineral balance	Vitamin D (nmol/l)	62.84 (33.92)	63.14 (27.52)	47.02 (23.25)	<0.001[a,c]
	Parathyroid hormone (PTH, pmol/l)	5.25 (2.22)	5.60 (2.24)	7.16 (3.18)	<0.001[a,b,c]
Inflammation	C-reactive protein (hsCRP, mg/l)	1.58 (2.82)	2.68 (4.9)	6.65 (7.82)	<0.001[a,b,c]
	Interleukin 6 (IL-6, pg/ml)	1.05 (1.18)	2.00 (23.44)	1.81 (3.67)	<0.001[a,b,c]
	Tumor necrosis factor alpha (TNF-α, pg/ml)	0.87 (0.94)	1.13 (1.54)	1.25 (1.32)	<0.001[a,b,c]
	Tryptophan (µmol/l)	58.75 (11.54)	65.57 (13.17)	65.58 (15.51)	<0.001[a,c]
	Kynurenine (µmol/l)	1.84 (0.53)	2.32 (0.59)	2.48 (0.84)	<0.001[a,c]
	3-hydroxykynurenine (nmol/l)	24.18 (11.42)	32.06 (13.08)	37.81 (19.79)	<0.001[a,b,c]
Stress response	Area under the curve with respect to the ground (AUCg, µg/dl/h)	19.98 (7.18)	19.34 (7.35)	17.48 (6.41)	0.012[a,c]
	Area under the curve with respect to the increase (AUCi, µg/dl/h)	3.30 (5.92)	2.43 (6.68)	1.61 (5.38)	0.047
	Evening cortisol (nmol/l)	5.92 (3.76)	5.29 (3.23)	5.36 (2.63)	0.017[a]
	Dexamethasone (DST) cortisol-suppression	0.38 (0.17)	0.39 (0.24)	0.31 (0.25)	0.001[a,c]
Heart rate Variability	Mean root mean square of successive differences (RMSSD)	47.53 (39.11)	41.60 (31.63)	31.91 (28.69)	<0.001[a,b,c]
	Difference from baseline to computer task	8.93 (18.94)	4.80 (23.4)	0.65 (22.5)	<0.001[a,b,c]
	Difference from resting baseline to interview	6.65 (14.13)	3.59 (17.5)	0.40 (16.7)	<0.001[a,b,c]
Hematology	Hemoglobin (Hob, nmol/l)	8.50 (0.80)	8.72 (0.78)	8.76 (0.89)	<0.001[a,b,c]
	Hematocrit (HT, L/L)	0.40 (0.04)	0.41 (0.03)	0.42 (0.04)	<0.001[a,b,c]
	Erythrocytes (×10^12/l)	4.48 (0.42)	4.65 (0.4)	4.75 (0.43)	<0.001[a,b,c]
Steroid hormones	Dehydroepiandrosterone (DHEA, nmol/ml)	25.50 (16.21)	22.49 (16.15)	19.08 (16.04)	<0.001[a,b,c]
	Dehydroepiandrosterone sulphate (DHEA-S, µmol/l)	6.55 (3.61)	6.25 (3.65)	5.39 (3.66)	<0.001[a,b,c]
	Sex hormone-binding globulin (SHBG, nmol/l)	85.38 (66.9)	60.69 (51.65)	45.29 (39.04)	<0.001[a,b,c]
	Estradiol (E2, nmol/l)	0.49 (5.37)	0.26 (2.69)	0.17 (0.18)	0.088
	Testosterone (nmol/l)	10.52 (11.3)	8.54 (8.62)	7.24 (7.54)	0.001[a,b]

Hypothesis	Variable (abbreviation, unit)	Lean (*n* = 311)	Average (*n* = 910)	Overweight (*n* = 239)	*P**
Kidney function	Creatinine (mg/dl)	0.73 (0.14)	0.76 (0.15)	0.76 (0.26)	0.007[a]
Liver function	Gamma-glutamyltransferase (gamma-GT, U/L)	22.19 (27.56)	22.66 (17.67)	43.67 (39.75)	<0.001[a,b,c]
	Alanine aminotransaminase (ALAT, U/L)	19.29 (11.74)	22.91 (16.43)	34.95 (28.54)	<0.001[a,b,c]
	Aspartate aminotransferase (ASAT, U/L)	23.83 (9.86)	25.45 (10.61)	29.69 (13.73)	<0.001[a,b,c]
Metabolic markers	Waist circumference (cm)	77.85 (9.30)	88.88 (10.65)	108.43 (14.06)	<0.001[a,b,c]
	Body Mass Index	21.45 (2.77)	25.33 (3.55)	33.13 (5.81)	<0.001[a,b,c]
	Glucose (mmol/l)	5.01 (1.15)	5.16 (0.77)	5.74 (1.58)	<0.001[a,b,c]
	Low-density lipoproteins (LDL, mmol/l)	2.73 (0.91)	3.22 (0.98)	3.40 (0.99)	<0.001[a,b,c]
	High-density lipoproteins (HDL, mmol/l)	1.85 (0.50)	1.59 (0.39)	1.38 (0.40)	<0.001[a,b,c]
	Triglycerides (mmol/l)	0.97 (0.51)	1.29 (0.81)	1.96 (1.12)	<0.001[a,b,c]
	Blood pressure, systolic (Bp, mm Hg)	130.41 (17.5)	136.37 (20.01)	144.04 (17.58)	<0.001[a,b,c]
	Blood pressure, diastolic (Bp, mm Hg)	78.45 (9.74)	82.11 (11.29)	88.48 (10.62)	<0.001[a,b,c]
	Ankle brachial pressure index (ABPI)	1.14 (0.15)	1.14 (0.16)	1.15 (0.16)	0.345
	Leptin (µg/l)	7.26 (6.60)	14.94 (10.49)	31.06 (18.28)	<0.001[a,b,c]
	Thyroid-stimulating hormone (TSH, mU/l)	2.59 (3.57)	2.49 (1.51)	2.80 (3.23)	0.245
	Thyroxine (T4, pmol/l)	15.69 (2.94)	15.44 (2.44)	15.04 (2.78)	0.008[a,c]
Amylase	Amylase (IU/l)	289 496.10 (259 769.97)	294 556.09 (325 513.49)	319 290.70 (253 538.20)	0.290

*Based on ANOVA for continuous variables (post-hoc = Tukey) and Kruskal–Wallis tests for non-normally distributed variables (post-hoc = Dunn test).

Significant differences by class (at $\alpha < 0.05$ corrected using the False Discovery Rate controlling method):

[a]'lean' v. 'overweight' class.

[b]'average' v. 'overweight' class.

[c]'lean' v. 'average' class.

Table 3. Cramer's V values for each biological variable in the 3-class model based on the sample of subjects with current psychopathology

Variable	Cramer's *V*	Variable	Cramer's *V*
Waist circumference	0.606	**LDL-cholesterol**	0.148
Body mass index	0.585	**Insulin-like growth factor 1**	0.142
Leptin	0.518	**Interleukin 6**	0.139
Kynurenine	0.335	**Dehydroepiandrosterone**	0.134
C-reactive protein	0.282	**Dehydroepiandrosterone-sulphate**	0.128
Sex hormone binding globulin	0.275	**25-hydroxyvitamin D**	0.118
Gamma-glutamyltransferase	0.270	**Dexamethasone suppression test**	0.107
Triglycerides	0.264	**AUCi**	0.091
Ankle-brachial index	0.260	**Free thyroxine**	0.091
OH-kynurenine	0.258	**Tumor necrosis factor alpha**	0.089
Glucose	0.233	**Estradiol**	0.082
Testosterone	0.214	**Thyroid stimulating hormone**	0.075
HDL-cholesterol	0.209	**Amylase**	0.074
Parathyroid hormone	0.173	**Brain-derived neurotrophic factor**	0.070
Liver values (combined ASAT/ALAT)	0.168	**Heart rate variability reactivity**	0.066
Blood pressure (combined systolic/diastolic)	0.164	**Evening cortisol**	0.053
Tryptophan	0.156	**Creatinine**	0.053
Heart rate reactivity (combined from both stress tests)	0.154	**AUCg**	0.046
Blood values (combined HB, HT, erythrocytes)	0.151		

ASAT, aspartate aminotransferase; ALAT, alanine aminotransferase; HB, hemoglobin; HT, hematocrit; AUCi, area under the curve with respect to increase; AUCg, area under the curve with respect to the ground

Table 4 shows the distribution of clinical characteristics across the three identified latent classes. As expected because of the pre-stratified processing of the data, the classes did not differ with respect to gender or age. The 'lean' and 'average' classes were similar in terms of diagnoses, whereas subjects in the 'overweight' class were more likely to suffer from MD. Persons in the 'overweight' class were also more likely to endorse the atypical depressive subtype and to use psychotropic medication, specifically tricyclic antidepressants. Overweight subjects also had higher scores on both the IDS and the BAI. The 'lean' class had a lower age of onset compared to the other classes (i.e. 18.1 v. 21.6 v. 22.6). There were no differences in course or diagnoses at 2- and 6-years follow-up (see Supplementary Table S2). Individual symptoms of depression and anxiety did not differ between classes (data not shown).

LCA in the combined sample of patients and healthy controls (see Supplementary Table S3) again showed a 3-class model to be optimal. The classes were very similar to the original model with regard to their respective biomarker profiles (see Supplementary Figure S1). When compared across the classes, the percentage of subjects with current psychopathology was higher in the 'overweight' class, whereas the percentages of patients and controls in the 'lean' class were equal (see Supplementary Table S4).

Table 4. Distribution of characteristics across identified classes in the sample of subjects with current psychopathology

	Lean (*n* = 311)	Average (*n* = 910)	Overweight (*n* = 239)	*P**
Demographics and covariates				
Age, mean (SD)	40.5 (12.23)	42.1 (12.59)	43.2 (11.49)	0.033
Sex, % female (n)	66.9 (208)	66.2 (602)	66.1 (158)	0.971
Metabolic syndrome, % (n) (medication adjusted)	2.7 (8)	18.8 (167)	65.3 (154)	<0.001[a,b,c]
Number of chronic diseases				<0.001[b,c]
0, n (%)	135 (43.4)	381 (41.9)	69 (28.9)	
1, n (%)	102 (32.8)	301 (33.1)	74 (31.0)	
2, n (%)	47 (15.1)	143 (15.7)	55 (23.0)	
3+, n (%)	27 (8.7)	85 (9.3)	41 (17.1)	
Lifestyle factors				
Smokers, n (%)	48.2 (150)	39.9 (363)	43.1 (103)	0.035
Alcohol abuse (audit), mean (SD)	5.8 (6.34)	4.7 (4.97)	3.7 (5.11)	<0.001[a,b,c]
Medication use				
Psychotropics, % (n)	67.2 (209)	71.5 (651)	77.4 (182)	0.018[b,c]
SSRI, % (n)	22.5 (70)	26.3 (241)	28.4 (68)	0.457
TCA, % (n)	0.6 (2)	4.1 (37)	8.4 (20)	<0.001[b]
Benzodiazepine, % (n)	22.2 (69)	24.0 (218)	28.0 (66)	0.410
Current diagnosis (last month)				0.300
Depression only, % (n)	24.8 (77)	24.1 (219)	26.4 (63)	
Anxiety only, % (n)	37.0 (115)	37.1 (338)	29.7 (71)	
Comorbidity, % (n)	38.3 (119)	38.5 (353)	43.9 (105)	
Age of onset, mean (SD)	18.1 (11.15)	21.6 (12.77)	22.64 (12.87)	<0.001[a,b]
Number of depression diagnoses				0.192
1, % (n)	46.6 (145)	48.6 (442)	50.2 (120)	
2, % (n)	16.4 (51)	14.3 (130)	20.1 (48)	
Number of anxiety diagnoses				0.425
1, % (n)	47.3 (147)	48.9 (445)	42.3 (101)	
2, % (n)	19.0 (59)	20.4 (186)	24.3 (58)	
3, % (n)	9.0 (28)	6.6 (60)	7.2 (17)	

	Lean (*n* = 311)	Average (*n* = 910)	Overweight (*n* = 239)	*P**
Months with depression symptoms, mean (SD)	21.4 (16.33)	20.3 (14.19)	21.7 (17.10)	0.206
Months with anxiety symptoms, mean (SD)	26.8 (19.46)	26.7 (19.57)	24.5 (18.29)	0.544
Months with avoidance symptoms, mean (SD)	30.3 (21.01)	30.6 (22.55)	26.6 (22.12)	0.192
Internalizing diagnoses				
MD, % (n)	53.7 (166)	53.3 (485)	63.2 (151)	0.020[b,c]
Minor depression, % (n)	6.4 (20)	5.9 (54)	4.2 (10)	0.496
Dysthymia, % (n)	19.6 (61)	17.9 (163)	23.0 (55)	0.197
Social phobia, % (n)	42.8 (133)	36.8 (335)	35.6 (85)	0.126
Panic w. agoraphobia, % (n)	21.9 (78)	25.3 (230)	23.8 (57)	0.473
Panic w.o. agoraphobia, % (n)	10.6 (33)	10.4 (95)	11.7 (28)	0.850
Agoraphobia, % (n)	9.6 (30)	10.1 (92)	13.0 (31)	0.378
GAD, % (n)	27.3 (85)	26.9 (245)	28.0 (67)	0.941
Depression severity (IDS), mean (SD)	29.70 (11.64)	30.01 (12.11)	34.39 (12.95)	<0.001[b,c]
Anxiety severity (BAI), mean (SD)	16.64 (9.98)	18.02 (10.48)	20.99 (12.57)	<0.001[b]
Number of MD episodes, mean (SD)	4.9 (8.31)	5.4 (11.84)	4.7 (8.46)	0.333
IDS subscale				
Atypical, % (n)	15.0 (45)	19.6 (175)	28.4 (66)	<0.001[b]
Melancholic, % (n)	14.3 (43)	10.5 (94)	16.0 (37)	0.035
BAI subscale, mean (SD)				
Somatic	9.5 (6.73)	10.6 (7.03)	13.0 (8.26)	<0.001[b,c]
Subjective	7.1 (4.37)	7.4 (4.46)	8.0 (5.39)	0.395

AUDIT, alcohol use disorder identification test; BAI, Beck Anxiety Inventory; BMI, body mass index; CIDI, Composite Interview Diagnostic Instrument; GAD, generalized anxiety disorder; IDS, Inventory of Depressive Symptomatology; MD, major depression, NA, not applicable; SSRI, selective serotonin reuptake inhibitors; TCA, tricyclic antidepressants

*Based on ANOVA for continuous variables (post-hoc = Tukey), Kruskal-Wallis tests for non-normally distributed variables (post-hoc = Dunn test) and chi-square test for categorical variables. Significant differences by class (at $\alpha < 0.05$ corrected using the False Discovery Rate controlling method): [a]'lean' v. 'average' class. [b]'lean' v. 'overweight' class. [c]'average' v. 'overweight' class.

Discussion

This study aimed to use LCA in to identify subgroups with different biomarker profiles in a large sample of patients with depressive and anxiety disorders. A lean (21.6%), an average (62.2%) and an overweight (16.2%) class were identified. Overall, the model seemed to reflect somatic health status, which was confirmed by the observation of similar classes when healthy controls were included in the sample. Still, the fact that the subjects with psychopathology were relatively likely to be in the 'overweight' class compared with healthy controls, indicates a possible connection between these disorders

and the overweight biomarker profile. Furthermore, we found that the lean group had a significantly lower age of onset compared to the other two groups (see Table 4). Although the mean ages of onset are all in early adulthood and relatively close to each other, it cannot be excluded that there may be subtle differences in etiological mechanisms.

These bottom-up findings align with findings from research that used a more top-down approach and showed an association between depression and anxiety and obesity/MetS in this sample, and with results from large-scale meta-analyses.[39,56-62] The exact mechanisms behind these associations are unclear, partially because the (bi) directional nature of the association is still a point of discussion.[40,58,61,62] Obesity might cause depression and anxiety through social or biological mechanisms, but it might also be that the excessive weight gain is caused by the unhealthy lifestyle habits of patients. Medication use is another risk factor, as it is becoming apparent that many commonly used psychotropic medications such as antidepressants are associated with cardiometabolic risk factors such as insulin resistance, obesity, and dyslipidemia.[63] Evidence from the current sample supports this hypothesis. We found that psychotropic medication use is higher in the overweight subtype, and previous research in this sample showed a directional relationship between medication use and waist circumference and MetS.[40] Another interesting hypothesis is that depression and the MetS share a common cause, for instance, genetic risk factors.[64] Indeed, some evidence indicates that MetS and depression are caused by similar alterations of the stress system, including the hypothalamus-pituitary-adrenal axis, the autonomic nervous system, the immune system, and platelet and endothelial function.[39,57,65-67]

However, we found no association between the overweight subtype and course or diagnoses at 2- and 6-year follow-up. It is possible that no such association exists, but there are alternative explanations. It is possible that other causal factors that were not shared between depression and the MetS obfuscated the results. Furthermore, in this sample, it has been shown that especially MD subjects, with and without comorbid anxiety, are more likely to change weight compared to controls.[68,69] If subsequent biomarker changes also occurred, these subjects might have switched subtypes, and it might be that this change in subtype would have been more informative with respect to course than baseline subtype membership. Unfortunately, the current data did not allow for investigation of subtype changes because biomarker data needed to do so was not available at follow-up.

In accordance with previous symptom-based subtyping research[16,70-72] this study does not provide a bottom-up validation for the DSM-diagnosis categories of depression and anxiety disorders as patients with different diagnoses did not cluster into distinct biological classes. Also, in contrast to previous research[41,47,73,74], we did not find an

association between the latent classes and atypical v. melancholic features of depression. When comparing melancholic and atypical symptomatology on the IDS-SR between classes, the only significant difference was that the overweight class showed a higher percentage of atypical specifiers compared to the lean class. However, in absolute terms, both atypical and melancholic IDS-SR counts were highest for the overweight class, although differences on the melancholic subscale were no longer significant after correction for multiple comparisons.

Overall, the results indicate that biomarker heterogeneity among depressive and/or anxiety patients mostly reflects variations in somatic health that extend into the part of the population without mental health problems. However, this does not mean that part of these biological variations is not also related to psychological health. As stated above, variations in somatic health are known to be strongly related to variations in psychopathology. Additionally, there may be smaller but still relevant associations between specific sets of biomarkers and depression/ anxiety that were not detected in the current analyses but could be of strong interest for the development of more personalized etiological models. A future methodological challenge lies in better investigating if generic somatic-health-related biological effects can be separated from more specific psychopathology related biological effects. Possibilities for this may lie in the use of more flexible clustering algorithms, but also in the combination of biomarker and clinical data in the identification of subtypes. With regard to the biomarkers that could be investigated deeper, the current results suggested that there were several biomarkers that had smaller effects in the LCA results than the MetS biomarkers, but could still be potentially interesting as targets for further research, such as inflammation-related markers (e.g., Kynurenine, hsCRP) or sex-related markers (e.g., SHGB, testosterone).

Strengths and limitations

Strengths of the current study included the large sample size, the broad range of available biomarkers, and the presence of thorough clinical assessments. However, the results should also be interpreted in the context of several limitations. First, the results apply to a group of outpatients and results cannot be directly translated to other groups (e.g., inpatients). Second, LCA makes very strong assumptions (e.g., local independence) that enable LCA to estimate interpretable but strongly simplified models. However, we believe that (the violation of) the assumption of local independence is not a driving force in our model, because we made sure not to include strongly correlated pairs of biomarkers and found classes that were not defined by specific clusters of correlated biomarkers but rather by a collection of biomarkers from different domains (see Table 2). Third, the possible

negative influence of underweight could not be evaluated because people with a BMI < 18.5 were rare compared with overweight persons. Fourth, to facilitate the analyses, continuous biomarker measurements were coded to a discrete scale, possibly leading to a loss of information. Latent Profile Analysis, using continuous variables, usually provides a more nuanced representation of the data. Unfortunately, this technique could not be applied to our dataset because it is very sensitive to non-normal distributions.[75] Fifth, coding was stratified for sex and age groups, but other unknown/unmeasured factors were not considered. Finally, the biological data needed to estimate the subtypes was not available at follow-up, making it impossible to investigate subtype stability over time and the effects of subtype changes over time.

Conclusion

Three biological subgroups were identified with LCA among depressive and/or anxiety patients. These subgroups showed classes that (1) were strongly tied to general (metabolic) health, (2) did not reflect any natural cut-offs along the lines of the traditional diagnostic classifications, and (3) showed that especially poor metabolic health had a strong relationship with depression and anxiety and could, therefore, be seen as a distal marker for these types of psychopathology.

References

1. Kessler, R. C. The costs of depression. *Psychiatr. Clin. North Am.* **35**, 1–14 (2012).
2. Whiteford, H., Ferrari, A. & Degenhardt, L. Global Burden Of Disease Studies: Implications For Mental And Substance Use Disorders. *Health Aff. (Millwood).* **35**, 1114–20 (2016).
3. World Health Organization. Investing in treatment for depression and anxiety leads to fourfold return. (2016). Accessed from: http://www.who.int/mediacentre/news/releases/2016/depression-anxiety-treatment/en/.
4. Nutt, D. J., Ballenger, J. C., Sheehan, D. & Wittchen, H. U. Generalized anxiety disorder: comorbidity, comparative biology and treatment. *Int. J. Neuropsychopharmacol.* **5**, 315–325 (2002).
5. Hettema, J. M. The nosologic relationship between generalized anxiety disorder and major depression. *Depress. Anxiety* **25**, 300–316 (2008).
6. Goldberg, D. The heterogeneity of 'major depression'. *World Psychiatry* **10**, 226–228 (2011).
7. Kendler, K. S. The structure of psychiatric science. *Am. J. Psychiatry* **171**, 931–938 (2014).
8. Monroe, S. M. & Anderson, S. F. Depression: The Shroud of Heterogeneity. *Curr. Dir. Psychol. Sci.* **24**, 227–231 (2015).
9. Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A. & Rosenthal, R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* **358**, 252–260 (2008).
10. Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D. & Andersson, G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol. Med.* **40**, 211–23 (2010).
11. Cooney, G. Exercise for depression. *J. Evid. Based. Med.* **6**, 307–308 (2013).
12. Roest, A. M. *et al.* Reporting Bias in Clinical Trials Investigating the Efficacy of Second-Generation Antidepressants in the Treatment of Anxiety Disorders: A Report of 2 Meta-analyses. *JAMA Psychiatry* **72**, 500–510 (2015).
13. Baumeister, H. & Parker, G. Meta-review of depressive subtyping models. *J. Affect. Disord.* **139**, 126–140 (2012).
14. Lubke, G. H. & Muthén, B. Investigating Population Heterogeneity With Factor Mixture Models. *Psychol. Methods* **10**, 21–39 (2005).
15. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. *Springer Series in Statistics* (Springer, 2011).
16. Van Loo, H. M., De Jonge, P., Romeijn, J.-W., Kessler, R. C. & Schoevers, R. A. Data-driven subtypes of major depressive disorder: a systematic review. *BMC Med.* **10**, 156 (2012).
17. Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J. & Beckmann, C. F. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 433–447 (2016).
18. Hasler, G., Drevets, W. C., Manji, H. K. & Charney, D. S. Discovering Endophenotypes for Major Depression. *Neuropsychopharmacology* **29**, 1765–1781 (2004).
19. Kendell, R. & Jablensky, A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am. J. Psychiatry* **160**, 4–12 (2003).
20. Hasler, G. Pathophysiology of depression: do we have any solid evidence of interest to clinicians? *World Psychiatry* **9**, 155–61 (2010).

21. Kapur, S., Phillips, a G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).

22. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).

23. Tang, S. W., Helmeste, D. M. & Leonard, B. E. Antidepressant compounds: A critical review. in *Depression: From Psychopathology to Pharmacotherapy* **27**, 1–19 (2010).

24. Meyer, J. M. & Ginsburg, G. S. The path to personalized medicine. *Curr. Opin. Chem. Biol.* **6**, 434–438 (2002).

25. Bartova, L., Berger, A. & Pezawas, L. Is there a personalized medicine for mood disorders? *Eur. Arch. Psychiatry Clin. Neurosci.* **260**, 121–126 (2010).

26. Leuchter, A. F. *et al.* Biomarkers to predict antidepressant response. *Curr. Psychiatry Rep.* **12**, 553–562 (2010).

27. Simon, G. E. & Perlis, R. H. Personalized medicine for depression: can we match patients with treatments? *Am. J. Psychiatry* **167**, 1445–1455 (2010).

28. Simon, G. What little we know about tailoring depression treatment for individual patients. *Depress. Anxiety* **28**, 435–438 (2011).

29. e Silva, J. A. C. Personalized medicine in psychiatry: new technologies and approaches. *Metabolism* **62**, S40–S44 (2013).

30. Miller, D. B. & O'Callaghan, J. P. Personalized medicine in major depressive disorder - Opportunities and pitfalls. *Metabolism* **62**, S34–S39 (2013).

31. Ozomaro, U., Wahlestedt, C. & Nemeroff, C. B. Personalized medicine in psychiatry: problems and promises. *BMC Med.* **11**, 132 (2013).

32. Cuijpers, P., Ebert, D. D., Acarturk, C., Andersson, G. & Cristea, I. A. Personalized Psychotherapy for Adult Depression: A Meta-Analytic Review. *Behav. Ther.* **47**, 966–980 (2016).

33. Korte, S. M. *et al.* The many different faces of major depression: It is time for personalized medicine. *Eur. J. Pharmacol.* **753**, 88–104 (2015).

34. Penninx, B. W. J. H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17**, 121–140 (2008).

35. Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B. & Trivedi, M. H. The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychol. Med.* **26**, 477–486 (1996).

36. Rush, A. J., Carmody, T. & Reimitz, P. The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. *Int. J. Methods Psychiatr. Res.* **9**, 45–59 (2000).

37. Beck, A. T., Epstein, N., Brown, G. & Steer, R. A. An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psychol.* **56**, 893 (1988).

38. Chaves, P. H. M. *et al.* What constitutes normal hemoglobin concentration in community-dwelling disabled older women? *J. Am. Geriatr. Soc.* **52**, 1811–1816 (2004).

39. Vancampfort, D. *et al.* Metabolic syndrome and metabolic abnormalities in patients with major depressive disorder: a meta-analysis of prevalences and moderating variables. *Psychol. Med.* **44**, 2017–28 (2014).

40. Hiles, S. A., Révész, D., Lamers, F., Giltay, E. & Penninx, B. W. J. H. Bidirectional prospective associations of metabolic syndrome components with depression, anxiety, and antidepressant use. *Depress. Anxiety* **33**, 754–764 (2016).

41. Lamers, F. *et al.* Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. *Mol. Psychiatry* **18**, 692–699 (2013).

42. Quak, J. *et al.* Does tryptophan degradation along the kynurenine pathway mediate the association between pro-inflammatory immune activity and depressive symptoms? *Psychoneuroendocrinology* **45**, 202–210 (2014).

43. Patas, K. *et al.* Association between serum brain-derived neurotrophic factor and plasma interleukin-6 in major depressive disorder with melancholic features. *Brain. Behav. Immun.* **36**, 71–79 (2014).

44. Milaneschi, Y. *et al.* The association between low vitamin D and depressive disorders. *Mol. Psychiatry* **19**, (2014).

45. Giltay, E. J. *et al.* Salivary testosterone: associations with depression, anxiety disorders, and antidepressant use in a large cohort study. *J. Psychosom. Res.* **72**, 205–213 (2012).

46. Morita, T. *et al.* Plasma dehydroepiandrosterone sulfate levels in patients with major depressive disorder correlate with remission during treatment with antidepressants. *Hum. Psychopharmacol.* **29**, 280–286 (2014).

47. Milaneschi, Y., Lamers, F., Bot, M., Drent, M. L. & Penninx, B. W. J. H. Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol. Psychiatry* **81**, 807–814 (2017).

48. van Santen, A. *et al.* Psychological traits and the cortisol awakening response: Results from the Netherlands Study of Depression and Anxiety. *Psychoneuroendocrinology* **36**, 240–248 (2011).

49. Veen, G. *et al.* Evening salivary alpha-amylase, major depressive disorder, and antidepressant use in the Netherlands Study of Depression and Anxiety (NESDA). *Psychiatry Res.* **208**, 41–46 (2013).

50. de Geus, E. J. C., Willemsen, G. H. M., Klaver, C. H. A. M. & van Doornen, L. J. P. Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. *Biol. Psychol.* **41**, 205–227 (1995).

51. Licht, C. M. M. *et al.* Association between major depressive disorder and heart rate variability in the Netherlands Study of Depression and Anxiety (NESDA). *Arch. Gen. Psychiatry* **65**, 1358–1367 (2008).

52. Nylund, K. L., Asparouhov, T. & Muthén, B. O. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Struct. Equ. Model.* **14**, 535–569 (2007).

53. Muthen, L. K. & Muthen, B. O. Mplus user's guide. 5th. *Los Angeles, CA Muthén Muthén* 197–200 (2007).

54. Cramér, H. The two-dimensional case. in *Mathematical Methods of Statistics* (Princeton University Press, 1946).

55. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

56. van Reedt Dortland, A. K. B. *et al.* Associations between serum lipids and major depressive disorder: results from the Netherlands Study of Depression and Anxiety (NESDA). *J. Clin. Psychiatry* **71**, 1,478-736 (2009).

57. Van Reedt Dortland, A. K. B., Giltay, E. J., Van Veen, T., Zitman, F. G. & Penninx, B. W. J. H. Metabolic syndrome abnormalities are associated with severity of anxiety and depression and with tricyclic antidepressant use. *Acta Psychiatr. Scand.* **122**, 30–39 (2010).

58. Blaine, B. Does Depression Cause Obesity? *J. Health Psychol.* **13**, 1190–1197 (2008).

59. Luppino, F. S. *et al.* Overweight, Obesity, and Depression: A Systematic Review and Meta-analysis of Longitudinal Studies. *Arch. Gen. Psychiatry* **67**, 220–229 (2010).

60. Sardinha, A. & Nardi, A. E. The role of anxiety in metabolic syndrome. *Expert Rev. Endocrinol. Metab.* **7**, 63–71 (2014).

61. Ghanei Gheshlagh, R., Parizad, N., Sayehmiri, K., Zamanian-Azodi, M. & Rashidy-pour, A. Is there a relationship between metabolic syndrome and depression? A systematic review and meta-analysis. *Koomesh* **16**, 488–494 (2015).

62. Mannan, M., Mamun, A., Doi, S. & Clavarino, A. Is there a bi-directional relationship between depression and obesity among adult men and women? Systematic review and bias-adjusted meta analysis. *Asian J. Psychiatr.* **21**, 51–66 (2016).

63. Abosi, O., Lopes, S., Schmitz, S. & Fiedorowicz, J. G. Cardiometabolic effects of psychotropic medications. *Horm. Mol. Biol. Clin. Investig.* **36**, 1–15 (2018).

64. Bornstein, S. R., Schuppenies, A., Wong, M. L. & Licinio, J. Approaching the shared biology of obesity and depression: the stress axis as the locus of gene-environment interactions. *Mol. Psychiatry* **11**, 892–902 (2006).

65. Luppino, F. S., Bouvy, P. F., Giltay, E. J., Penninx, B. W. J. H. & Zitman, F. G. The metabolic syndrome and related characteristics in major depression: Inpatients and outpatients compared metabolic differences across treatment settings. *Gen. Hosp. Psychiatry* **36**, 509–515 (2014).

66. Marazziti, D., Rutigliano, G., Baroni, S., Landi, P. & Dell'Osso, L. Metabolic syndrome and major depression. *CNS Spectr.* **19**, 293–304 (2014).

67. Vogelzangs, N. *et al.* Inflammatory and metabolic dysregulation and the 2-year course of depressive disorders in antidepressant users. *Neuropsychopharmacology* **39**, 1624–1634 (2014).

68. de Wit, L. M., van Straten, A., Lamers, F., Cuijpers, P. & Penninx, B. W. J. H. Depressive and anxiety disorders: Associated with losing or gaining weight over 2 years? *Psychiatry Res.* **227**, 230–237 (2015).

69. Gibson-Smith, D. *et al.* Major depressive disorder, antidepressant use, and subsequent 2-year weight change patterns in the netherlands study of depression and anxiety. *J. Clin. Psychiatry* **77**, e144–e151 (2016).

70. Rhebergen, D. *et al.* An examination of generalized anxiety disorder and dysthymic disorder by latent class analysis. *Psychol. Med.* **44**, 1–12 (2013).

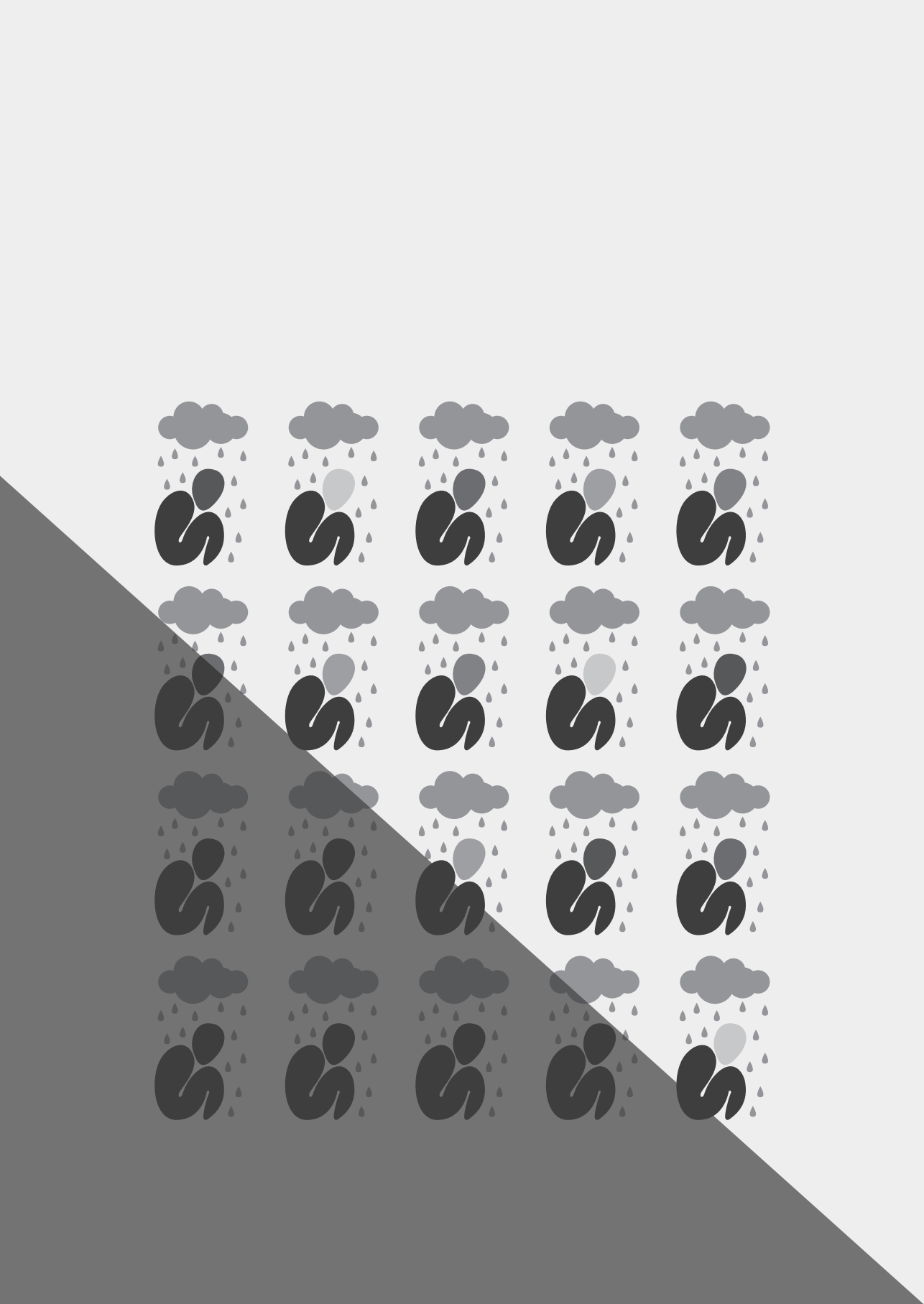
71. de Vos, S., Wardenaar, K. J., Bos, E. H., Wit, E. C. & de Jonge, P. Decomposing the heterogeneity of depression at the person-, symptom-, and time-level: latent variable models versus multimode principal component analysis. *BMC Med. Res. Methodol.* **15**, 88 (2015).

72. Wardenaar, K. J., Wanders, R. B. K., ten Have, M., de Graaf, R. & de Jonge, P. Using a hybrid subtyping model to capture patterns and dimensionality of depressive and anxiety symptomatology in the general population. *J. Affect. Disord.* **215**, 125–134 (2017).

73. Cizza, G. *et al.* Clinical subtypes of depression are associated with specific metabolic parameters and circadian endocrine profiles in women: the power study. *PLoS One* **7**, e28912 (2012).

74. Łojko, D., Buzuk, G., Owecki, M., Ruchała, M. & Rybakowski, J. K. Atypical features in depression: Association with obesity and bipolar disorder. *J. Affect. Disord.* **185**, 76–80 (2015).

75. Morgan, G. B., Hodge, K. J. & Baggett, A. R. Latent profile analysis with nonnormal mixtures: A Monte Carlo examination of model selection using fit indices. *Comput. Stat. Data Anal.* **93**, 146–161 (2016).



Chapter 6

Investigating data-driven biological subtypes of psychiatric disorders using Specification-Curve Analysis

Lian Beijers, Hanna M. van Loo, Jan-Willem Romeijn, Femke Lamers, Robert A. Schoevers and Klaas J. Wardenaar

Psychological Medicine, 1-12 (2020).

The supplementary material for this article can be found at https://doi.org/10.1017/s0033291720002846.

Abstract

Background

Cluster analyses have become popular tools for data-driven classification in biological psychiatric research. However, these analyses are known to be sensitive to the chosen methods and/or modelling options, which may hamper generalizability and replicability of findings. To gain more insight into this problem, we used Specification-Curve Analysis (SCA) to investigate the influence of methodological variation on biomarker-based cluster-analysis results.

Methods

Proteomics data (31 biomarkers) were used from patients with Major Depression (MD) (n = 688) and healthy controls (n = 426) in the Netherlands Study of Depression and Anxiety. In SCAs, consistency of results was evaluated across 1,200 k-means and hierarchical clustering analyses, each with a unique combination of clustering algorithm, fit-index, and distance metric. Next, SCAs were run in simulated datasets with varying cluster numbers and noise/outlier levels to evaluate the effect of data properties on SCA outcomes.

Results

The real data SCA showed no robust patterns of biological clustering in either the MD or a combined MD/healthy dataset. The simulation results showed that the correct number of clusters could be identified quite consistently across the 1,200 model specifications, but that correct cluster identification became harder when number of clusters and noise levels increased.

Conclusion

SCA can provide useful insights into the presence of clusters in biomarker data. However, SCA is likely to show inconsistent results in real-world biomarker datasets that are complex and contain considerable levels of noise. Here, the number and nature of the observed clusters may depend strongly on the chosen model-specification, precluding conclusions about the existence of biological clusters among psychiatric patients.

Introduction

Heterogeneity is a key feature of almost all psychiatric disorders.[1,2] Psychiatric patients usually present with a wide variety of clinical features (e.g., symptom patterns or treatment response[3–7]), and different underlying biological disturbances could be at play for patients with the same diagnosis[8]. Identification of more homogeneous diagnostic (sub)groups within larger diagnostic groups (e.g., depression, developmental disorders, psychosis) is often proposed as a starting point for increasing our understanding of more patient-specific etiological mechanisms, and thus, to advance the development of more biologically-informed, patient-specific diagnoses and personalized treatment.[2,8,9]

Identification of psychiatric diagnoses and subtypes has traditionally been based on clinical judgement and consensus.[10] Data-driven cluster analyses can be used to further reduce psychopathological heterogeneity by identifying patterns in data that are missed by clinical pattern recognition.[11] Although the call to apply data-driven approaches to psychiatric disease classification has been around for decades[12], their popularity rose notably in recent years[11,13–17]. This is likely due to a combination of factors, including the availability of suitable datasets, increased computational capabilities and ongoing advances in the fields of statistics and machine learning that make it possible to extract information from complex and high-dimensional data.[11,18,19] Data-driven clustering techniques have been used to gather evidence about possible subtypes in a broad range of psychiatric patient populations, including depression[13,17], psychosis[20–22], bipolar disorder[15] and developmental disorders (attention deficit hyperactivity disorder[23], autism spectrum disorder[16]).

The predominant approach used in psychiatry has been unsupervised learning in the form of finite mixture models (FMMs) and clustering algorithms (i.e., k-means clustering, hierarchical clustering, and community detection).[11] Unsupervised methods have been widely used for discovering subtypes within clinical groups because supervised learning, which aims to correctly predict the subject labels (e.g., patients vs. healthy control), is fundamentally limited by the quality of the clinical labels and cannot be used to investigate the validity of these labels.[24] Unsupervised learning does not use labels, but rather attempts to find subgroups based on data structure and heuristics used by each algorithm. Although the use of data-driven clustering techniques seems promising, there is also reason for caution. Scientific results are known to not always be robust and specifics of a chosen analytical method can have a significant influence on research outcomes.[25–27] In case of cluster analyses, however, there is usually no way of knowing if the results of a presented analysis would have been the same if different model specifications had been

used, as researchers will generally perform only one or two separate analyses.[11] Better insight into the effects of model specifications on unsupervised clustering results could greatly improve our understanding of data-driven psychiatric subtyping. In addition, it could provide leads for data-driven subtypes of Major Depression (MD) by identification of patterns that are robust to methodological variation.

In unsupervised learning, analytical variations across studies are a realistic risk because of the large availability of different model specifications for unsupervised learning algorithms. This is likely due to the lack of straightforward way to judge the quality of unsupervised learning results because there is no outcome measure, as opposed to supervised learning, which either succeeds at predicting a predefined outcome or not.[28] We decided to focus on k-means and hierarchical clustering because these have been shown to be the most commonly used methods across disorders, and FMMs have previously been shown to have a number of issues.[11,29–31] Within k-means/hierarchical clustering, there are three main aspects of the method that can vary: (1) algorithm, (2) distance metric (used to determine dissimilarity between data points) and (3) fit index (decides which is the optimal number of clusters). When investigating the 13 studies mentioned by Marquand et al. (2016) we found that k-means clustering was used most often, but that a specific rationale or justification for this choice was generally not given (8/13). This is likely due to the fact that because of the aforementioned lack of gold standard, we rely on simulation studies for algorithms[32–35] as well as distances[34,35] and fit indices[36,37]. These studies are performed only rarely and generally have mixed results.[32–37]

The current study aimed to identify clusters in in a psychiatric sample and to gain insight into the effects of different model specifications on the results by applying a *Specification-Curve Analysis* (SCA)[38] to a selected group of unsupervised machine learning algorithms (k-means clustering and six hierarchical clustering algorithms). SCA was developed to investigate the effects of methodological variations on regression results in psychology but can be also applied to study the effect of different model specifications in unsupervised machine learning analyses. When applied to the current case of cluster analysis, SCA considers the results of a large range of model specifications jointly, instead of using cluster analysis with just one or two model specifications. Because SCA has never been applied to cluster analysis before, we also investigated the influence of data properties such as the true number of existing clusters in the data and varying levels of noise on the SCA outcomes.

For this study, we focused on the identification of biological proteomics-based subtypes of MD. There have been increasing efforts to identify homogeneous clusters of MD patients, mainly based on clinical data. The results of these studies tend to be

unstable, or find subtypes mainly based on severity.[13] Fewer efforts have been based on biological measures.[17] There are some indications that biology-based clustering suffers from a similar degree of variation, likely due (at least in part) to the large variability in used methodology.[17] In this study we investigated if proteomic-based subtypes are indeed sensitive to different model specifications, or that we could find robust subtypes using proteomics data. Our specific aims were to (1) evaluate the influence of model specifications on the number of identified data-driven biological clusters in MD, (2) to investigate if SCA identifies clusters with distinct biological patterns that are robust to variations in model specifications, and (3) to run simulations to investigate how data properties influence SCA cluster results.

Methods and materials

For a visual overview of the complete analytical process, see Figure 1.

Participants and procedures

NESDA is a multisite naturalistic cohort study that examines the long-term course of depressive and anxiety disorders. A detailed description of the NESDA design can be found elsewhere.[39] In brief, the NESDA cohort consists of 2,981 subjects aged 18—65 years, including those with a lifetime anxiety and/or depressive disorder and a subgroup of healthy controls. The research protocol was approved by the Medical Ethical Committees of participating institutes, and after a complete description of the study, all respondents provided written informed consent. For the present study, all 688 subjects with a current (past 6 months) diagnosis of MD according to the Composite International Diagnostic Interview (CIDI; WHO version 2.1) as well as 426 healthy controls were selected. The SCA was first run in the MD patient sample and then repeated in the combined MD and healthy control sample (see below).

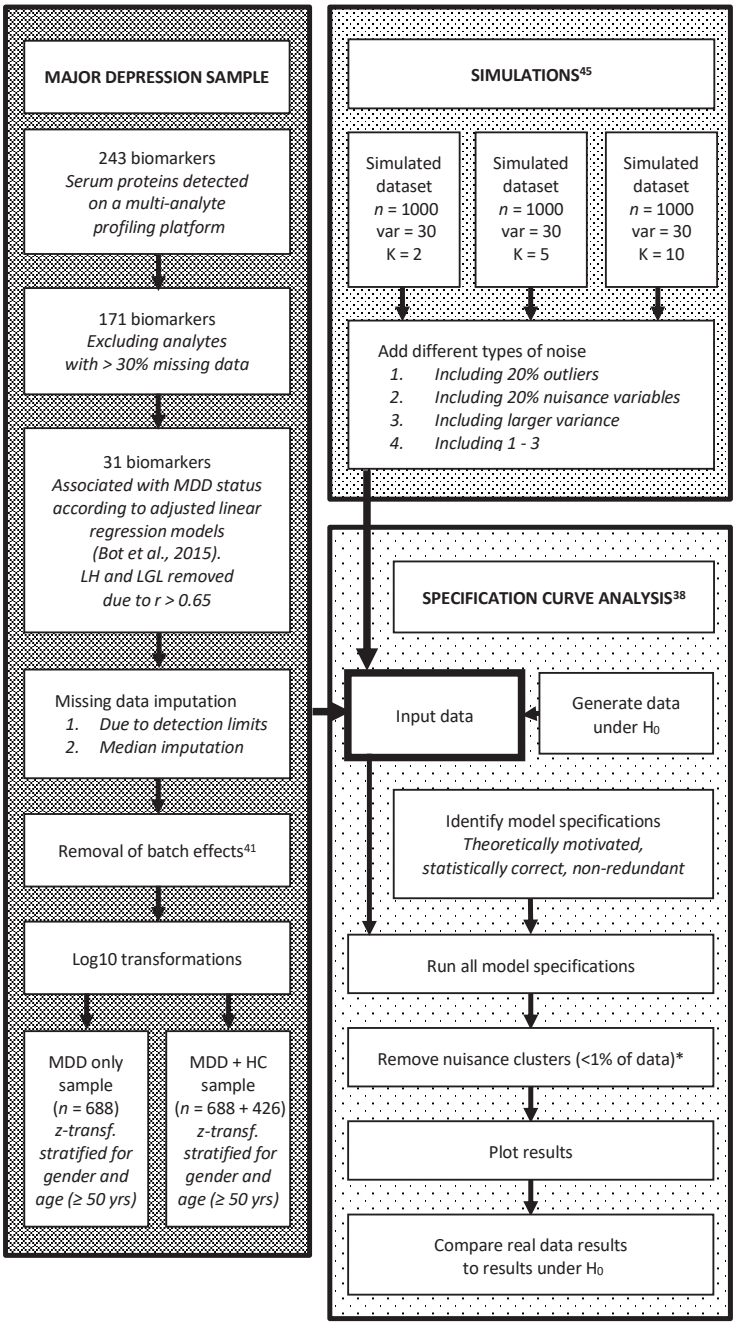


Figure 1. Flowchart of the complete analytical process, including real data preparation, data simulation and specification curve analysis.

Measurements

Extensive information was gathered through face-to-face interviews, a medical examination, a cognitive computer task and collection of blood samples.[39] DSM-IV diagnoses of depressive (minor depression, dysthymia and MD) and anxiety disorders (Generalized Anxiety Disorder, Social Phobia, Agoraphobia and Panic Disorder) were established using the CIDI. Those without any diagnosis according to the CIDI were included as healthy controls.

Proteomic analytes

Blood was sampled after an overnight fast in five research centers throughout the Netherlands (Amsterdam, Leiden, Groningen, Emmen and Heerenveen), and stored at −80 °C. All samples were shipped on dry ice and processed from frozen in a Clinical Laboratory Improvement Amendments-certified laboratory (Myriad RBM; Austin, TX, USA), where a panel of 243 analytes (Myriad RBM DiscoveryMAP 250+) involved in various hormonal, immunological, and metabolic pathways were assessed in serum using multiplexed microbead immunoassays. Each batch also contained three duplicate control samples with different protein concentrations, giving an average inter- and intra-assay variability of 10.6% (range 5.5–32.5%) and 5.6% (range 2.5–15.8%), respectively.

Analyte data selection

To reduce the likelihood that identified clusters would merely reflect degrees of general somatic health rather than psychopathology, only biomarkers were included that were previously shown to differ between current MD patients and healthy controls.[40,41] We excluded the luteinizing hormone and lactoylglutathione lyase because of correlations > 0.65 with follicle-stimulating hormone and macrophage migration inhibitory factor, respectively. A total of 31 biomarkers related to immune response, protein metabolism, and diverse cell communication and signal transduction processes were included in the study (See Table 1 and Supplementary Table S1). Because biomarkers were selected based on their ability to discriminate between MD and healthy controls, the combined MD and healthy sample was expected to contain at least two clusters ($K \geq 2$).

Table 1. Biochemical analytes and associated biological processes

Analyte	Biological process*
Alpha-1-antichymotrypsin	PM
Alpha-1-antitrypsin	PM
CD40 antigen	CC,ST
Complement factor H-related protein 1	IM
ENRAGE	CC,ST
Growth-regulated alpha protein	IM
Interleukin-12p40	IM
Interleukin-1 receptor antagonist	CC,ST
Macrophage migration inhibitory factor	CC,ST
Lactoylglutathione lyase (not included because of high correlation with MIF)	M
Insulin growth factor-binding protein-5	CC,ST
Urokinase-type plasminogen activator receptor	CC,ST
Cathepsin D	PM
Receptor tyrosine-protein kinase erbB-3	CC,ST
Hepsin	PL
Cellular fibronectin	CG
Matrix metalloproteinase-10	PM
Matrix metalloproteinase-3	PM
Tenascin C	CC,ST
Carcinoembryonic antigen	IM
Angiogenin	M
Angiopoietin 2	CC,ST
Vascular endothelial growth factor	CC,ST
Apolipoprotein A4	T
Apolipoprotein D	T
Fatty acid-binding protein, adipocyte	CC,ST
Pancreatic polypeptide	CC,ST
Von Willebrand factor	PM
Luteinizing hormone (not included because of high correlation with FSH)	CC,ST
Follicle-stimulating hormone	CC,ST
Cystatin C	PM
Fetuin-A	CC,ST
Prostasin	PM

Abbreviations: CC, cell-cell communication; CG, cell growth/maintenance; ENRAGE, advanced glycation end-products binding protein; FSH, Follicle-stimulating hormone; IM, immune response; M, metabolism; MIF, Macrophage migration inhibitory factor; PL, proteolysis and peptidolysis; PM, protein metabolism; ST, signal transduction; T, transport

* From the Human Protein Reference Database, according to Bot et al 2015.

Data processing

Missing values due to biomarker values being below or above the detection limits were imputed with the values of the lower and upper detection limit, respectively. Other missing values were imputed by the median value (see Supplementary Table S1 for missing value percentages). We applied the ComBat function, including all covariates used previously by Bot et al. (2015), to remove any potential plate effects.[42] Data were log10-transformed to normalize the variance distributions. Because various clustering techniques are sensitive to the relative scaling of variables, we performed z-score transformations, separately for the MD sample and the combined patient and control sample. Transformations were stratified for gender and age (≥ 50 years vs. < 50 years) to prevent these variables from driving the model solutions.

Statistical analyses

SCA consists of three steps.[38] First, the researcher identifies a set of theoretically justified, statistically correct, and non-redundant analytic specifications. Second, the analysis is run with each specification and the results (i.e., number of identified clusters; y-axis) are plotted as a function of analysis specification (x-axis), which allows for the identification of (in)consistency across specifications. Third, the researcher determines whether the resulting curve is inconsistent with the null hypothesis (H_0: no clusters present). It is difficult to test the results of any SCA with a statistical test because the specifications are neither statistically independent nor part of a single model.[38] Therefore, this is done by bootstrapping. The researcher generates many datasets that are in accordance with the null hypothesis (i.e., no clusters present), and runs the complete set of specifications on each of these H_0 dataset. If the curve based on the real dataset falls outside of the range of expected results based on the bootstrapped H_0 datasets, H_0 can be rejected.

Analytic specifications

Using the package *NbClust_3.0*[43] in *R_3.6.1*, we performed an SCA with 1,200 individual cluster analyses representing all possible model specifications within the most popular non-parametric clustering algorithms (i.e., agglomerative hierarchical clustering and k-means cluster analysis[44]). The 1,200 specifications (see Supplementary Table S2) each represented a unique combination of a *clustering algorithm* (7 options), *distance metric* (determines the distance between data points; 6 options), and *fit index* (identifies the optimal number of clusters; 21 options). Graphical or computationally expensive fit indices were not included. The current large range of available options was included, because there is currently very little evidence to prefer one over the other.[37]

Model selection

In order to approximate what researchers would do when conducting a cluster analysis, we tested 1-15 clusters in each of the 1,200 cluster analyses, and then selected the best model based on the fit index. In addition, we excluded small clusters (≤ 1% of subjects), whilst retaining the other clusters in each model, because small clusters usually include only one or two subjects with extreme values (outliers), and the other clusters may still hold interesting information.

Evaluating the null

In order to generate datasets that were in accord with the null hypothesis (H_0: no clusters present), we created 500 datasets, in which all variables were statistically independent. This was done by selecting a random value from every biological variable for each participant. Next, we ran the SCA in each of these datasets and created the range of expected results. First, the results based on every dataset were ordered from smallest to largest number of clusters (*K*). Then we combined the 500 results, and the 2.5^{th} and 97.5^{th} percentile for each position 1 to 1200 were identified, representing the lower and upper of the expected number of clusters (*K*) under H_0. Therefore, these results do not give the expected range of the specific combination of options, but rather the range of the m^{th} smallest *K*. In order for a real-data SCA to reject H_0, the results must fall outside this range.

Cluster stability

Between models with the same *K*, the cluster sizes and allocation of subjects can differ. If *K* clusters truly exist in the data, we expect the model solutions to be relatively stable with respect to these characteristics across different model specifications that yielded *K* clusters. Cluster stability was assessed with a few simple metrics. First, we identified the number of unique model solutions for each group of models with the same number of clusters (*K*). Second, we ranked the models based on the number of times they occurred. Third, we checked the number of solutions that occurred only once in the group of models with the same *K*. Finally, we assessed the stability of subject allocation to clusters by comparing the most often occurring model with the second and third ranking model solutions. We then quantified the number of subjects that switched classes between these model solutions.

Simulations

We performed a simulation study, aiming to investigate if a known cluster structure is indeed detected as the most consistent in an SCA, and to evaluate the effects of noise

and outliers. We simulated datasets using the R-package *clusterlab_0.0.2.6*.[45] Data were simulated with 2, 5 and 10 clusters, with subjects equally distributed across clusters (total n = 1000). The data were simulated with Gaussian variance 1 and circle circumference K+1 to create data without cluster overlap (baseline data). In addition, we simulated noisy datasets with different characteristics:

1. Including 20% outliers (distance 4)
2. Including 20% nuisance variables (randomly selected values with the same mean/standard deviation as the other variables)
3. Including a larger variance (v = 2), in order to have ~30% overlap
4. Including all of the above

For the first and second principal component coordinates of these datasets, see Supplementary Figure S1.

Results

Specification curve analysis in MD sample

Figure 2 shows the descriptive specification curve for the MD sample (Supplementary Table S3 shows sample characteristics). Forty-two specifications resulted in an error (see Supplementary Table S4). More than half of specifications (60.2%) resulted in models containing one or more small clusters (n ≤ 1%) that were excluded (see Supplementary Table S5). The resulting number of valid clusters was variable, although most models indicated no cluster structure (median = 1, IQR = 1-2). Interestingly, all analyses using the centroid, median or single-linkage algorithms indicated no clustering (K = 1), whereas single-cluster results were relatively uncommon for k-means, Ward and complete-linkage clustering algorithms (m = 5/150, m = 43/144 m = 44/144 respectively).

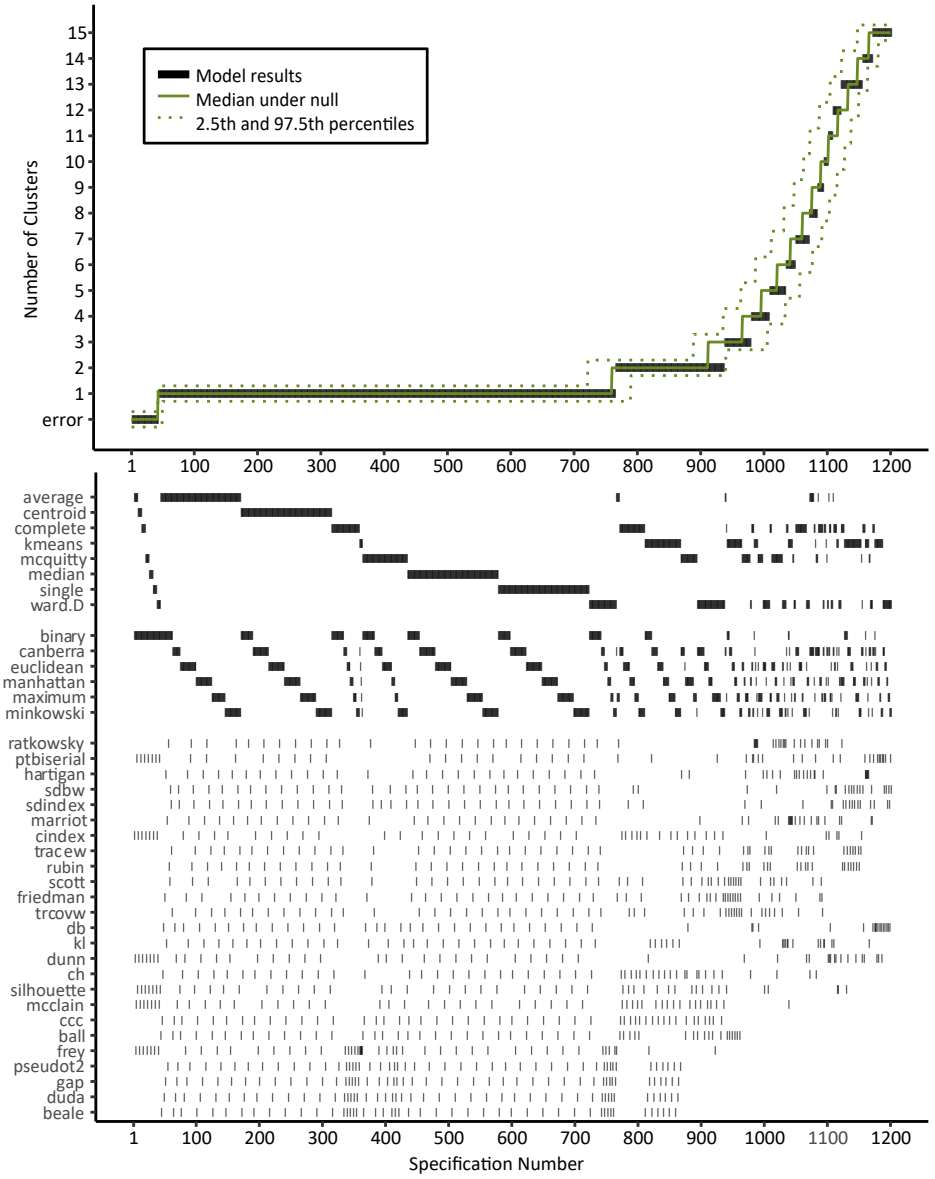


Figure 2. Descriptive Specification Curve in the sample with MD subjects only, with small clusters (≤ 1% of subjects) removed

Each black dot in the top panel depicts an estimate of the optimal number of clusters (*K*) from a different specification; the dots vertically aligned in the lower panel indicate the analytic decisions behind those estimates. The green lines indicate the expected range of results at each position. N.B. this is not the expected range of the specific combination of options, but rather the range of the m^{th} smallest K.

Based on Figure 2, we cannot readily conclude that any cluster structure is present, because the observed curve overlaps strongly with the curves based on the randomly drawn data. More specifically, although many specifications resulted in a solution with $K \geq 2$, this did not provide solid evidence for existing clusters as no result K was found more often in the real data compared to the random data.

Subject allocation showed limited stability, as indicated by different cluster sizes between model solutions and multiple distinct model solutions within each group of specifications with the same K (see Table 2). For example, for $K = 2$, the stability of subjects' cluster allocations between the most common two-cluster model (33.1%) and the second most common two-cluster solutions (11.6%) was only 56.8%.

Table 2. Stability measures of models with different numbers of clusters (K) for the MD dataset

K	Number of models % of 1200, (*n*)	Distinct solutions,	Dominant solution*, % (*n*)	Unique solutions⁺, % (*n*)
1	60.2 (722)			
2	14.3 (172)	15	33.1 (57)	0.6 (1)
3	3.5 (42)	8	57.1 (24)	2.4 (1)
4	2.4 (29)	7	34.5 (10)	3.4 (1)
5	2.2 (26)	9	38.5 (10)	11.5 (3)
6	1.2 (15)	7	46.7 (7)	26.7 (4)
7	1.8 (22)	6	36.4 (8)	0 (0)
8	1.1 (13)	5	53.8 (7)	23.1 (3)
9	0.8 (10)	5	30 (3)	10 (1)
10	0.6 (7)	6	28.6 (2)	71.4 (5)
11	0.6 (7)	5	28.6 (2)	42.9 (3)
12	1.1 (13)	7	30.8 (4)	30.8 (4)
13	2.8 (34)	4	79.4 (27)	5.9 (2)
14	1.3 (16)	7	37.5 (6)	18.8 (3)
15	2.5 (30)	6	43.3 (13)	0 (0)
Error	3.5 (42)			

*the model solution (i.e., specific division of subjects) that occurs most often within the group of models containing K clusters

⁺number of model solutions that occur only once

When healthy controls were included, the SCA was very similar (see Supplementary Fig. S2 and Supplementary Tables S6 and S7). Although $K = 2$ was expected here, 2-cluster solutions were not found more often in this dataset compared to the random datasets.

Simulated data

Figures 3-5 show the specification curves for simulated datasets. These showed that it is possible to detect the true number of clusters as the most consistent in the SCA, but that this is harder with larger number of clusters. In the noise-free two-cluster data, most model specifications (65.5%) resulted in two clusters (see Figure 2 and Supplementary Tables S8 and S9). For the dataset with five and 10 clusters, these percentages were 33.6% (see Figure 3 and Supplementary Tables S10 and S11) and 25.4% (see Figure 4 and Supplementary Tables S12 and S13), respectively. Within specifications with the correct results, the classification accuracy was almost 100% for the three most common model solutions in each of the three noise-free datasets. Consequently, the stability of subject allocation was high between models.

Increasing the level of noise in the simulated datasets led to a decrease in correctly identified results in the SCA. SCAs in data with 20% noise variables showed a similar number of correct results as in datasets without noise (62.3%, 33.6%, and 25.4% respectively). However, transforming 20% of the sample to outliers did have a larger effect, especially in the two-and five-cluster datasets, where the number of correctly identified clusters in the SCA was similar to the SCA results obtained in datasets with added noise, outliers and cluster overlap (22.8% vs. 19.1% and 2.6% vs. 2.1% respectively). Increasing the variance especially influenced the number of correctly identified results in the ten-cluster dataset: 3.6% correct results compared to 0.9% with added noise, outliers and increased variance.

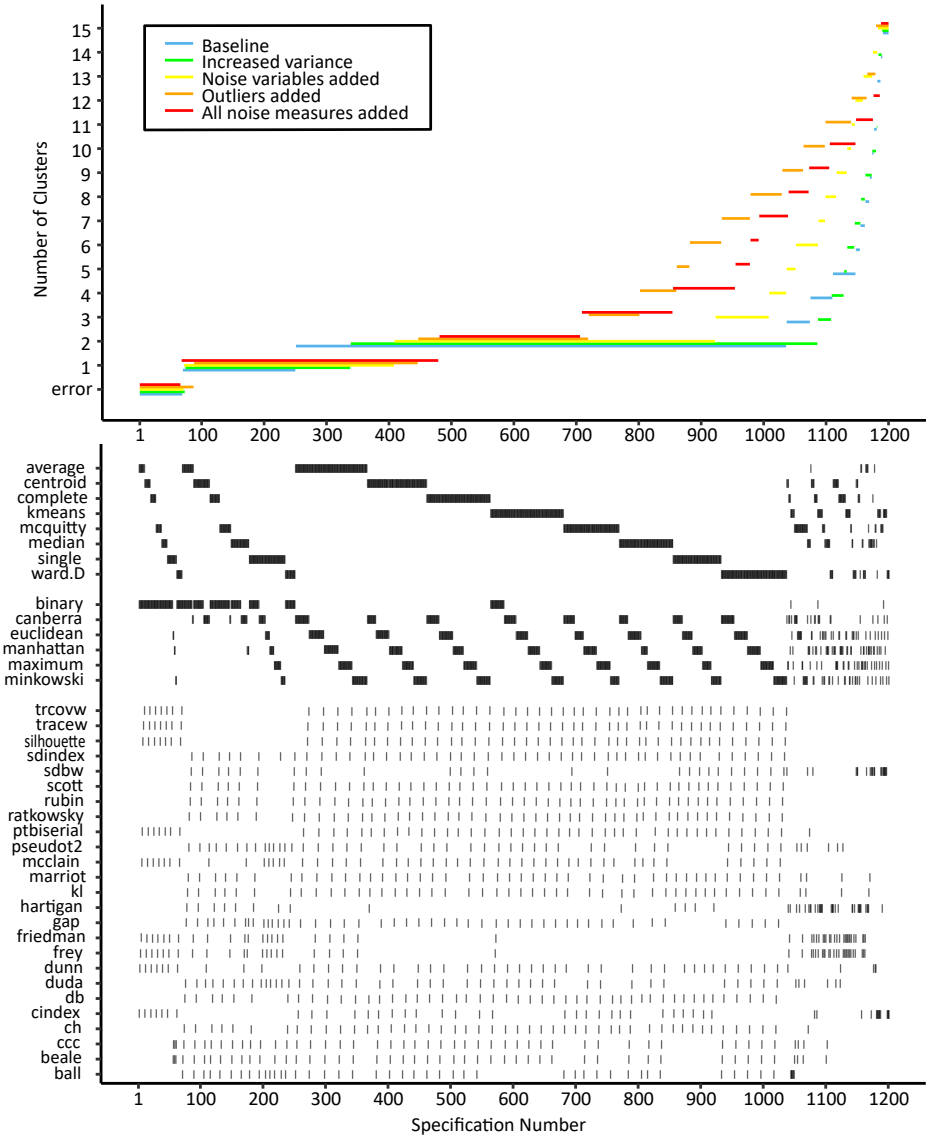


Figure 3. Specification curves based on simulated datasets with K = 2, with small clusters (≤ 1% of subjects) removed

Each dot in the top panel depicts an estimate of the optimal number of clusters (K) from a different specification; the dots vertically aligned in the lower panel indicate the analytic decisions behind the estimates of the baseline analysis. N.B. the analytic decisions behind the other analyses are not presented here.

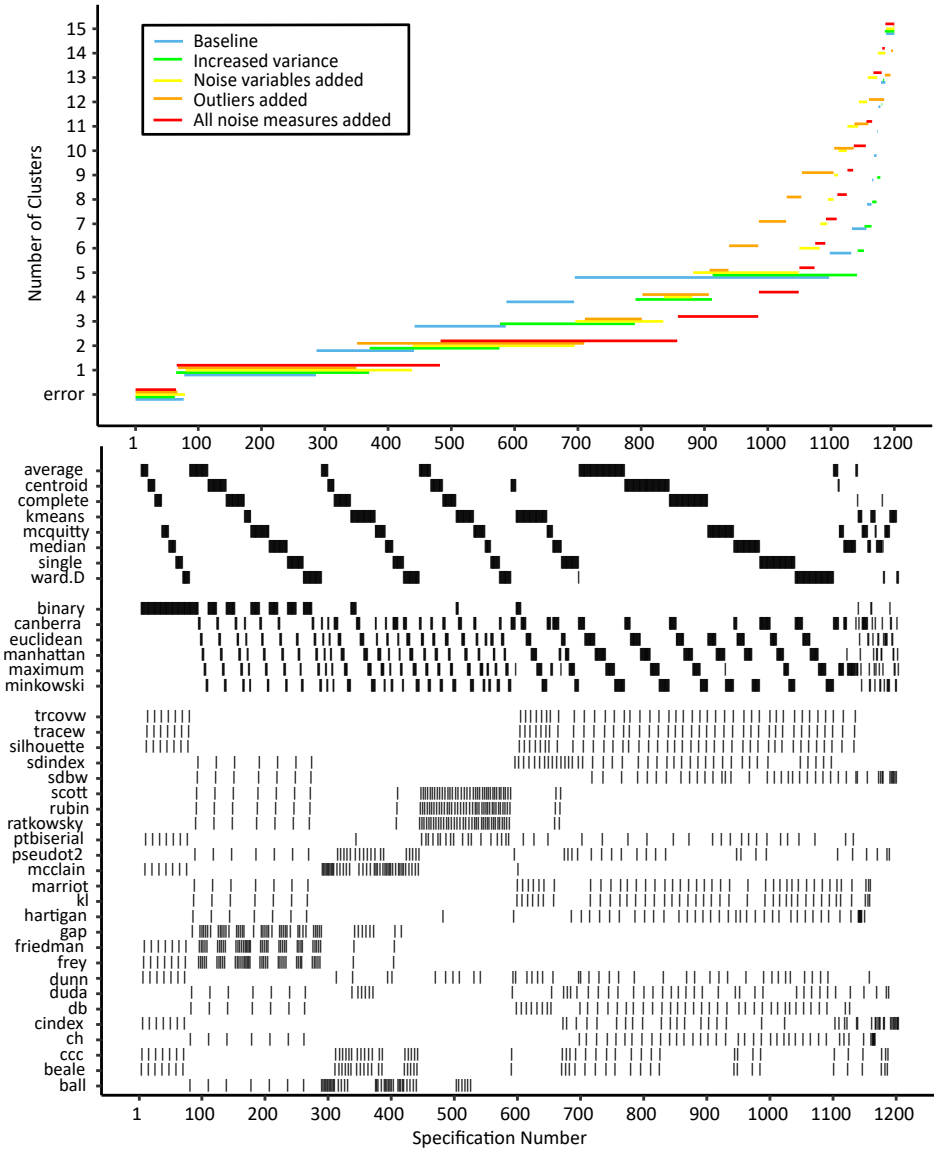


Figure 4. Specification curves based on simulated datasets with K = 5, with small clusters (≤ 1% of subjects) removed

Each dot in the top panel depicts an estimate of the optimal number of clusters (K) from a different specification; the dots vertically aligned in the lower panel indicate the analytic decisions behind the estimates of the baseline analysis. N.B. the analytic decisions behind the other analyses are not presented here.

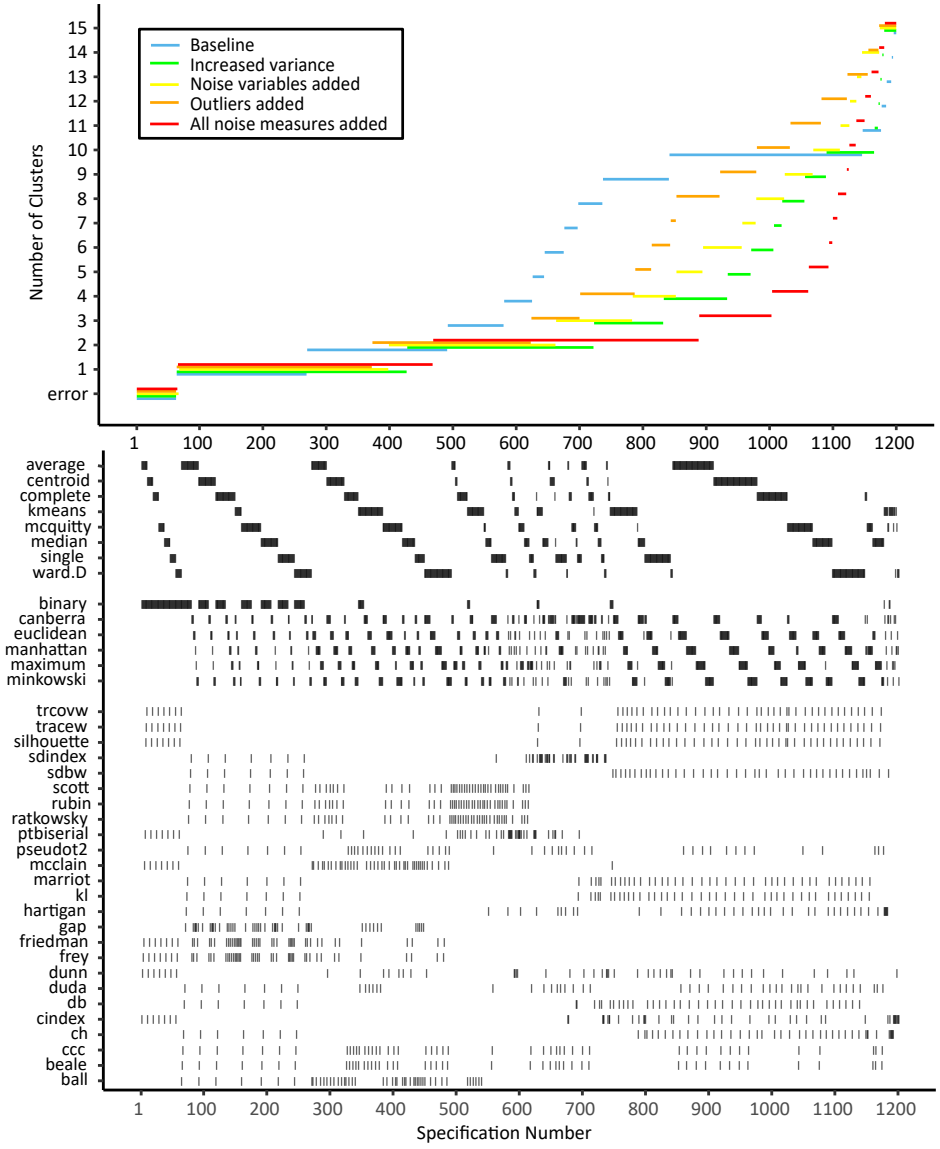


Figure 5. Specification curves based on simulated datasets with K = 10, with small clusters (≤ 1% of subjects) removed

Each dot in the top panel depicts an estimate of the optimal number of clusters (K) from a different specification; the dots vertically aligned in the lower panel indicate the analytic decisions behind the estimates of the baseline analysis. N.B. the analytic decisions behind the other analyses are not presented here.

Discussion

We investigated the presence of data-driven biological clusters of depression and evaluated the effect of different model specifications on these findings. The cluster-analysis results based on our sample of MD patients were very sensitive to the model specifications used. The SCA showed that the number of identified clusters was inconsistent, and that cluster allocation stability was low. Together, these observations indicated no robust cluster structure in the real dataset. This was also the case for the sample including healthy controls. Moreover, our analyses showed that many specifications will result in a cluster solution even when no structure is present in the data. The simulation study showed that it is possible for SCA to correctly identify clusters as the most consistent solution if they are present in the data, but that this becomes more difficult with large number of clusters and/or higher noise levels. Below, implications of these results are discussed.

As discussed in the introduction, the variability in results of previous cluster analyses raise inevitable questions about how much confidence we should put in results from a single cluster analysis, especially when this single analysis lacks replication in independent samples and clinical validation (e.g., differences in risk factors or course).[11,13,17] Our study aimed to investigate if the faith in model results improves when SCA is applied. The simulation results are somewhat encouraging, but the lack of a robust cluster structure in the real dataset including the one with both MD patients and healthy controls raises several concerns. How can we explain that the NESDA study found differences in biomarkers between cases and controls, but we do not find them in cluster analyses using the same biomarkers? Should the results bring into question the applicability of cluster techniques to biological data and therefore caution against any future use of such techniques?

It is possible that we did not find clusters in the real dataset because of technical issues. It could be, for instance, that the differences between cases and controls are too small to be picked up by cluster analysis, or that there is not sufficient correlation between the biomarkers, or that the signal-to-noise ratio is insufficient for cluster detection.

Alternatively, the fact that the SCA was not able to distinguish between MD patients and controls could indicate that the DSM categories cannot be validated using this specific type of biological data. Some, but not all, of the used biomarkers have been shown to be associated with depression before. For example, macrophage migration inhibitory factor, a pleiotropic cytokine, has been shown to be higher in MD patients compared with controls in five out of six studies.[46] Interleukin-1 receptor antagonist has also been shown to be increased in patients compared to controls.[47,48] The von Willebrand factor, a marker involved in haemostasis, was previously found to be increased in one study[49], which is

supported by earlier genetic findings of an association between depressive symptoms and a specific von Willebrand allele in cardiac patients.[50] Pancreatic polypeptide, which was elevated in patients, has been linked to anorexia nervosa [51], and another member of the pancreatic polypeptide family, peptide YY, was (marginally) positively related to depressive symptoms in older adults[52]. The other individual markers that were identified by Bot et al. (2015) were not associated with MD in previous studies or have not previously been investigated. For instance, the lower levels of growth-regulated alpha protein were in contrast with a study that found higher levels - although this result was not significant in the validation cohort.[52]

The simulation results indicated that it is difficult to identify stable/robust clusters, even when they do in fact exist, as they showed the analyses' sensitivity to data complexity (i.e., number of clusters), increased noise, and/or the presence/number of outliers. This is also the case for analyses based on single specification simulations.[32] In some cases (i.e., low numbers of clusters, little noise) it is likely still possible to identify any robust clusters present with SCA. In that case, results should be considered much more reliable than that of a single analysis, because the former is robust to differences in model specifications. This has already been shown in social psychology, where for example the negative impact of racial bias on call-back rates in job application processes has been shown to be robust, whereas increased death toll of female-named hurricanes was not.[38]

Limitations

Our study should be considered in the light of the following limitations. First, we used 31 biomarkers that were previously shown to differ between patients with current MD and healthy controls using adjusted linear regression.[41] It is possible that other biochemical markers are more suitable for finding clusters of MD patients. Currently, it is unknown which measures are best suited for biological subtyping of depression[17], so it could also be that brain structure or functional connectivity[53], or genetic background[54] could be more suitable for clustering MD patients. Furthermore, it could be that inter-personal variations in psychiatric samples are better captured by continuous distributions (e.g., severity dimension[s]) rather than discrete clusters.[31,55–57]

Second, SCA has traditionally been used in psychology to investigate the effects of using alternative regression models.[38,58,59] Cluster techniques are more complex. Two three-cluster solutions may be completely different in size and subject allocation, whereas a two- and a three-cluster solution may be partially overlapping. It is therefore important to keep in mind that this application of SCA focuses mainly on the resulting number of clusters and cluster stability, rather than the substantive interpretation of the clusters.

Had we found an optimal number of clusters ($K_{optimal}$) with a stable model solution, we would have investigated if the movement of subjects between models with $K_{optimal}$ -1 and with $K_{optimal}$ was stable. If this would have been the case, we would have investigated the movement of subjects between models with ever decreasing *K*, in order to investigate if there was a stable division tree to be made all the way from $K = 1$ to $K_{optimal}$.

Third, we used a limited number of model specifications for unsupervised learning. We focused on k-means clustering and hierarchical clustering because these are among the most commonly used methods across disorders[11], and FMMs have been shown to have a number of issues that limit their usefulness for psychiatric classification. FMMs tend to detect groups with different severity levels, which is not always the aim of cluster analysis, and local dependence between variables can obfuscate the results.[29–31] Because there is insufficient evidence on which model clustering algorithms, distances, and fit indices are most useful for a study like ours, we decided to study all of the potential model specifications, and not to exclude any a priori. We decided to use the exhaustive list of options in the NbClust R-package, which was designed to gather all indices available in SAS and R packages together into a single one package as well as some newer indices that are not implemented anywhere else yet.[43]

Fourth, we did not perform a Monte Carlo SCA but rather used SCA to evaluate the result obtained in a single simulation study. There is no Monte Carlo element in our procedure as we did not seek to quantify clustering quality of SCA or a single specification per se. Rather, our simulations aimed to evaluate whether, in the presence of a known number of clusters in a population, SCA can robustly show this number across different model specifications. Therefore, we used simulated datasets to illustrate the use of SCA under different circumstances (different numbers of clusters, noise levels). In total, we only simulated 15 datasets (i.e., 2, 5 and 10 clusters with 5 different noise levels). We chose to simulate different noise levels by increasing the number of outliers[34], varying the number of informative variables[35] and different degrees of separation between the clusters[33,35,60] (i.e., increasing variance), but other methods of simulating noisy datasets also exist[60].

Finally, it is important to remember that there are still many sources of variation left in our analyses, as can be seen in Figure 1. For example, we limited our analysis to a single MD dataset with a limited set of markers, because the primary focus was on the influence of model specifications on the results and not on the effects of different data-processing choices. Furthermore, we chose to exclude clusters smaller than 1% of the data, under the assumption that these are likely to represent methodological artifacts or outliers rather than true cluster structure in the data. Arguably, other approaches to such 'nuisance

clusters' could have been equally valid. The same goes for the way we chose to estimate the model results under the null hypothesis for the real datasets.

Conclusion

Clustering methods are important statistical techniques for psychiatric science to improve mental health care by identifying more homogeneous and biologically informed diagnostic categories. This study used SCA to investigate data-driven biological subtypes of MD and showed that results of cluster analyses were heavily dependent on different model specifications. SCA can help to investigate robustness of cluster analyses and identify stable clusters. As such, SCA is a useful technique that could aid the development of robust and replicable subtyping models in psychiatric disorders.

References

1. Kendell, R. & Jablensky, A. Distinguishing between the validity and utility of psychiatric diagnoses. *Am. J. Psychiatry* **160**, 4–12 (2003).
2. Kapur, S., Phillips, a G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).
3. Monroe, S. M. & Anderson, S. F. Depression: The Shroud of Heterogeneity. *Curr. Dir. Psychol. Sci.* **24**, 227–231 (2015).
4. Georgiades, S., Szatmari, P. & Boyle, M. Importance of studying heterogeneity in autism. *Neuropsychiatry* **3**, 123–125 (2013).
5. Picardi, A. *et al.* Heterogeneity and symptom structure of schizophrenia. *Psychiatry Res.* **198**, 386–394 (2012).
6. Volavka, J. & Citrome, L. Oral antipsychotics for the treatment of schizophrenia: heterogeneity in efficacy and tolerability should drive decision-making. *Expert Opin. Pharmacother.* **10**, 1917–1928 (2009).
7. Kofler, M. J. *et al.* Heterogeneity in ADHD: Neurocognitive predictors of peer, family, and academic functioning. *Child Neuropsychol.* **23**, 733–759 (2017).
8. Ozomaro, U., Wahlestedt, C. & Nemeroff, C. B. Personalized medicine in psychiatry: problems and promises. *BMC Med.* **11**, 132 (2013).
9. e Silva, J. A. C. Personalized medicine in psychiatry: new technologies and approaches. *Metabolism* **62**, S40–S44 (2013).
10. Kendler, K. S. An historical framework for psychiatric nosology. *Psychol. Med.* **39**, 1935–1941 (2009).
11. Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J. & Beckmann, C. F. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 433–447 (2016).
12. Kendell, R. E. Clinical validity. *Psychol. Med.* **19**, 44–55 (1989).
13. Van Loo, H. M., De Jonge, P., Romeijn, J.-W., Kessler, R. C. & Schoevers, R. A. Data-driven subtypes of major depressive disorder: a systematic review. *BMC Med.* **10**, 156 (2012).
14. Schnack, H. G. Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophr. Res.* **214**, 34–42 (2017).
15. Librenza-Garcia, D. *et al.* The impact of machine learning techniques in the study of bipolar disorder: A systematic review. *Neurosci. Biobehav. Rev.* **80**, 538–554 (2017).
16. Lombardo, M. V, Lai, M.-C. & Baron-Cohen, S. Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol. Psychiatry* **24**, 1435–1450 (2019).
17. Beijers, L., Wardenaar, K. J., van Loo, H. M. & Schoevers, R. A. Data-driven biological subtypes of depression: systematic review of biological approaches to depression subtyping. *Mol. Psychiatry* **24**, 888–900 (2019).
18. Ahmad, A. & Fröhlich, H. Integrating Heterogeneous omics Data via Statistical Inference and Learning Techniques. *Genomics Comput. Biol.* **2**, e32 (2016).

19. Lin, E. & Hsien-Yuan, L. Machine learning and systems genomicsapproaches for multi-omics data. *Biomark. Res.* **5**, (2017).

20. Lewandowski, K. E., Baker, J. T., McCarthy, J. M., Norris, L. A. & Öngür, D. Reproducibility of Cognitive Profiles in Psychosis Using Cluster Analysis. *J. Int. Neuropsychol. Soc.* **24**, 382–390 (2018).

21. Reser, M. P., Allott, K. A., Killackey, E., Farhall, J. & Cotton, S. M. Exploring cognitive heterogeneity in first-episode psychosis: What cluster analysis can reveal. *Psychiatry Res.* **229**, 819–827 (2015).

22. Chand, G. B. *et al.* T195. Neuroanatomical Heterogeneity of Schizophrenia Quantified via Semi-Supervised Machine Learning Reveals Two Distinct Subtypes: Results From the PHENOM Consortium. *Biol. Psychiatry* **85**, S205–S206 (2019).

23. Mostert, J. C. *et al.* Similar Subgroups Based on Cognitive Performance Parse Heterogeneity in Adults With ADHD and Healthy Controls. *J. Atten. Disord.* **22**, 281–292 (2018).

24. Wolfers, T., Buitelaar, J. K., Beckmann, C. F., Franke, B. & Marquand, A. F. From estimating activation locality to predicting disorder: A review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci. Biobehav. Rev.* **57**, 328–349 (2015).

25. Simmons, J. P., Nelson, L. D. & Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **22**, 1359–1366 (2011).

26. Silberzahn, R. *et al.* Many analysts, one dataset: Making transparent how variations in analytical choices affect results Authors. *Adv. Methods Pract. Psychol. Sci.* **1**, 337–356 (2018).

27. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).

28. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. *Springer Series in Statistics* (Springer, 2011).

29. Hagenaars, J. A. Latent Structure Models with Direct Effects between Indicators: Local Dependence Models. *Sociol. Methods Res.* **16**, 379–405 (1988).

30. Borsboom, D. *et al.* Kinds versus continua: A review of psychometric approaches to uncover the structure of psychiatric constructs. *Psychol. Med.* **45**, 1567–1579 (2016).

31. van Loo, H. M., Wanders, R. B. K., Wardenaar, K. J. & Fried, E. I. Problems with latent class analysis to detect data-driven subtypes of depression. *Mol. Psychiatry* **23**, 495–496 (2016).

32. Hands, S. & Everitt, B. A Monte Carlo Study of the Recovery of Cluster Structure in Binary Data by Hierarchical Clustering Techniques. *Multivariate Behav. Res.* **22**, 235–243 (1987).

33. Ferreira, L. & Hitchcock, D. B. A comparison of hierarchical methods for clustering functional data. *Commun. Stat. Simul. Comput.* **38**, 1925–1949 (2009).

34. Saraçli, S., Doğan, N. & Doğan, I. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J. Inequalities Appl.* **203**, (2013).

35. Clifford, H., Wessely, F., Pendurthi, S. & Emes, R. D. Comparison of clustering methods for investigation of genome-wide methylation array data. *Front. Genet.* **2**, 1–11 (2011).

36. Milligan, G. W. & Cooper, M. C. An examination of procedures for determining the number of clusters in a data set. *Psychometrika* **50**, 159–179 (1985).

37. Islam, A., Alizadeh, B. Z., van den Heuvel, E. R. & GROUP investigators. A comparison of indices for identifying the number of clusters in hierarchical clustering: A study on cognition in schizophrenia patients. *Commun. Stat. Case Stud. Data Anal. Appl.* **1**, 98–133 (2015).

38. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve: Descriptive and inferential statistics on all reasonable specifications. *Nat. Hum. Behav.* **4**, 1208–1214 (2020).

39. Penninx, B. W. J. H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17**, 121–140 (2008).

40. Beijers, L. *et al.* Biomarker-based subtyping of depression and anxiety disorders using Latent Class Analysis. A NESDA study. *Psychol. Med.* **49**, 617–627 (2019).

41. Bot, M. *et al.* Serum proteomic profiling of major depressive disorder. *Transl. Psychiatry* **5**, e599 (2015).

42. Johnson, W., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

43. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. **NbClust** : An *R* Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **61**, (2014).

44. Jain, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **31**, 651–666 (2010).

45. John, C. R. *et al.* M3C: A Monte Carlo reference-based consensus clustering algorithm. *Sci. Rep.* **10**, 1816 (2018).

46. Bloom, J. & Al-Abed, Y. MIF: Mood Improving/Inhibiting Factor? *J. Neuroinflammation* **11**, 11 (2014).

47. Maes, M. *et al.* Increased serum IL-6 and IL-1 receptor antagonist concentrations in major depression and treatment resistant depression. *Cytokine* **9**, 853–858 (1997).

48. Milaneschi, Y. *et al.* Interleukin-1 Receptor Antagonist and Incident Depressive Symptoms Over 6 Years in Older Persons: The InCHIANTI Study. *Biol. Psychiatry* **65**, 973–978 (2009).

49. Domenici, E. *et al.* Plasma protein biomarkers for depression and schizophrenia by multi analyte profiling of case-control collections. *PLoS One* **5**, e9166 (2010).

50. J.M., M. *et al.* Genetic predictors of depressive symptoms in cardiac patients. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **105B**, 381–388 (2009).

51. Batterham, R. L. *et al.* Pancreatic polypeptide reduces appetite and food intake in humans. *J. Clin. Endocrinol. Metab.* **88**, 3989–3992 (2003).

52. Powell, T. R. *et al.* Putative transcriptomic biomarkers in the inflammatory cytokine pathway differentiate major depressive disorder patients from control subjects and bipolar disorder patients. *PLoS One* **9**, e91076 (2014).

53. Hasler, G. & Northoff, G. Discovering imaging endophenotypes for major depression. *Mol. Psychiatry* **16**, 604–619 (2011).

54. Flint, J. & Kendler, K. S. The Genetics of Major Depression. *Neuron* **81**, 484–503 (2014).

55. Wanders, R. B. K. *et al.* Casting wider nets for anxiety and depression: disability-driven cross-diagnostic subtypes in a large cohort. *Psychol. Med.* **46**, 3371–3382 (2016).

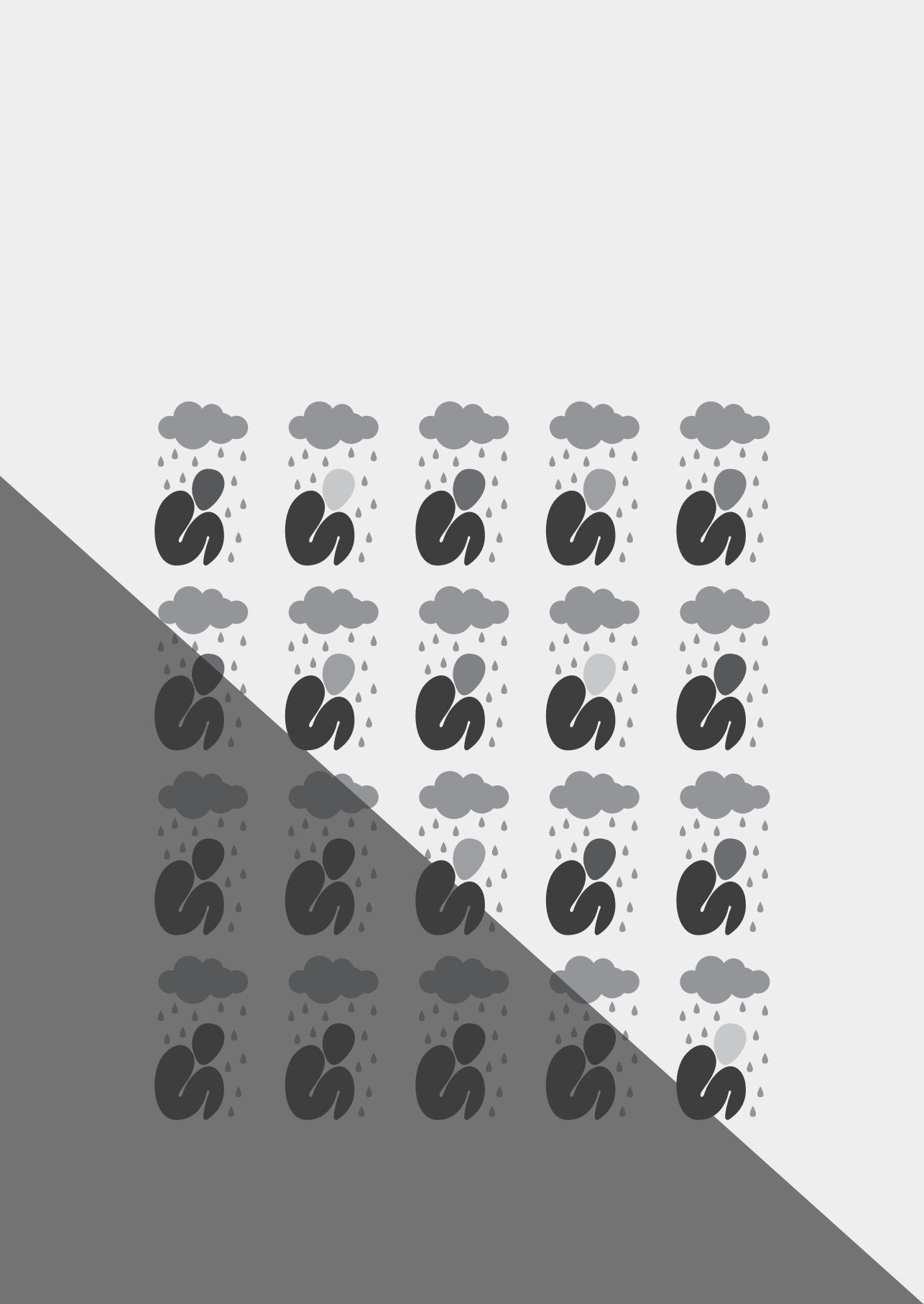
56. Wardenaar, K. J., Wanders, R. B. K., ten Have, M., de Graaf, R. & de Jonge, P. Using a hybrid subtyping model to capture patterns and dimensionality of depressive and anxiety symptomatology in the general population. *J. Affect. Disord.* **215**, 125–134 (2017).

57. Islam, M. A. *et al.* Long-term cognitive trajectories and heterogeneity in patients with schizophrenia and their unaffected siblings. *Acta Psychiatr. Scand.* **138**, 591–604 (2018).

58. Rohrer, J. M., Egloff, B. & Schmukle, S. C. Probing Birth-Order Effects on Narrow Traits Using Specification-Curve Analysis. *Psychol. Sci.* **28**, 1821–1832 (2017).

59. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**, 173–182 (2019).

60. Milligan, G. W. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* **45**, 325–342 (1980).



Chapter 7

General discussion

Preface

The aims of this dissertation were to gain more insight into the etiology of MD by (1) using rich datasets and novel statistical methodology to take a more detailed look at risk factors of MD and (2) to investigate if and how well bottom-up subtyping approaches might enable the discovery of more homogeneous subtypes of MD. I will first summarize the main findings and their implications for both aims separately. Subsequently, I discuss the strengths and limitations of my contributions to the literature. Then, based on my main findings and the challenges I encountered during the writing of this dissertation, I will formulate starting-points for future research into the etiology and pathophysiology of MD, including possible directions for research into bottom-up data-driven subtypes of MD.

Main findings

The first part of this dissertation investigated the risk factors of MD using rich datasets and novel statistical methodology. In **Chapter 2**, Relative Importance Analysis was performed using the Lifelines cohort, a large longitudinal population study, in order to identify key risk factors of MD onset and recurrence over a 6-year period. We included 21 risk factors that have previously been found to be related to MD, such as socio-demographic variables, neuroticism, family history, stressful life events, childhood trauma, health behaviors, general health status, and metabolic and inflammatory markers. A family history of anxiety and depression, childhood trauma, higher neuroticism, female sex, younger age, long-term difficulties, lower physical quality of life and current anxiety disorders were all important predictors of MD onset, but family history had the largest effect. Most risk factors for MD onset also predicted MD recurrence, but a higher number of anxiety disorders at baseline predicted first onset only. Lower education levels did not predict onset, but they were the largest predictor of recurrence. Risk of recurrence was highest in men. In **Chapter 3**, cross-sectional data from the same sample was used to take a closer look at the relationship between sex, age and internalizing disorders (i.e., MD, dysthymia, generalized anxiety disorder (GAD), social phobia and panic disorder), symptoms of depression and anxiety, negative affect (NA), and neuroticism. Generalized additive models were used to identify nonlinear patterns of these internalizing disorders, symptoms, and traits over the participants' lifetimes, and to investigate sex differences. Women reported more internalizing disorders, symptoms and traits than men, but the

relative difference remained stable across age (relative risk ~ 1.7). The prevalence of internalizing disorders generally increased between the ages of 18-30 years, stabilized between 30-50, and decreased after age 50 again. Internalizing symptoms and traits showed different patterns.

The second part of this dissertation focused on methodological and empirical questions about bottom-up MD subtyping. **Chapter 4** presented the results of a systematic review of current evidence available for data-driven biological subtypes of MD from studies that identified (1) data-driven subtypes of MD based on biological variables (i.e., biochemical, neurological, or genetic data), or (2) data-driven subtypes based on clinical features such as symptom patterns and validated these with biological variables post-hoc, in order to gain insight into existing knowledge about the role of biological factors in MD heterogeneity. Twenty-nine publications including 24 separate analyses in 20 unique samples were identified, including a total of ~4000 subjects. Five out of six biochemical studies indicated that there might be depression subtypes with and without disturbed neurotransmitter levels, and one indicated there might be an inflammatory subtype. Seven symptom-based studies identified subtypes, which were mainly determined by severity and by weight gain vs. loss. Two studies compared subtypes based on medication response. These symptom-based subtypes were associated with differences in biomarker profiles and functional connectivity, but results have not sufficiently been replicated. Four out of five neuroimaging studies found evidence for groups with structural and connectivity differences, but the methods and results were inconsistent. The single genetic study found a subtype with a distinct pattern of SNPs, but this subtype has not been replicated in an independent test sample. One study combining all aforementioned types of data discovered subtypes with different levels of functional connectivity, childhood abuse, and treatment response, but the sample size was small. **Chapter 5** showed that it is possible to successfully apply data-driven clustering techniques, which are commonly used in studies based on clinical data, to a set of biochemical biomarkers. Latent Class Analysis was performed on data from the Netherlands Study of Depression and Anxiety, a large, multi-site naturalistic cohort study. Thirty-six biomarkers (e.g., leptin, brain-derived neurotrophic factor, tryptophan) were used as input variables. Once estimated, latent classes were compared on sociodemographic and clinical characteristics. The analyses resulted in three classes, which were primarily characterized by different levels of metabolic health and were labeled as: 'lean' (21.6%), 'average' (62.2%) and 'overweight' (16.2%). The identified classes were strongly related to general (metabolic) health and did not reflect any traditional diagnostic classifications such as generalized anxiety disorder vs. MD or the metabolic subtype vs. the atypical subtype of MD. In **Chapter 6,** Specification-Curve

Analysis (SCA) was performed using 31 proteomic biomarkers that had previously been found to be related to MD in the NESDA sample, in order to gain more insight into the influence of methodological variation on biomarker-based cluster-analysis results. This analysis evaluated the consistency of the model results across 1,200 k-means and hierarchical clustering analyses, each with a unique model specification (i.e., combination of clustering algorithm, fit-index, and distance metric). The results were inconsistent, meaning that no robust patterns of biological clustering were discovered in either the MD only sample or the combined MD/healthy dataset. Next, SCAs were run in simulated datasets with known cluster numbers and noise/outlier levels to evaluate the effect of data properties on SCA outcomes. The simulation results showed that the correct number of clusters could be identified quite consistently across the 1,200 model specifications used, but that correct cluster identification became harder when number of clusters and noise levels increased.

Implications

A closer look at risk factors for MD

The first part of this dissertation was intended to refine our understanding of MD etiology by taking both a broad perspective, looking at the relative importance of a large group of risk factors for depression assessed at the same moment in time, and a deep perspective, by zooming in further on the relationship between sex, age, and MD over the lifecourse.

Chapter 2 studied key risk factors relevant for population screening to identify subjects at risk of onset and recurrence of MD. For example, screening for MD among family members of depressed individuals may lead to more timely interventions. The importance of lower education levels as a predictor for recurrence of MD in this chapter suggests that awareness of the effect of educational inequality in MD is needed especially in relation to the course of the disorder.

Chapter 3, which used a cross-sectional approach to investigate the relationship between sex, age and internalizing disorder prevalence, provides an interesting addition to the prospective perspective on life-course epidemiology of MD from Chapter 2. Since different patterns of incidence, chronicity, and recurrence can lead to the same patterns of prevalence, the results of both chapters cannot be compared directly, but some general inferences can be made. Chapter 2 assumed that there is a linear relationship between age and onset and recurrence of MD, and found that MD risk in both samples decreased with age, although the decrease was about twice as strong for onset. This is in line with a review

that showed that the development of the point prevalence of MD over the lifetime follows an inversed U-shape, because the mathematical derivative of this shape is an decreasing line.[1] Still, since this review is based on models that are unable to identify more complex non-linear patterns, it is possible that the assumption of linearity in Chapter 2 was an oversimplification. Indeed, Chapter 3 used generalized additive models to show that the pattern of MD prevalence over age is more intricate than this.

Chapter 3 suggested similar curves of MD prevalence across the lifespan for men and women, which is in accordance with the aforementioned review.[1] This means that men and women have the same rates of incidence, chronicity and recurrence, or that if these rates differ between the sexes, that their average result is similar. Still, Chapter 3 also showed a sizable gender gap in MD prevalence. Previous studies show this gap arises in puberty, due to higher incidence rates in women.[2–4] This sex difference in incidence likely did not show up in the results of Chapter 3 because this study was based on an adult sample. Chapter 2 on the other hand did show that women had a higher risk of onset of MD, but the study was not designed to include interaction effects between sex and age. Interestingly, this chapter also showed that men were more likely to suffer from recurrent episodes, unlike previous population studies, which often showed no effect of sex on MD recurrence.[5–11] This could be due to a lack of power, since these studies generally had smaller sample sizes. Although the effects of sex on onset and recurrence of MD pointed in opposite directions, the findings of Chapter 2 are in line with the results of Chapter 3, because the effect size was similar, meaning that averaging over these effects would result in a similar change of prevalence in men and women.

Are there robust biological subtypes of MD?

Although all of the studies reviewed in Chapter 4 provided interesting leads for future research, the methodological differences across studies and lack of replication precluded definitive conclusions about the existence of clinically useful and generalizable biological subtypes of MD. One type of data that seems to hold some promise for data-driven biological subtypes of MD is data related to immunometabolic dysregulations. Four early symptom-based subtyping studies reviewed in Chapter 4 identified a severe depression subtype that showed some overlap with the DSM melancholic specifier and had lower L-TRP scores and more dysregulation of the stress and metabolic systems.[12–16] Two later studies used data from the NESDA sample to identify two severe subtypes, characterized either by appetite/weight loss or appetite/weight gain, which mainly differed on biomarkers that are related to weight, metabolism, inflammation, and stress.[17–23] Based on these results, researchers have postulated that some patients might suffer from an immunometabolic

subtype of MD. Patients with immunometabolic depression (IMD) are thought to suffer from altered energy intake/expenditure balance, which is why they show more atypical behavioral symptoms (i.e., hyperphagia, weight gain, hypersomnia, fatigue, and leaden paralysis).[24,25] Therefore, IMD patients would be expected to respond better to novel and/or alternative therapeutic approaches such as exercise and anti-inflammatory drugs, compared to patients who do show immunometabolic dysregulations.[25–27] The relationship between immunometabolic dysregulations and treatment response has been investigated a number of studies, but the majority of these have focused on comparing different types of traditional medication (e.g., tricyclic antidepressants vs. selective serotonin re-uptake inhibitors).[25]

The findings from this dissertation with regards to a putative immunometabolic subtype of MD are mixed. For example, in Chapter 3, (auto)immune diseases, low-grade inflammation, and the metabolic syndrome had little predictive value for either incidence or recurrence of MD. However, it is possible that this a result of averaging over people with and without immunometabolic dysregulations. It is also possible that this is due to the fact that whereas the aforementioned studies looked at cross-sectional associations between variables, Chapter 3 was designed to look at longitudinal associations between risk factors and MD. However, Chapter 5 did not show a clear IMD subtype either, even though the cluster analysis was based on a cross-sectional collection of mostly metabolic biomarker. In fact, there were no differences in individual symptoms of anxiety and depression like increased appetite or psychomotor retardation between the classes in Chapter 5. Still, the fact that the subjects with psychopathology were relatively likely to be in the 'overweight' class compared with healthy controls, indicates a possible connection between depressive and anxiety disorders and the overweight biomarker profile, and the overweight class showed a higher percentage of atypical specifiers compared to the lean class. Furthermore, when comparing the classes from Chapter 5 to the symptom-based subtypes from Lamers et al. (2010), patients from the 'severe atypical' class were less likely to be in the lean class and more likely to be in the average class compared to the 'severe melancholic' and the 'moderate' classes.

Together with previous reviews these results indicate that although there is some overlap between data-driven subtypes based on symptoms and those based on metabolic markers, atypical depression defined according to DSM criteria does not appear to be an effective stratification criterion.[28–30] Whether another specifier with a modified symptom profile may provide different results is unknown and needs to be properly tested in dedicated studies.[25] Alternatively, the final IMD classification might include a combination of symptoms, clinical characteristics such as non-response to conventional

treatment and a history of childhood maltreatment, and markers related to, for example, low-grade inflammation (e.g., c-reactive protein, interleukin-6) or insulin resistance (e.g., triglyceride to high-density lipoprotein-cholesterol ratio, Quantitative Insulin Sensitivity Check Index).[27,31–33]

It is important to note that the clinical features of IMD are unlikely to be limited to patients diagnosed with MD. Symptoms such as hyperphagia, weight gain, hypersomnia, fatigue, and leaden paralysis might also be present in patients diagnosed with psychiatric disorders that show overlap with MD, such as anxiety, bipolar, or psychotic disorders, as well as in patients who primarily present with somatic conditions such as diabetes, cardiovascular disease, pain, osteoarthritis, or neurodegenerative diseases.[25] Since this subtype is likely transdiagnostic, future research into the IMD specifier should ideally include a wide range of psychiatric and somatic patients. Furthermore, the association between immunometabolic dysregulations and psychopathology seen in IMD might also involve multiple other types of dysregulations such as oxidative stress or disturbed mitochondrial biogenesis, so a wide range of potential biomarkers needs to be included in future research in order to be able to investigate the pathophysiological characteristics of this subtype.[34]

The influence of methodology on cluster results

One of the reasons no consistent evidence has been found for the IMD subtype or other biological subtypes of MD is likely to be the considerable amount of methodological variation present in the previous literature. Chapter 4 aimed to evaluate if consistent biological distinctions can be made between subtypes of MD, but the varying sample sizes, the methodological differences across studies, and the lack of replication precluded definitive conclusions about the existence of clinically useful and generalizable biological subtypes. For example, the older symptom-based subtyping studies[12–16] generally had smaller sample sizes compared to the newer studies[17–23], and used parametric mixture methods instead of nonparametric clustering methods. The biochemical subtyping studies used similar techniques, but generally had small sample sizes, increasing the potential influence of random error on the results.[35–42] The brain connectivity studies' methods also varied considerably with regards to measurement setting (e.g., rest vs. task), regions of interest used, and analytical and statistical methods, which resulted in a large variation in the model results. The influence of methodological choices on model results also became evident in the genetic study; a subtype was only identified when using a heavily restricted set of selected SNPs. Finally, the study that combined multiple types of data included many controls, as did some other studies.[43–45] This might lead the final model to focus on

the difference between patients and controls, rather than subtypes of MD, especially when separation between these groups is used as a selection criterion.[46]

Chapter 6 aimed to take a closer look at the influence of such methodological variations by evaluating the effects of different model specifications on the number of identified data-driven biological clusters in MD. The cluster-analysis results based on our sample of MD patients were very sensitive to the model specifications used. This means subtyping results of a single analysis should be interpreted with caution until they are proven to be robust to methodological variation. This has already been shown in social psychology, where for example the negative impact of racial bias on callback rates in job application processes has been shown to be robust, whereas increased death toll of female-named hurricanes was not.[47] The simulation results indicated that it is possible to identify robust clusters, but that this becomes more difficult as the data complexity (i.e., number of clusters), levels of noise, and/or the number of outliers increases. Measurement error and a smaller sample size might also make it more difficult to show consistent results. As such, SCA is a useful technique that could aid the development of robust and replicable subtyping models in psychiatric disorders, but a lack of robust clusters in an SCA does not always mean that there are no clusters in the data.

Chapter 6 also showed that while SCA might have a tendency to indicate a lack of cluster structure when there are high levels of noise, many individual cluster analyses have the opposite problem – they will result in a cluster solution even when no structure is present in the data. This is especially troublesome since, as became evident in Chapter 4, there is often a lack of validation of subtyping models against the null hypothesis that clusters are not present in the data. In fact, this was only done explicitly in a single study.[48] It is true that many clustering techniques are not capable of performing this test, but some post-hoc tests exist[49,50], and there are R-packages available that can test if data can be modelled as coming from a single multivariate Gaussian distribution (e.g., SigClust[51]). Comparing the final model to the null hypothesis was also more difficult in some studies because cluster separation was artificially enhanced by, for example, removing edge cases[52] or intermediate clusters[20,53]. In practice, not testing the null hypothesis means that most studies will always result in at least two clusters. Provided there are enough variables to test for differences between the putative clusters, there will always be some significant differences, but that does not mean that the clusters represent relevant subtypes.

In summary, there is a lot of methodological variation in the field of data-driven bottom-up subtyping of psychopathology/MD. Model results that not been evaluated for robustness to methodological variation should be interpreted with caution until replication, and any study that engages in artificially enhancing cluster separation or

that presents two clusters without testing the null hypothesis should be regarded with a healthy skepticism.

Strengths and limitations

Sample

Strengths of both the Lifelines (Chapters 2 and 3) and NESDA (Chapter 5 and 6) cohorts included their longitudinal design, large sample size, including both men and women and a wide age range, the high number of available risk factors, and the presence of thorough assessments with validated structured questionnaires. In NESDA and the first wave of Lifelines, psychiatric disorders were assessed with structured interviews by trained research assistants, and focused on current psychopathology to minimize recall bias.[54–56]

However, a number of limitations regarding the samples that were used need to be taken into account when evaluating the results of this dissertation. For example, although some variables related to inflammation were available, neither sample included data related to all of the main biochemical hypotheses of MD (e.g., serotonin and other neurotransmitters[57], components of the hypothalamus-pituitary-adrenal axis[58]).[59] Collecting this kind of data is complicated and costs a lot of resources, which means that datasets that include all of these biochemical markers and sufficient sample size currently do not exist. This means that it is not (yet) possible to investigate if there are biochemical subtypes of MD related to the main biochemical hypotheses of MD in the same study, which despite the large set of available variables somewhat limited the scope of Chapters 5 and 6.

Furthermore, although both samples have a longitudinal design, Chapters 3 and 6 used cross-sectional data only. In Chapter 4, the biological data needed to estimate the subtypes was not available at follow-up, making it impossible to investigate subtype stability over time and the effects of subtype changes over time. Furthermore, the DSM criterion of disability due to MD or GAD was not assessed for MD and GAD in Lifelines, in NESDA the disability variables were not included in the diagnosis calculations. In Lifelines, dysthymia was not assessed in subjects who satisfied criteria for MD in Lifelines, which could have biased prevalence rates upwards and downwards, respectively. However, given that our estimates of MD, GAD, and dysthymia are comparable to previous estimates, these biases are likely minor.[1,60–62]

Finally, both NESDA and Lifelines are subject to a number of limitations that are common to longitudinal cohort studies.[63–66] For example, it is possible that the NESDA and Lifelines samples are subject to selection bias, which occurs when older individuals with MD are relatively less often participating in population studies than younger individuals with MD due to higher morbidity and mortality, difficulty in establishing contact or increased refusals.[67–69] However, in Lifelines, we found no interaction effect between age and the presence of an internalizing disorder at baseline when predicting participation at follow up (2014-2017, data not shown). This means that the impact of having an internalizing disorder on attrition did not differ between older and younger subjects, so selection bias is also not a likely explanation for the reduction in prevalence of internalizing pathology after age 50 observed in Chapter 3. Additionally, there might have been selective attrition in either sample due to MD or other factors such as higher age or lower education levels.[63–65,70–72] This might have led to an underestimation of the incidence and recurrence rates, which was mainly a limitation in Chapter 2. It is difficult to ascertain the effects of selective attrition on our analyses in this chapter, because we cannot be sure which subjects dropped out due to developing MD after baseline.

Variable selection and data processing

The data-driven approach of this dissertation should be considered a strength, since one important advantage of data-driven subtyping is that it is less sensitive to human bias compared to clinically defined subtypes.[73] However, it is important to keep in mind that although the algorithms have no opinions and expectations with regards to what they are going to find, the model results are only ever going to contain the data that was provided to the algorithm. This means that the opinions and preferences of the researcher can still affect the potential model outcomes by influencing the methodological choices made during a data-driven subtyping study. One of the strengths of this dissertation is the variety of approaches to variable selection. Each approach has the potential to tell us something interesting about MD, and a certain robustness to methodological variation could be discovered if multiple approaches point in the same direction. However, each approach also comes with its own set of strengths and limitations.

In Chapter 5, all available biochemical variables were included, which could be considered a strength, since it is means the variable selection is relatively unbiased. However, it is important to remember that this does not mean variable selection is completely without bias, but rather that the choice of what to include is dependent on the state of knowledge, budgetary aspects, and possibly also specific interests of researchers at the moment a study is designed. Furthermore, another limitation of this approach is that

most biochemical variables show variation in the general population, but the majority of this variation is not necessarily related to MD. This means that including all available data leads to increased risks of finding clusters that are not MD-specific. For example, a cluster might be identified that has low amounts of one or more vitamins, but although these people will be at higher risk for different somatic health issues, such clusters might be uninformative with regards to MD. Still, the fact that subtypes also exist in the general population does not necessarily render the resulting subtypes uninformative, as these different clusters might still represent subgroups with dysregulations in different etiological pathways that result in differences in MD incidence, course or treatment response. For example, in Chapter 5, the people in the cluster with metabolic issues were more likely to be depressed.

The variable selection in Chapter 6 was more hypothesis-driven. A total of 31 variables were included, all of which had previously been shown to differ between patients with current MD and healthy controls using adjusted linear regression.[23] This approach of including a limited set of variables known to be related to MD could be considered a strength, since it increases the chances of finding MD-specific clusters.[74–76] However, since there might be different subtypes that have either high or low values on the same variable, the average relationship between this variable and MD in the total sample might approach zero, meaning this selection criterion may potentially lead to exclusion of interesting variables that can differentiate between different subtypes, which is a limitation. Indeed, the range of available markers in NESDA is much larger (> 250), so the fact that no meaningful subtypes of MD were identified in Chapter 6 might indicate that some of the excluded markers would have been more suitable for finding clusters among MD patients.

One major limitation to both of these approaches is that it is possible that other biochemical markers that are not included in NESDA would have been more informative (e.g., related to the main biochemical hypotheses of MD: serotonin and other neurotransmitters[57], components of the hypothalamus-pituitary-adrenal axis[58], inflammation[32]). Furthermore, as became apparent from Chapter 4, it is not at all certain that biochemical data is the best choice when it comes to discovering biological subtypes of depression.[77] It could also be that genetic background or neuroimaging variables related to brain structure or functional connectivity are more suitable for clustering MD patients.[78,79] Ideally, all of these variables would be combined in a single dataset suitable for cluster analysis. Unfortunately, efforts to integrate clinical, genetic, biochemical, and neuroimaging data are logistically complicated. This means that a lack of large datasets that

include data from all of these levels can be a major obstacle when it comes to investigating data-driven subtypes of MD.[80,81]

Methodology

One major strength of this dissertation is that it was able to investigate previously unexplored questions with regards to the etiology and heterogeneity of MD by using novel and advanced statistical techniques. For example, Chapter 2 describes the first study to apply relative importance analyses to identify key risk factors for onset and recurrence of MD, although the technique has been used in other fields of medicine (e.g., predicting wound healing[82] or cardiovascular problems[83,84]). Chapter 3 describes the first study that used advanced nonlinear models to investigate the development of internalizing disorders over lifetime in a large sample from the general population. This provided important insights into the relationships of sex, age, and internalizing disorder prevalence, particularly with regards to the hypotheses about hormonal causes of the gender gap in MD prevalence. Chapter 5 describes the first study to ever perform LCA using biochemical markers. Finally, Chapter 6 discusses the first ever application of SCA, a novel technique designed to investigate the influence of methodological variation on model results, to a statistical inquiry (i.e., cluster analysis of psychopathology) outside of its field of origin (i.e., social psychology). SCA has traditionally been used in psychology to investigate the effects of using alternative regression models, but has been adapted in this dissertation to make it suitable for cluster analysis.[47,85,86]

There are also a number of methodological limitations that need to be taken into account when evaluating the results of this dissertation. For example, like the studies reviewed in Chapter 4, none of the studies presented here performed out-of-sample validation. This is largely due to the lack of suitable validation samples with high sample sizes and similar available variables, although the Lifelines sample is large enough to include validation in a hold-out sample. Furthermore, the LCA study in Chapter 5 did not include explicit null hypothesis testing (e.g., bootstrap likelihood ratio test or Lo–Mendell–Rubin test[87]), so it is not certain that the 3-class solution presented in this chapter is a better fit than a model that includes no classes at all. However, the lack of null hypothesis testing becomes a much larger issue when the model selection criteria indicate that there is minimal separation, i.e., the minimum of two number of clusters is the best fit.

In addition, it is important to remember that although the effect of methodological variation on bottom-up data-driven subtyping were investigated in Chapter 6, it is possible that small methodological variations in Chapters 2, 3 and 5 would also have led

to different results. And even in Chapter 6, there are still many sources of variation left out of the SCA. For example, the analysis was limited to a single MD dataset with a restricted set of markers, because the primary focus was on the influence of model specifications on the results and not on the effects of different data-processing choices (i.e., multiverse analysis[88]). One example of potential source of variation related to a data-processing is the way clusters with small numbers of participants, or even singletons, were handled. In Chapter 6, clusters smaller than 1% of the data were excluded, under the assumption that these are likely to represent methodological artifacts or outliers rather than true cluster structure in the data, but other approaches to such 'nuisance clusters' could have been equally valid.

Finally, in line with the scope of this dissertation, potentially suitable methods have been excluded. For example, Chapter 6 includes a limited number of model specifications for unsupervised learning. However, it does include the exhaustive list of options in the NbClust R-package, which focuses on k-means clustering and hierarchical clustering, two of the most commonly used clustering methods.[73] This package was designed to gather all indices available in SAS and R packages together into a single one package, as well as some newer indices that are not implemented anywhere else yet.[89]

Clinical implications

The first part of this dissertation has provided more insight into the etiology of MD, and identified a number of key risk factors that could be used to identify people at higher risk for onset or recurrence of MD. The high recurrence rates indicate that intervention aiming to prevent MD recurrence, such as post-remission CBT or mindfulness-based cognitive therapy, will potentially benefit over a third of the patient population.[90–92] This might warrant efforts to offer such interventions to all patients, but given the required investments, clinical practice will likely benefit from being able to better predict which patients are at highest risk of recurrence. For example, the findings of Chapter 2 indicate that people with lower education levels are more likely to suffer from recurrent episodes. The presence of co-morbid anxiety should be used as a dichotomous indicator, since there was no dose-response effect of the number of anxiety disorders for predicting MD recurrence. Importantly, although the results of Chapters 2 and 3 indicate that prevention of first MD onset is especially important in young women, women are not necessarily at higher risk for recurrence, meaning sex should not be used as a universal indicator of poor MD outcome.

The second part of this dissertation investigated the potential for bottom-up subtyping approaches to enable the discovery of more homogeneous subtypes of MD. The results of Chapter 4 show that although subtyping based on etiology and pathophysiology is a promising research avenue, it is still in its infancy in some ways. This means that no definitive conclusions regarding clinically useful subtypes can be drawn as of yet. Furthermore, no consistent biological subtypes were identified in Chapters 5 and 6, even though we used large datasets and advanced statistical methods. Even if this had been the case, the results of a cluster analysis should not be seen as equivalent or directly translatable to DSM-style subtype classifications for a number of reasons. For example, cluster analyses are not meant to identify broadly applicable and clinically useful categories, but to identify structures in datasets. The results of this dissertation indicate they will almost always provide a solution, but there is no guarantee that the results are always clinically meaningful.[73] Also, the results of cluster analyses can vary considerably depending on the used algorithms, sample populations, and input variables. Thus, results from a single cluster analysis can never provide clinically useful subtypes. Rather, if the results of sufficiently replicated and robust, cluster analysis results could be useful contributions to the total evidence base for new subtypes of MD that can be applied in clinical studies to assess its importance for mental healthcare. Currently, the IMD subtype has the largest evidence base, but these results are in need of external validation and must be evaluated for robustness to methodological variation. Before this classification can be used in clinical practice, additional research is needed to (1) develop tests to diagnose IMD patients, which might include questionnaires to identify specific symptom profiles but also other clinical tests for IMD such as physical measurements (e.g., body mass index) and blood tests (e.g., low-grade inflammation, cholesterol) and (2) determine the resulting subtypes can indeed be used to provide more effective treatments that are focused on the underlying pathophysiology.[25,27]

Future directions

The presented findings have provided more insight into the etiology of MD as well as the potential for bottom-up subtyping approaches to discover more homogeneous subtypes of MD. However, the results of this dissertation also raise some important new questions.

Importantly, the findings of this dissertation with regards to the risk factors of MD are in need of independent replication in other general population studies and clinical populations. Future studies to identify the key risk factors of MD would for example

need to include markers related to pathophysiological mechanisms underlying MD. According to this dissertation, research into selective prevention should give priority to young adolescent women, and people with a family history of MD. Since rearing experiences are thought to contribute just as much to trans-generational transmission of MD risk as genetic risk, it might be interesting to investigate targeted interventions to improve parenting skills in families with a history of MD as a means of reducing MD prevalence.[93–98]

The results related to bottom-up data-driven subtypes of MD are also in need of independent replication. The immunometabolic subtype discussed above is one example, since most of the evidence for IMD is based on NESDA, but the putative subtypes with and without low levels of monoamines identified in Chapter 4 also are in urgent need of further investigation. If it turns out that the moderate effect size of SSRIs in the total population is a result of averaging of effects across both of these subgroups, a sizable proportion of the population of MD patients could be receiving this treatment without actually needing it, which is problematic given the sizable side-effects of this type of medication.[99–102] Should the monoamine and IMD subtypes of MD be validated and proven robust to methodological variation in the future, subsequent research should focus on developing symptom profiles and/or other clinical tests to help identify which patients are suffering from each subtype of MD.

From a methodological perspective, this dissertation provides several important leads for researchers as well as reviewers and editors, who work with any type of data-driven subtyping studies. For example, editors and reviewers should ideally always request at least a post-hoc evaluation of the null hypothesis of no cluster structure, and discourage the use of techniques to artificially enhance cluster separation. Furthermore, any presented model should be evaluated for robustness to methodological variation, or, if that is not possible, interpretation of model results should be done with caution until the models have been proven to be robust. In order to stimulate the use of SCA, future research into bottom-up data-driven subtyping should aim to devise ways to streamline the presented methodology.[47] This would also enable a more detailed quantification of the effects of methodological variation on results from different unsupervised learning methods, which could help identify the methods that are most reliable for a given research question and set of data characteristics. For example, a certain algorithm might always give the same result regardless of the distance metric, where another algorithm does not. In order to make the data selection procedure more objective, and to allow for subjects to belong to multiple clusters, future methodological improvements should also include an increased focus on development of technologies like multiple co-clustering. Multiple co-clustering allows

for multiple subject cluster solutions that are associated with different sets of variables, without limiting the number of groups to which a subject or a variable can belong.[45] For example, a person who belongs to a cluster of subjects with metabolic issues might or might not fall into a cluster with high neuroticism and a history of abuse.

Finally, given the multifactorial nature of MD, research into data-driven subtypes of MD will probably have more potential for scientific breakthroughs if it connects psychological phenotyping with environmental and biological factors involved in the origins of MD. Large scale datasets have included clinical phenotyping, and genetic and neuroimaging data is being added.[103–105] The most pressing gap in the currently available data is arguably the lack of datasets that include biochemical data related to all of the main existing biological hypotheses of MD like monoamine functioning[57], inflammation[32] and HPA-axis functioning[58]. As long as such comprehensive data is lacking, it will not be possible to investigate if the small effect sizes of many of the identified risk factors are due to the fact that individual risks of MD consist of many small effects or if they are a consequence of averaging over subtypes that make up the complete population of MD patients instead.

Conclusion

This dissertation aimed to gain more insight into the etiology of MD, and to investigate if and how well bottom-up subtyping approaches might enable the discovery of more homogeneous subtypes of MD. The first part of this dissertation provided a number of leads for research into potential interventions to prevent first onset and recurrence of MD, as well as some interesting insights into the gender gap of MD. The second part of this dissertation showed that bottom-up data-driven subtyping research is a very complex endeavor that requires elaborate and costly data collection, including many variables from different levels, as well as intricate statistical models that enable the evaluation of the robustness of the model results. Finally, some interesting starting points for future research with regards to bottom-up data-driven subtypes of MD were identified. Specifically, the results of this dissertation indicate that there might be an immunometabolic subtype of MD, although the exact clinical and biochemical characteristics of this subtype still need to be elucidated, and that further research into monoamine subtypes of MD is needed to improve the opportunities for personalized treatment of MD.

References

1. Baxter, A. J. *et al.* Challenging the myth of an 'epidemic' of common mental disorders: trends in the global prevalence of anxiety and depression between 1990 and 2010. *Depress. Anxiety* **31**, 506–516 (2014).
2. Kessler, R. C. Epidemiology of women and depression. *J. Affect. Disord.* **74**, 5–13 (2003).
3. Kuehner, C. Why is depression more commen among women than among men? *The Lancet Psychiatry* **4**, 146–58 (2017).
4. Altemus, M., Sarvaiya, N. & Neill Epperson, C. Sex differences in anxiety and depression clinical perspectives. *Front. Neuroendocrinol.* **35**, 320–330 (2014).
5. Spijker, J. *et al.* Determinants of persistence of major depressive episodes in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *J. Affect. Disord.* **81**, 231–240 (2004).
6. Hardeveld, F. *et al.* Recurrence of major depressive disorder across different treatment settings: Results from the NESDA study. *J. Affect. Disord.* **147**, 225–231 (2013).
7. ten Have, M. *et al.* Recurrence and chronicity of major depressive disorder and their risk indicators in a population cohort. *Acta Psychiatr. Scand.* **137**, 503–515 (2018).
8. Mattisson, C., Bogren, M., Horstmann, V., Munk-Jörgensen, P. & Nettelbladt, P. The long-term course of depressive disorders in the Lundby Study. *Psychol. Med.* **37**, 883–891 (2007).
9. Nöbbelin, L., Bogren, M., Mattisson, C. & Brådvik, L. Risk factors for recurrence in depression in the Lundby population, 1947–1997. *J. Affect. Disord.* **228**, 125–131 (2018).
10. Murphy, J. M., Olivier, D. C., Sobol, A. M., Monson, R. R. & Leighton, A. H. Diagnosis and outcome: depression and anxiety in a general population. *Psychol. Med.* **16**, 117–126 (1986).
11. Hoertel, N. *et al.* A comprehensive model of predictors of persistence and recurrence in adults with major depression: Results from a national 3-year prospective study. *J. Psychiatr. Res.* **95**, 19–27 (2017).
12. Maes, M., Cosyns, P., Maes, L., D'Hondt, P. & Schotte, C. Clinical subtypes of unipolar depression: Part I. A validation of the vital and nonvital clusters. *Psychiatry Res.* **34**, 29–41 (1990).
13. Maes, M., Maes, L., Schotte, C. & Cosyns, P. A clinical and biological validation of the DSM-III melancholia diagnosis in men: Results of pattern recognition methods. *J. Psychiatr. Res.* **26**, 183–196 (1992).
14. Schotte, C. K. W., Maes, M., Cluydts, R. & Cosyns, P. Cluster analytic validation of the DSM melancholic depression. The threshold model: Integration of quantitative and qualitative distinctions between unipolar depressive subtypes. *Psychiatry Res.* **71**, 181–195 (1997).
15. Orsel, S., Karadag, H., Turkcapar, H. & Karaoglan Kahilogullari, A. Diagnosis and classification subtyping of depressive disorders: Comparison of three methods. *Klin. Psikofarmakol. Bul.* **20**, 57–65 (2010).
16. Maes, M. *et al.* Clinical subtypes of unipolar depression: Part III. Quantitative differences in various biological markers between the cluster-analytically generated nonvital and vital depression classes. *Psychiatry Res.* **34**, 59–75 (1990).
17. Milaneschi, Y. *et al.* Polygenic dissection of major depression clinical heterogeneity. *Mol. Psychiatry* **21**, 516–22 (2016).

18. Milaneschi, Y. *et al.* The effect of FTO rs9939609 on major depression differs across MDD subtypes. *Mol. Psychiatry* **19**, 960–962 (2014).

19. Lamers, F. *et al.* Evidence for a differential role of HPA-axis function, inflammation and metabolic syndrome in melancholic versus atypical depression. *Mol. Psychiatry* **18**, 692–699 (2013).

20. Milaneschi, Y., Lamers, F., Bot, M., Drent, M. L. & Penninx, B. W. J. H. Leptin Dysregulation Is Specifically Associated With Major Depression With Atypical Features: Evidence for a Mechanism Connecting Obesity and Depression. *Biol. Psychiatry* **81**, 807–814 (2017).

21. Bus, B. A. A. *et al.* Low serum BDNF levels in depressed patients cannot be attributed to individual depressive symptoms or symptom cluster. *World J. Biol. Psychiatry* **15**, 561–569 (2014).

22. Rodgers, S. *et al.* Serum testosterone levels and symptom-based depression subtypes in men. *Front. Psychiatry* **6**, (2015).

23. Bot, M. *et al.* Serum proteomic profiling of major depressive disorder. *Transl. Psychiatry* **5**, e599 (2015).

24. Penninx, B. W. J. H. Depression and cardiovascular disease: Epidemiological evidence on their linking mechanisms. *Neuroscience and Biobehavioral Reviews* vol. 74 277–286 (2017).

25. Milaneschi, Y., Lamers, F., Berk, M. & Penninx, B. W. J. H. Depression Heterogeneity and Its Biological Underpinnings: Toward Immunometabolic Depression. *Biological Psychiatry* vol. 88 369–380 (2020).

26. Capuron, L., Lasselin, J. & Castanon, N. Role of Adiposity-Driven Inflammation in Depressive Morbidity. *Neuropsychopharmacology* **42**, 115–128 (2017).

27. Ioannou, M. *et al.* Towards precision medicine: What are the stratification hypotheses to identify homogeneous inflammatory subgroups. *Eur. Neuropsychopharmacol.* **45**, 108–121 (2021).

28. Davidson, J. R. T. A history of the concept of atypical depression. *J. Clin. Psychiatry* **68**, 10–15 (2007).

29. Parker, G. B. Atypical Depression: A Valid Subtype? *J Clin Psychiatry* **68**, 18–22 (2007).

30. Thase, M. E. Atypical depression: Useful concept, but it's time to revise the dsm-iv criteria. *Neuropsychopharmacology* **34**, 2633–2641 (2009).

31. Mansur, R. B., Brietzke, E. & McIntyre, R. S. Is there a 'metabolic-mood syndrome'? A review of the relationship between obesity and mood disorders. *Neurosci. Biobehav. Rev.* **52**, 89–104 (2015).

32. Valkanova, V., Ebmeier, K. P. & Allan, C. L. CRP, IL-6 and depression: A systematic review and meta-analysis of longitudinal studies. *J. Affect. Disord.* **150**, 736–744 (2013).

33. Watson, K. T. *et al.* Association of Insulin Resistance with Depression Severity and Remission Status: Defining a Metabolic Endophenotype of Depression. *JAMA Psychiatry* vol. 78 439–441 (2020).

34. Morris, G. & Berk, M. The many roads to mitochondrial dysfunction in neuroimmune and neuropsychiatric disorders. *BMC Med.* **13**, 1–24 (2015).

35. Asberg, M., Bertilsson, L., Tuck, D., Cronholm, B. & Sjöqvist, F. Indoleamine metabolites in the cerebrospinal fluid of depressed patients before and during treatment with nortriptyline. *Clin. Pharmacol. Ther.* **14**, 277–86 (1973).

36. Asberg, M., Thoren, P., Traskman, L., Bertilsson, L. & Ringberger, V. 'Serotonin depression'--a biochemical subgroup within the affective disorders? *Science (80-.).* **191**, 478–480 (1976).

37. Gibbons, R. D. & Davis, J. M. Consistent evidence for a biological subtype of depression characterized by low CSF monoamine levels. *Acta Psychiatr. Scand.* **74**, 8–12 (1986).

38. Westenberg, H. G. M. & Verhoeven, W. M. A. CSF monoamine metabolites in patients and controls: support for a bimodal distribution in major affective disorders. *Acta Psychiatr. Scand.* **78**, 541–549 (1988).

39. Maas, J. W. *et al.* Pre-treatment neurotransmitter metabolites and response to imipramine or amitriptyline treatment. *Psychol Med* **12**, 37–43 (1982).

40. Åsberg, M. *et al.* CSF monoamine metabolites in melancholia. *Acta Psychiatr. Scand.* **69**, 201–219 (1984).

41. Azorin, J. M. *et al.* Plasma levels of 3-methoxy-4-hydroxyphenylglycol in depressed patients compared with normal controls. *Neuropsychobiology* **23**, 18–24 (1990).

42. Davis, J. M. *et al.* Cerebrospinal fluid and urinary biogenic amines in depressed patients and healthy controls. *Arch. Gen. Psychiatry* **45**, 705–717 (1988).

43. Maglanoc, L. A. *et al.* Data-Driven Clustering Reveals a Link Between Symptoms and Functional Brain Connectivity in Depression. *Biol. psychiatry. Cogn. Neurosci. neuroimaging* **4**, 16–26 (2019).

44. Yu, C., Arcos-Burgos, M., Licinio, J. & Wong, M. L. A latent genetic subtype of major depression identified by whole-exome genotyping data in a Mexican-American cohort. *Transl. Psychiatry* **7**, (2017).

45. Tokuda, T. *et al.* Identification of depression subtypes and relevant brain regions using a data-driven approach. *Sci. Rep.* **8**, 14082 (2018).

46. Kapur, S., Phillips, a G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).

47. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve: Descriptive and inferential statistics on all reasonable specifications. *Nat. Hum. Behav.* **4**, 1208–1214 (2020).

48. Feder, S. *et al.* Sample heterogeneity in unipolar depression as assessed by functional connectivity analyses is dominated by general disease effects. *J. Affect. Disord.* **222**, 79–87 (2017).

49. Tibshirani, R., Walther, G. & Hastie, T. Estimating the Number of Data Clusters via the Gap Statistic. *J. R. Stat. Soc.* **63**, 411–423 (2001).

50. Kou, J. Estimating the number of clusters via the GUD statistic. *J. Comput. Graph. Stat.* **23**, 403–417 (2014).

51. Liu, Y., Hayes, D. N., Nobel, A. & Marron, J. S. Statistical significance of clustering for high-dimension, low-sample size data. *J. Am. Stat. Assoc.* **103**, 1281–1293 (2008).

52. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2017).

53. Lamers, F. *et al.* Serum proteomic profiles of depressive subtypes. *Transl. Psychiatry* **6**, e851 (2016).

54. Kruijshaar, M. E. *et al.* Lifetime prevalence estimates of major depression: An indirect estimation method and a quantification of recall bias. *Eur. J. Epidemiol.* **20**, 103–111 (2005).

55. Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M. & Wittchen, H. U. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *Int. J. Methods Psychiatr. Res.* **21**, 169–184 (2012).

56. Penninx, B. W. J. H. *et al.* The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *Int. J. Methods Psychiatr. Res.* **17**, 121–140 (2008).

57. Albert, P. R., Benkelfat, C. & Descarries, L. The neurobiology of depression-revisiting the serotonin hypothesis. I. cellular and molecular mechanisms. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 2378–2381 (2012).

58. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).

59. Jentsch, M. C. *et al.* Biomarker approaches in major depressive disorder evaluated in the context of current hypotheses. *Biomark. Med.* **9**, 277–297 (2015).

60. Charlson, F. J., Ferrari, A. J., Flaxman, A. D. & Whiteford, H. A. The epidemiological modelling of dysthymia: Application for the Global Burden of Disease Study 2010. *J. Affect. Disord.* **151**, 111–120 (2013).

61. Bijl, R. V, Ravelli, A. & van Zessen, G. Prevalence of psychiatric disorder in the general population: Results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Soc Psychiatry Psychiatr Epidemiol* **33**, 587–595 (1998).

62. de Graaf, R., ten Have, M., van Gool, C. & van Dorsselaer, S. Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Soc. Psychiatry Psychiatr. Epidemiol.* **47**, 203–213 (2012).

63. De Graaf, R., Bijl, R. V., Smit, F., Ravelli, A. & Vollebergh, W. A. M. Psychiatric and sociodemographic predictors of attrition in a longitudinal study: The Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Am. J. Epidemiol.* **152**, 1039–1047 (2000).

64. Eaton, W. W., Anthony, J. C., Tepper, S. & Dryman, A. Psychopathology and attirition in the epidemiologic catchment area surveys. *Am. J. Epidemiol.* **135**, 1051–1059 (1992).

65. De Graaf, R., Van Dorsselaer, S., Tuithof, M. & Ten Have, M. Sociodemographic and psychiatric predictors of attrition in a prospective psychiatric epidemiological study among the general population. Result of the Netherlands Mental Health Survey and Incidence Study-2. *Compr. Psychiatry* **54**, 1131–1139 (2013).

66. Eerola, M., Huurre, T. & Aro, H. The problem of attrition in a Finnish longitudinal survey on depression. *Eur. Jouranl Epidemiol.* **20**, 113–120 (2005).

67. Holwerda, T. J. *et al.* The relationship between generalized anxiety disorder, depression and mortality in old age. *Int. J. Geriatr. Psychiatry* **22**, 241–249 (2007).

68. Schoevers, R. A. *et al.* Association of depression and gender with mortality in old age: Results from the Amsterdam Study of the Elderly (AMSTEL). *Br. J. Psychiatry* **177**, 366–342 (2000).

69. Beekman, A. T. F. *et al.* The natural history of late-life depression: A 6-year prospective study in the community. *Arch. Gen. Psychiatry* **59**, 605–611 (2002).

70. Marcellus, L. Are we missing anything? Pursuing research on attrition. *Can. J. Nurs. Res.* **36**, 82–98 (2004).

71. Lamers, F. *et al.* Sociodemographic and psychiatric determinants of attrition in the Netherlands Study of Depression and Anxiety (NESDA). *Compr. Psychiatry* **53**, 63–70 (2012).

72. Chatfield, M. D., Brayne, C. E. & Matthews, F. E. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J. Clin. Epidemiol.* **58**, 13–19 (2005).

73. Marquand, A. F., Wolfers, T., Mennes, M., Buitelaar, J. & Beckmann, C. F. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 433–447 (2016).

74. Moffitt, T. E. & Caspi, A. Gene–environment interactions in psychiatry: joining forces with neuroscience. *Nat. Rev. Neurosci.* **7**, 583–590 (2006).

75. Jaffee, S. R. & Price, T. S. Gene-environment correlations: A review of the evidence and implications for prevention of mental illness. *Mol. Psychiatry* **12**, 432–442 (2007).

76. Molenaar, P. C. M. On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Dev. Psychobiol.* **50**, 60–69 (2008).

77. Beijers, L., Wardenaar, K. J., van Loo, H. M. & Schoevers, R. A. Data-driven biological subtypes of depression: systematic review of biological approaches to depression subtyping. *Mol. Psychiatry* **24**, 888–900 (2019).

78. Flint, J. & Kendler, K. S. The Genetics of Major Depression. *Neuron* **81**, 484–503 (2014).

79. Hasler, G. & Northoff, G. Discovering imaging endophenotypes for major depression. *Mol. Psychiatry* **16**, 604–619 (2011).

80. Buch, A. M. & Liston, C. Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. *Neuropsychopharmacology* **46**, 156–175 (2021).

81. Lynch, C. J., Gunning, F. M. & Liston, C. Causes and Consequences of Diagnostic Heterogeneity in Depression: Paths to Discovering Novel Biological Depression Subtypes. *Biological Psychiatry* vol. 88 83–94 (2020).

82. Tonidandel, S. & LeBreton, J. M. Relative Importance Analysis: A Useful Supplement to Regression Analysis. *J. Bus. Psychol.* **26**, 1–9 (2011).

83. Hayek, S. S. *et al.* Depression and chest pain in patients with coronary artery disease. *Int. J. Cardiol.* **230**, 420–426 (2017).

84. Moon, I. *et al.* Association of arterial stiffness with aortic calcification and tortuosity. *Med. (United States)* **98**, (2019).

85. Rohrer, J. M., Egloff, B. & Schmukle, S. C. Probing Birth-Order Effects on Narrow Traits Using Specification-Curve Analysis. *Psychol. Sci.* **28**, 1821–1832 (2017).

86. Orben, A. & Przybylski, A. K. The association between adolescent well-being and digital technology use. *Nat. Hum. Behav.* **3**, 173–182 (2019).

87. Nylund, K. L., Asparouhov, T. & Muthén, B. O. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Struct. Equ. Model.* **14**, 535–569 (2007).

88. Steegen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. Increasing Transparency Through a Multiverse Analysis. *Perspect. Psychol. Sci.* **11**, 702–712 (2016).

89. Charrad, M., Ghazzali, N., Boiteau, V. & Niknafs, A. **NbClust** : An *R* Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **61**, (2014).

90. Tickell, A. *et al.* Recovery from recurrent depression with mindfulness-based cognitive therapy and antidepressants: A qualitative study with illustrative case studies. *BMJ Open* **10**, e033892 (2020).

91. Biesheuvel-Leliefeld, K. E. M. *et al.* Effectiveness of psychological interventions in preventing recurrence of depressive disorder: Meta-analysis and meta-regression. *Journal of Affective Disorders* vol. 174 400–410 (2015).

92. Bockting, C. L., Hollon, S. D., Jarrett, R. B., Kuyken, W. & Dobson, K. A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. *Clinical Psychology Review* vol. 41 16–26 (2015).

93. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. The Rearing Environment and Risk for Major Depression: A Swedish National High-Risk Home-Reared and Adopted-Away Co-Sibling Control Study. *Am. J. Psychiatry* **177**, 447–453 (2020).

94. Kendler, K. S., Ohlsson, H., Lichtenstein, P., Sundquist, J. & Sundquist, K. The Nature of the Shared Environment. *Behav. Genet.* **49**, 1–10 (2019).

95. Kendler, K. S., Ohlsson, H., Sundquist, K. & Sundquist, J. Sources of parent-offspring resemblance for major depression in a national Swedish extended adoption study. *JAMA Psychiatry* **75**, 194–200 (2018).

96. Weissman, M. M. Is Depression Nature or Nurture? Yes. *American Journal of Psychiatry* vol. 177 376–377 (2020).

97. Ormel, J., Cuijpers, P., Jorm, A. F. & Schoevers, R. Prevention of depression will only succeed when it is structurally embedded and targets big determinants. *World Psychiatry* **18**, 111 (2019).

98. Agerbo, E. *et al.* Risk of Early-Onset Depression Associated With Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. *JAMA psychiatry* **75**, 387–397 (2021).

99. Ferguson, J. M. SSRI antidepressant medications: Adverse effects and tolerability. *Prim. Care Companion J. Clin. Psychiatry* **3**, 22–27 (2001).

100. Bala, A., Nguyen, H. M. T. & Hellstrom, W. J. G. Post-SSRI Sexual Dysfunction: A Literature Review. *Sexual Medicine Reviews* vol. 6 29–34 (2018).

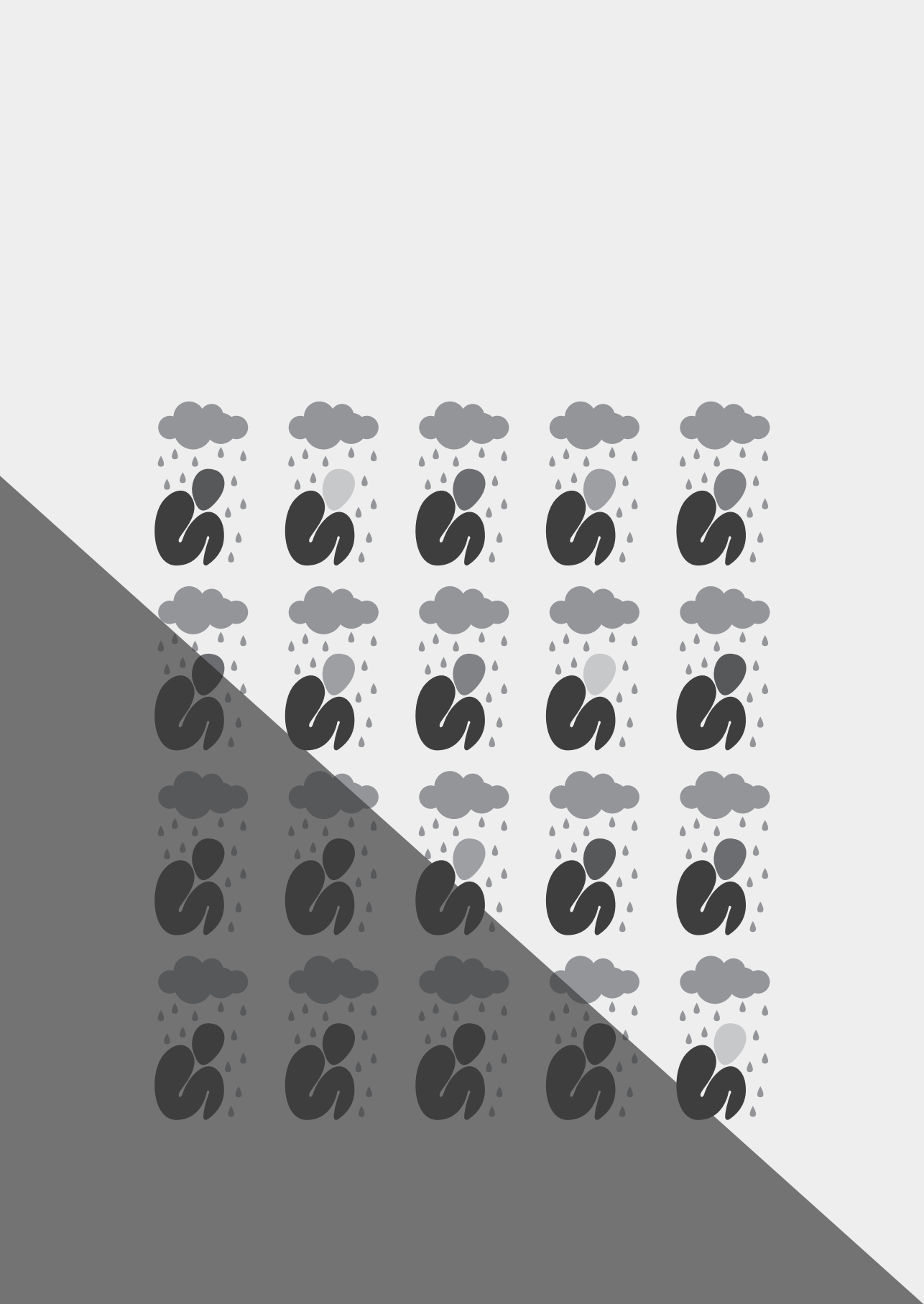
101. Renoir, T. Selective serotonin reuptake inhibitor antidepressant treatment discontinuation syndrome: A review of the clinical evidence and the possible mechanisms involved. *Frontiers in Pharmacology* vol. 4 APR 45 (2013).

102. Horowitz, M. A. & Taylor, D. Tapering of SSRI treatment to mitigate withdrawal symptoms. *The Lancet Psychiatry* vol. 6 538–546 (2019).

103. Smith, D. J. *et al.* Prevalence and characteristics of probable major depression and bipolar disorder within UK Biobank: Cross-sectional study of 172,751 participants. *PLoS One* **8**, (2013).

104. Van Loo, H. M. *et al.* Prevalence of internalizing disorders, symptoms, and traits across age using advanced nonlinear models. *Psychol. Med.* 1–10 (2021)

105. Frässle, S. *et al.* Predicting individual clinical trajectories of depression with generative embedding. *NeuroImage. Clin.* **26**, 102213 (2020).



General summary

Major Depression (MD) is a common and debilitating disorder, which currently makes it the single largest contributor to the global burden of disease. Unfortunately, most treatments for MD have moderate effect sizes in clinical trials. Better understanding of the causal mechanisms of MD is expected to facilitate the development of biologically informed, patient-specific diagnoses, which in turn should enable psychiatrists to provide treatments that are tailored to a patients' etiological and pathophysiological background.

Since the effectiveness of current therapies relative to placebo is modest, other approaches might be needed to address the public health burden of MD, such as preventive interventions for both first onsets and recurrent episodes of MD. Identifying key risk factors for MD could help provide focal points for these interventions, but relative importance analyses have not been performed before due to a lack of datasets including enough participants and sufficient numbers of risk factors. Also, most studies report models that are easier to calculate like a single multivariable model including all variables that are significant in univariable analyses. Other limitations of commonly used models include their inability to detect patterns more complex than a u-curve. More complex non-linear models might provide more insight into the relationships between risk factors such as age and sex, and MD prevalence, potentially improving opportunities for preventative interventions by identifying sub-populations with higher risks of MD. The first aim of this thesis was to gain more insight into the etiology of MD by using rich datasets and novel methodology to take a more detailed look at MD risk factors.

It might also be beneficial to identify groups of patients that respond better to a specific treatment. The heterogeneity of MD makes it difficult to predict treatment response for an individual patient based on the average response from large groups of MD patients, so looking for more homogeneous groups within the patient population (i.e., subtypes) might help. Unfortunately, the first subtypes of MD, which were largely based on consensus, have not performed much better than the original classification with regards to the prediction of onset, course, and treatment response. Data-driven approaches address the issue of intra-class heterogeneity by using clustering methods in patient datasets to identify subtypes that might be missed when relying on clinical observation alone. Although data-driven approaches to psychiatric diagnostics have long been used, they have recently gained more popularity. However, because of several methodological issues, it is still unclear how much of an improvement can be made with data-driven subtypes. More insight into the effects of methodological variation on clustering results could greatly improve our understanding of data-driven subtyping, helping us to not over-

interpret the results of a single study. In addition, it could provide leads for data-driven subtypes of MD by identification of patterns that are robust to methodological variation. It is also unknown which type of data will deliver the best results when it comes to data-driven subtyping. Research into diagnostic subtypes of depression has predominantly focused on subtyping based on symptom patterns, but there is little evidence showing that heterogeneity in etiology and pathophysiology are best explained by variations at this level. Therefore, it might be necessary to perform subtyping based on other sources of heterogeneity, including clinical risk factors, biochemical markers, genetic variations, and brain region activity/connectivity. This might enable us to identify groups of people that share a similar etiology and/or similar pathophysiology, which could mean that they are more likely to respond to similar treatments. The second aim of this thesis was to investigate if and how well bottom-up subtyping approaches might enable the discovery of more homogeneous subtypes of MD.

The first part of this thesis was intended to refine our understanding of MD etiology by taking both a broad perspective, including a large group of risk factors at once, and a deep perspective, by zooming in further on the relationship between sex, age, and MD. In **Chapter 2**, Relative Importance Analysis was performed using the Lifelines cohort, a large longitudinal population study, in order to identify key risk factors of MD onset and recurrence over a 6-year period. We included 21 risk factors that have previously been found to be related to MD, such as socio-demographic variables, neuroticism, family history, stressful life events, childhood trauma, health behaviors, general health status, and metabolic and inflammatory markers. A family history of anxiety and depression, childhood trauma, higher neuroticism, female sex, younger age, long-term difficulties, lower physical quality of life and current anxiety disorders were all important predictors of MD onset, but family history had the largest effect. Most risk factors for MD onset also predicted MD recurrence, but a higher number of anxiety disorders at baseline predicted first onset only. Lower education levels did not predict onset, but they were the largest predictor of recurrence. Risk of recurrence was highest in men. These results identify a number of key risk factors relevant for population screening to identify subjects at risk of onset or recurrence of MD. For example, screening for MD among family members of depressed individuals may lead to more timely interventions, and the importance of lower education levels as a predictor for recurrence of MD suggests that strategies for tackling educational inequality in MD are needed, especially in relation to the course of the disorder. Finally, these results call for awareness of the potential detrimental course of MD in *both* men and women. In **Chapter 3**, cross-sectional data from the same sample was used to take a closer look at the relationship between sex, age and internalizing disorders (i.e., MD, dysthymia,

generalized anxiety disorder, social phobia and panic disorder), symptoms of depression and anxiety, negative affect (NA), and neuroticism. Generalized additive models were used to identify nonlinear patterns of these internalizing disorders, symptoms, and traits over the participants' lifetimes, and to investigate sex differences. Women reported more internalizing disorders than men, but the relative difference remained stable across age (relative risk ~ 1.7). For both sexes, depressive symptoms decreased slightly from age 18 until the age of 35, increased until the age of 50, and then decreased again until the age of 65, after which symptoms increased again. Anxiety symptoms increased until the age of 40, and then decreased, with a stabilization after age 70. NA and neuroticism gradually decreased after age 18. The patterns of internalizing disorders were different. There were small differences between the disorders, but prevalence generally increased between the ages of 18-30 years, stabilized between 30-50, and decreased after age 50. These results indicate that there might be differences in etiology, and that internalizing symptom scores cannot readily replace the diagnostic criteria. Furthermore, this chapters showed that changes in sex hormones around the menopause do not significantly influence women's risk of internalizing disorders. The findings from these two Chapters together imply that the gender gap is mainly a consequence of differences in incidence early in life, rather than increased recurrence, which is in line with previous studies that showed that the gender gap in MD prevalence arises in puberty, due to higher incidence rates in women. This means that prevention of new onsets of MD should pay specific attention to adolescent women particularly.

The second part of this thesis focused on methodological and empirical questions about bottom-up MD subtyping. **Chapter 4** presents the results of a systematic review of current evidence available for data-driven biological subtypes of MD from studies that identified (1) data-driven subtypes of MD based on biological variables, or (2) data-driven subtypes based on clinical features such as symptom patterns and validated these with biological variables post-hoc, in order to gain insight into existing knowledge about the role of biological factors in MD heterogeneity. Twenty-nine publications including 24 separate analyses in 20 unique samples were identified, including a total of ~4000 subjects. Five out of six biochemical studies indicated that there might be depression subtypes with and without disturbed neurotransmitter levels, and one indicated there might be an inflammatory subtype. Seven symptom-based studies identified subtypes, which were mainly determined by severity and by weight gain vs. loss. Two studies compared subtypes based on medication response. These symptom-based subtypes were associated with differences in biomarker profiles and functional connectivity, but results have not sufficiently been replicated. Four out of five neuroimaging studies found evidence for

groups with structural and connectivity differences, but the methods and results were inconsistent. The single genetic study found a subtype with a distinct pattern of single-nucleotide polymorphisms, but this subtype has not been replicated in an independent test sample. One study combining all aforementioned types of data discovered subtypes with different levels of functional connectivity, childhood abuse, and treatment response, but the sample size was small. Although all of these studies provided interesting leads for future research, the methodological differences across studies and lack of replication precluded definitive conclusions about the existence of clinically useful and generalizable biological subtypes of MD.

Chapter 5 shows that it is possible to successfully apply clustering techniques commonly used in studies based on clinical data to a set of biochemical biomarkers. Latent Class Analysis was performed on data from the Netherlands Study of Depression and Anxiety, a large, multi-site naturalistic cohort study. Thirty-six biomarkers (e.g., leptin, brain-derived neurotrophic factor, tryptophan) were measured, as well as sociodemographic and clinical characteristics. The analyses resulted in three classes, which were primarily characterized by different levels of metabolic health and were labeled as: 'lean' (21.6%), 'average' (62.2%) and 'overweight' (16.2%). The identified classes were strongly related to general (metabolic) health and did not reflect any traditional diagnostic classifications. These results suggest that there might be a subtype with poor metabolic health could be seen as a distal marker for depression and anxiety, which is in line with previous symptom-based subtyping studies. **Chapter 6** details the use of Specification-Curve Analysis (SCA), using 31 proteomic biomarkers previously related to MD in the NESDA sample, to gain more insight into the influence of methodological variation on biomarker-based cluster-analysis results. This analysis evaluated the consistency of the model results across 1,200 k-means and hierarchical clustering analyses, each with a unique model specification (i.e., combination of clustering algorithm, fit-index, and distance metric). The results were inconsistent, meaning that no robust patterns of biological clustering were discovered in either the MD only sample or the combined MD/healthy dataset. Next, SCAs were run in simulated datasets with known varying cluster numbers and noise/outlier levels to evaluate the effect of data properties on SCA outcomes. The simulation results showed that the correct number of clusters could be identified quite consistently across the 1,200 model specifications, but that correct cluster identification became harder when number of clusters and noise levels increased. These results indicated SCA is a useful technique that could aid the development of robust and replicable subtyping models in psychiatric disorders, but that a lack of robust clusters in an SCA does not always mean that there are no clusters in the data. Together, the results of the second part of this thesis show

that although subtyping based on etiology and pathophysiology is a promising research avenue, it is still in its infancy in some ways. If cluster analysis results are sufficiently replicated and proven to be robust to methodological variation, they could provide useful contributions to the total evidence base for new subtypes of MD that can be applied in clinical practice. Currently, the IMD subtype has the largest evidence base, but a lot of research still needs to be done before this classification can be used in clinical practice.

This research project had several strengths such as the large sample sizes, large numbers of included variables, and advanced statistical techniques. However, the results should also be interpreted in the light of a number of limitations, including the limited availability of data related to the main hypotheses about the pathophysiology of MD, the cross-sectional nature of many of the analyses, the lack of out-of-sample validation, and the many sources of methodological variation that were left unexplored.

This dissertation aimed to gain more insight into the etiology of MD, and to investigate if and how well bottom-up subtyping approaches might enable the discovery of more homogeneous subtypes of MD. The first part of this dissertation provided a number of leads for research into potential interventions to prevent of first onset and recurrence of MD, as well as some interesting insights into the gender gap of MD. The second part of this thesis showed that bottom-up data-driven subtyping research is a very complex endeavor that requires elaborate and costly data collection, including many variables from different levels, as well as intricate research designs that enable the evaluation of the robustness of the model results. Furthermore, the results indicated that artificial enhancement of cluster separation and not testing the null results can lead to overestimation of number of clusters, and that clustering results are very sensitive to model specifications. This suggest that most studies will consistently result in at least two clusters, which might not always represent relevant subtypes. Researchers as well as reviewers and editors need to be aware of these issues and evaluate model results accordingly.

Finally, some interesting starting points for future research with regards to bottom-up data-driven subtypes of MD were identified. Specifically, my results indicate that there might be an immunometabolic subtype of MD, although the exact clinical and biochemical characteristics of this subtype still need to be elucidated. In addition, further research into monoamine subtypes of MD is needed to improve the opportunities for personalized treatment of MD. Another promising direction for future research is clustering based on multilevel data, including genetic, neurological, and biochemical as well as phenotypical data.

Curriculum Vitae

Lian Beijers (1990) was born in Winterswijk, the Netherlands. After completing secondary education (Gymnasium, SG Arnhem), she moved to Wageningen in 2011 to study Biology, majoring in Molecular Biology. In 2015 she switched masters, graduating with a degree in Clinical Psychosocial Epidemiology (CPE) from the University of Groningen, including the Honours College track on leadership. During her studies she was involved in teaching bachelor students (statistics, presentation skills), several research projects, and the Educational Committee for CPE.

In 2017, she started her PhD research at the Interdisciplinary Center Psychopathology and Emotion regulation (ICPE, Department of Psychiatry, UMCG) under the supervision of Robert Schoevers, Klaas Wardenaar, and Hanna van Loo. During her PhD program, she continued her extracurricular activities; teaching (science elective on publication bias for bachelor medical students, and psychiatric epidemiology for CPE students), collaborating with other researchers on different projects and representing PhD students in the SHARE PhD Council and the SHARE and GSMS Educational Committees.

List of publications

1. de Vries, Y. A., Roest, A. M., Beijers, L., Turner, E. H. & de Jonge, P. Bias in the reporting of harms in clinical trials of second-generation antidepressants for depression and anxiety: A meta-analysis. *Eur. Neuropsychopharmacol.* **26,** 1752–1759 (2016).

2. Beijers, L., Jeronimus, B. F., Turner, E. H., De Jonge, P. & Roest, A. M. Spin in RCTs of anxiety medication with a positive primary outcome: A comparison of concerns expressed by the US FDA and in the published literature. *BMJ Open* **7,** e012886 (2017).

3. Beijers, L., Wardenaar, K. J., Bosker, F. J., Lamers, F., Van Grootheest, G., De Boer, M. K., Penninx, B. W. J. H. & Schoevers, R. A. Biomarker-based subtyping of depression and anxiety disorders using Latent Class Analysis. A NESDA study. *Psychol. Med.* **49,** 617–627 (2019).

4. Van Bronswijk, S., Moopen, N., Beijers, L., Ruhe, H. G. & Peeters, F. Effectiveness of psychotherapy for treatment-resistant depression: A meta-analysis and meta-regression. *Psychol. Med.* **49,** 366–379 (2019).

5. Jentsch, M. C., Burger, H., Meddens, M. B. M., Beijers, L., Van Den Heuvel, E. R., Meddens, M. J. M. & Schoevers, R. A. Gender differences in developing biomarker-based major depressive disorder diagnostics. *Int. J. Mol. Sci.* **21,** 3039 (2020).

6. Beijers, L., Van Loo, H. M., Romeijn, J. W., Lamers, F., Schoevers, R. A. & Wardenaar, K. J. Investigating data-driven biological subtypes of sychiatric disorders using specification-curve analysis. *Psychol. Med.* 1–12 (2020).

7. Van Loo, H. M., Beijers, L., Wieling, M., De Jong, T. R., Schoevers, R. A. & Kendler, K. S. Prevalence of internalizing disorders, symptoms, and traits across age using advanced nonlinear models. *Psychol. Med.* 1–10 (2021).

Submitted for publication

Beijers, L., Wardenaar, K. J., De Jong, T. R., Schoevers, R. A. & Van Loo, H. M. Key risk factors for onset and recurrence of Major Depression: results from Lifelines, a large representative population cohort.

Acknowledgments

This dissertation would not have been completed without the support of many people. I would like to thank all these people and a few in particular.

First of all, I would like to thank my supervisors. As I have shared time and time again with co-workers, friends, and family, I feel very privileged to have worked with such a great team. I could not have wished for better co-supervisors than Klaas Wardenaar and Hanna van Loo, who both took a genuine interest in not just my research but also my personal well-being. Our weekly meetings were generally very useful, but better still, always very 'gezellig'. I am also very grateful to my promotor, Robert Schoevers. Despite the fact that his schedule seems to fill up more quickly than any human being can work, he always came up with interesting suggestions in our monthly meetings, and always supported my personal development.

The past four years would have been a lot less enjoyable without the many colleagues I got to know at the ICPE. To my colleagues Anna and Melissa, it's a shame we didn't get to spend our final years in the office together.

I want to thank my family for their unconditional love and support. I've been through many ups and downs in these past four years, and they were there for all of it. I would like to thank my parents again for sharing their living space with me and allowing me the use of their bedroom during office hours, so I could finish this thesis from the safety of my parental home.

Finally, I want to thank the Plum Village community, particularly the lovely Wake Up Sangha (Groningen and international), and the Mind and Life/Mindful Researchers community, for providing a sense of safety, purpose and connection in these trying times. This support network has been invaluable to the realization of this thesis, and my growth as a person.

Nederlandstalige samenvatting

Depressie is een veel voorkomende en slopende aandoening. Helaas zijn de meeste behandelingen voor depressie maar matig effectief in klinische onderzoeken. Een beter begrip van de oorzaken van depressie zou psychiaters in staat kunnen stellen depressieve patiënten een meer gedetailleerde diagnose te geven, waardoor ze behandelingen zouden kunnen bieden die zijn afgestemd op de etiologische en pathofysiologische achtergrond van een patiënt.

Aangezien de effectiviteit van de huidige therapieën in vergelijking met placebo bescheiden is zijn er wellicht andere benaderingen nodig om de effecten van depressie op de volksgezondheid aan te pakken, zoals bijvoorbeeld preventieve interventies voor zowel de eerste als terugkerende episodes van depressie. Het identificeren van de belangrijkste risicofactoren voor depressie zou kunnen helpen om te bepalen op welke patiënten of welke ziektemechanismen deze interventies zich zouden moeten richten. Helaas wordt dit soort onderzoek bemoeilijkt door een gebrek aan datasets met voldoende deelnemers en een groot aantal risicofactoren. Daarnaast vragen dit soort analyses veel rekenkracht, dus rapporteren de meeste onderzoeken andere modellen als belangrijkste uitkomst, zoals bijvoorbeeld een model waarin alle variabelen die significant waren in univariabele analyses zijn opgenomen. Andere beperkingen van veelgebruikte modellen zijn onder meer hun onvermogen om patronen te detecteren die complexer zijn dan een u-curve. Complexere niet-lineaire modellen zouden meer inzicht kunnen geven in de relaties tussen risicofactoren zoals leeftijd en geslacht, en de prevalentie van depressie. Hierdoor zouden wellicht subpopulaties met een hoger risico op depressie geïdentificeerd kunnen worden, wat mogelijkheden oplevert voor preventieve interventies. Het eerste doel van dit proefschrift was om meer inzicht te krijgen in de etiologie van depressie door gebruik te maken van rijke datasets en nieuwe methodologie om een meer gedetailleerde kijk te krijgen op risicofactoren van depressie.

Het kan ook nuttig zijn om groepen patiënten te identificeren die beter reageren op een specifieke behandeling. De heterogeniteit van depressie maakt het moeilijk om de behandelrespons voor een individuele patiënt te voorspellen op basis van de gemiddelde respons van grote groepen depressieve patiënten, dus zoeken naar meer homogene groepen binnen de patiëntenpopulatie (d.w.z. subtypes) zou kunnen helpen. Helaas hebben de eerste subtypes van depressie, die grotendeels gebaseerd waren op consensus, niet veel beter gepresteerd dan de oorspronkelijke classificatie met betrekking tot het voorspellen van het begin en het verloop van de stoornis, of de behandelrespons. Data-gedreven benaderingen pakken het probleem van heterogeniteit aan door met statistische

clusteringstechnieken subtypen te identificeren die mogelijk over het hoofd worden gezien wanneer alleen op klinische observatie wordt vertrouwd. Hoewel data-gedreven benaderingen al lang worden gebruikt binnen de psychiatrische nosologie zijn ze de laatste jaren steeds populairder geworden. Vanwege een aantal methodologische problemen is het echter nog steeds onduidelijk hoeveel verbetering er verwacht kan worden van dit soort data-gedreven subtypen. Meer inzicht in de effecten van methodologische variatie op clusteringresultaten zou ons begrip van data-gedreven subtypering aanzienlijk kunnen verbeteren, waardoor we de resultaten van een enkele studie niet overschatten. Bovendien zou het aanknopingspunten kunnen bieden voor data-gedreven subtypes van depressie door patronen te identificeren die robuust zijn voor methodologische variatie. Het is ook onbekend welk type data de beste resultaten zal opleveren als het gaat om data-gedreven subtypering. Onderzoek naar diagnostische subtypes van depressie heeft zich tot nu toe voornamelijk gericht op subtypering op basis van symptoompatronen, maar er is weinig bewijs dat heterogeniteit in etiologie en pathofysiologie het best kan worden verklaard door variaties op dit niveau. Daarom kan het nodig zijn om subtypering uit te voeren op basis van andere bronnen van heterogeniteit, waaronder klinische risicofactoren, biochemische markers, genetische variaties en activiteit/connectiviteit van de hersenen. Dit zou ons in staat kunnen stellen om groepen mensen te identificeren die een dezelfde etiologie en/of pathofysiologie hebben, wat waarschijnlijk betekent dat ze meer kans hebben om goed te reageren op dezelfde behandelingen. Het tweede doel van dit proefschrift was om te onderzoeken of, en hoe goed, data-gedreven subtypering op basis van biologische data de ontdekking van meer homogene subtypes van depressie mogelijk maken.

Het eerste deel van dit proefschrift was bedoeld om ons begrip van de etiologie van depressie te verfijnen door zowel een verbredend perspectief als een verdiepend perspectief in te nemen, door te kijken naar een grote groep risicofactoren tegelijk en daarna verder in te zoomen op de relatie tussen geslacht, leeftijd en depressie. In **Hoofdstuk 2** werden de belangrijkste risicofactoren voor het ontstaan en terugkeren van depressie over een periode van 6 jaar geïdentificeerd in Lifelines, een groot longitudinaal populatieonderzoek. We hebben 21 risicofactoren bekeken waarvan eerder is vastgesteld dat ze verband houden met depressie, zoals sociaal-demografische variabelen, neuroticisme, familiegeschiedenis, acute en chronische stress, jeugdtrauma, leefstijl, algemene gezondheidstoestand en metabole en inflammatoire markers. Een familiegeschiedenis van angst en depressie, jeugdtrauma, hoger neuroticisme, vrouwelijk geslacht, jongere leeftijd, chronische stress, lagere fysieke kwaliteit van leven en het al hebben van een of meer angststoornissen waren allemaal belangrijke voorspellers van het ontstaan van depressie, maar familiegeschiedenis had het grootste effect. De meeste risicofactoren voor het ontstaan voorspelden ook het terugkeren

van depressie, maar een hoger aantal angststoornissen bij de eerste meting voorspelde alleen het eerste optreden van depressie. Een lager opleidingsniveau was niet van groot belang bij het voorspellen van nieuwe depressies, maar was de sterkste voorspeller van het terugkeren van depressie. Het risico op terugkerende depressies was het hoogst bij mannen. Deze resultaten laten zien dat een aantal belangrijke risicofactoren mogelijk relevant kunnen zijn voor bevolkingsonderzoek, om daardoor personen te identificeren met een risico op het ontstaan of terugkeren van depressie. Screening op depressie onder familieleden van depressieve personen kan bijvoorbeeld leiden tot meer tijdige interventies, en het belang van een lager opleidingsniveau als voorspeller voor herhaling van depressie suggereert dat strategieën om onderwijsongelijkheid bij depressie aan te pakken noodzakelijk zijn, vooral met betrekking tot het verloop van de stoornis. Ten slotte roepen deze resultaten op tot bewustwording van het mogelijk schadelijke beloop van depressie bij *zowel* mannen als vrouwen. In **Hoofdstuk 3** werden cross-sectionele gegevens uit dezelfde steekproef gebruikt om gedetailleerd onderzoek te doen naar de relatie tussen sekse, leeftijd en internaliserende stoornissen (d.w.z., depressie, dysthymie, gegeneraliseerde angststoornis, sociale fobie en paniekstoornis), symptomen van depressie en angst, negatief affect (NA), en neuroticisme. Gegeneraliseerde additieve modellen werden gebruikt om niet-lineaire patronen van deze internaliserende stoornissen, symptomen en eigenschappen gedurende het leven van de deelnemers te identificeren en om sekseverschillen te onderzoeken. Vrouwen rapporteerden meer internaliserende stoornissen dan mannen, maar het relatieve verschil bleef stabiel over de leeftijd (relatief risico ~ 1,7). Voor beide geslachten namen de depressieve symptomen licht af van de leeftijd van 18 tot de leeftijd van 35 jaar, namen toe tot de leeftijd van 50 jaar en namen daarna weer af tot de leeftijd van 65 jaar, waarna de symptomen weer toenamen. Angstsymptomen namen toe tot de leeftijd van 40 jaar en namen daarna af, met een stabilisatie na de leeftijd van 70 jaar. NA en neuroticisme namen geleidelijk af na de leeftijd van 18 jaar. De patronen van internaliserende stoornissen waren anders. Er waren kleine verschillen tussen de stoornissen, maar de prevalentie nam over het algemeen toe tussen 18-30 jaar, stabiliseerde tussen 30-50 en nam af na de leeftijd van 50 jaar. Deze resultaten geven aan dat er verschillen kunnen zijn in etiologie en dat de diagnostische criteria voor internaliserende stoornissen niet makkelijk vervangen kunnen worden door symptoomscores. Verder toonden dit hoofdstuk aan dat veranderingen in geslachtshormonen rond de menopauze geen significante invloed hebben op het risico van vrouwen op internaliserende stoornissen. De bevindingen van deze twee hoofdstukken samen impliceren dat de genderkloof voornamelijk een gevolg is van verschillen in incidentie op jonge leeftijd, en niet zozeer van een toegenomen terugval, wat in overeenstemming is met eerdere studies die aantoonden dat de genderkloof in de

prevalentie van depressie in de puberteit ontstaat door hogere incidentiecijfers bij vrouwen. Dit betekent dat bij de preventie van het ontstaan van depressie specifiek aandacht moet worden besteed aan adolescente vrouwen.

Het tweede deel van dit proefschrift was gericht op methodologische en empirische vragen over subtypering van depressie gebaseerd op biologische gegevens. **Hoofdstuk 4** presenteert het huidige bewijs dat beschikbaar is voor data-gedreven biologische subtypes van depressie, om inzicht te krijgen in bestaande kennis over de rol van biologische factoren in depressie-heterogeniteit. Hierbij werd gekeken naar onderzoeken die (1) data-gedreven subtypes van depressie identificeerden op basis van biologische variabelen, of (2) data-gedreven subtypes identificeerden op basis van klinische kenmerken zoals symptoompatronen en deze post-hoc valideerden met biologische variabelen. Negenentwintig publicaties, waaronder 24 afzonderlijke analyses in 20 unieke steekproeven, werden geïdentificeerd, met een totaal van ~4000 proefpersonen. Vijf van de zes biochemische onderzoeken gaven aan dat er mogelijk subtypes van depressie zijn met en zonder verstoorde neurotransmitterniveaus, en één gaf aan dat er een inflammatoir subtype zou kunnen zijn. Zeven op symptomen gebaseerde onderzoeken identificeerden subtypen die voornamelijk werden bepaald door de ernst van de depressie, en door gewichtstoename versus gewichtsverlies. Twee studies vergeleken subtypes op basis van medicatierespons. Deze op symptomen gebaseerde subtypes waren geassocieerd met verschillen in biomarkers en functionele connectiviteit, maar de resultaten zijn nog niet voldoende gerepliceerd om duidelijke conclusies te kunnen trekken. Vier van de vijf neuroimaging-onderzoeken vonden bewijs voor groepen met structurele en connectiviteitsverschillen, maar de methoden en resultaten waren inconsistent. De enkele genetische studie vond een subtype met een duidelijk patroon van genetische verschillen (EN: single-nucleotide polymorphisms), maar dit subtype is nog niet gerepliceerd. Een studie die alle bovengenoemde soorten gegevens combineerde ontdekte subtypen met verschillende niveaus van functionele connectiviteit, kindermishandeling en behandelingsrespons, maar de steekproefomvang was klein. Hoewel al deze onderzoeken interessante aanknopingspunten bieden voor toekomstig onderzoek kunnen we nog geen definitieve conclusies trekken over het bestaan van klinisch bruikbare en generaliseerbare biologische subtypes van depressie, vanwege de methodologische verschillen tussen de onderzoeken en het gebrek aan replicatie.

Hoofdstuk 5 laat zien dat het mogelijk is om clustertechnieken die vaak worden gebruikt in studies op basis van klinische gegevens met succes toe te passen op een reeks biochemische biomarkers. Latente klassen-analyse werd uitgevoerd op zesendertig biomarkers (bijv. leptine en tryptofaan) gemeten in de Nederlandse Studie naar Depressie

en Angst (NESDA), een groot naturalistisch cohortonderzoek met meerdere locaties. De analyses resulteerden in drie klassen die voornamelijk werden gekenmerkt door verschillende niveaus van metabole gezondheid. De klassen werden gelabeld als: 'mager' (21,6%), 'gemiddeld' (62,2%) en 'overgewicht' (16,2%). De geïdentificeerde klassen waren sterk gerelateerd aan de algemene (metabole) gezondheid van de deelnemers en weerspiegelden geen traditionele diagnostische classificaties. Deze resultaten suggereren dat er mogelijk een subtype is met een slechte metabole gezondheid, wat in overeenstemming is met eerdere subtyperings-studies gebaseerd op symptomen **Hoofdstuk 6** beschrijft het gebruik van Specificatie-Curve Analyse (SCA) om meer inzicht te krijgen in de invloed van methodologische variatie op cluster-analyse resultaten gebaseerd op biomarkers, waarbij gebruik wordt gemaakt van 31 proteomische biomarkers die eerder gerelateerd waren aan depressie in NESDA. Deze analyse evalueerde de consistentie van de modelresultaten over 1.200 verschillende k-means en hiërarchische clusteranalyses, elk met een unieke modelspecificatie (d.w.z., een unieke combinatie van clusteringalgoritme, fit-index en afstandsmaat). De resultaten waren inconsistent, wat betekent dat er geen robuuste patronen van biologische clustering werden ontdekt in het de groep depressieve patiënten of de dataset waar ook gezonde controles aan toegevoegd waren. Vervolgens werden SCAs uitgevoerd in gesimuleerde datasets met verschillende clusteraantallen en verschillende niveaus van ruis om het effect van deze eigenschappen op SCA-resultaten te evalueren. De simulatieresultaten toonden aan dat het juiste aantal clusters vrij consistent kon worden geïdentificeerd over de 1.200 modelspecificaties, maar dat correcte clusteridentificatie moeilijker werd naarmate het aantal clusters en het niveau van ruis toenamen. Deze resultaten gaven aan dat SCA een bruikbare techniek is die zou kunnen helpen bij de ontwikkeling van robuuste en repliceerbare subtyperingsmodellen bij psychiatrische stoornissen, maar dat een gebrek aan robuuste clusters in een SCA niet altijd betekent dat er geen clusters in de gegevens zijn. Samen laten de resultaten van het tweede deel van dit proefschrift zien dat, hoewel subtypering op basis van etiologie en pathofysiologie een veelbelovende onderzoeksrichting is, het in sommige opzichten nog in de kinderschoenen staat. Als de resultaten van clusteranalyse voldoende worden gerepliceerd en robuust blijken te zijn voor methodologische variatie, kunnen ze een nuttige bijdrage leveren aan de totale wetenschappelijke basis voor nieuwe subtypes van depressie die in de klinische praktijk kunnen worden toegepast. Momenteel heeft het imuun-metabole subtype de grootste wetenschappelijke basis, maar er moet nog veel onderzoek worden gedaan voordat deze classificatie in de klinische praktijk kan worden gebruikt.

Dit proefschrift had verschillende sterke punten, zoals de grote aantallen deelnemers, het grote aantal variabelen en geavanceerde statistische technieken. De resultaten moeten

echter ook worden geïnterpreteerd in het licht van enkele beperkingen, waaronder de beperkte beschikbaarheid van gegevens met betrekking tot de belangrijkste hypothesen over de pathofysiologie van depressie, het cross-sectionele karakter van veel van de analyses en de vele bronnen van methodologische variatie niet zijn onderzocht.

Dit proefschrift had tot doel meer inzicht te krijgen in de etiologie van depressie, en om te onderzoeken of, en hoe goed, data-gedreven subtypering op basis van biologische gegevens de ontdekking van meer homogene subtypes van depressie mogelijk zouden kunnen maken. Het eerste deel van dit proefschrift leverde een aantal aanknopingspunten op voor onderzoek naar mogelijke interventies om het ontstaan en terugkeren van depressie te voorkomen, en enkele interessante inzichten in de genderkloof van depressie. Het tweede deel van dit proefschrift toonde aan dat data-gedreven subtyperingsonderzoek een zeer complexe onderneming is die uitgebreide en kostbare gegevensverzameling vereist, inclusief een grote hoeveelheid variabelen van verschillende niveaus, evenals ingewikkelde methoden die de evaluatie van de robuustheid van de modelresultaten mogelijk maken. Bovendien gaven de resultaten aan dat kunstmatige verbetering van scheiding tussen clusters en het niet testen van de nulhypothese kan leiden tot overschatting van het aantal clusters, en dat clusteringresultaten erg gevoelig zijn voor modelspecificaties. Dit suggereert dat de meeste onderzoeken altijd ten minste twee clusters zullen presenteren, die mogelijk niet altijd relevante subtypen vertegenwoordigen. Zowel onderzoekers als recensenten en redacteuren moeten op de hoogte zijn van deze problemen en zouden de resultaten van clusteranalyses idealiter dienovereenkomstig evalueren.

Tot slot werden er in dit proefschrift een aantal interessante uitgangspunten voor toekomstig onderzoek naar data-gedreven subtypes van depressie geïdentificeerd. De resultaten geven aan dat er mogelijk een immuun-metabool subtype van depressie is, hoewel de exacte klinische en biochemische kenmerken van dit subtype nog moeten worden onderzocht. Daarnaast is verder onderzoek naar monoamine-subtypes van depressie nodig om de mogelijkheden voor gepersonaliseerde behandeling van depressie te verbeteren. Een andere veelbelovende richting voor toekomstig onderzoek is clustering op basis van gegevens van meerdere niveaus, zoals genetische, neurologische, biochemische en fenotypische gegevens.

Lekensamenvatting

Inleiding

Depressie is een veel voorkomende en slopende aandoening, die vaak als meer belastend ervaren wordt dan ziekten zoals bijvoorbeeld kanker of hartproblemen. Helaas zijn de meeste behandelingen voor depressie niet zo effectief als we graag zouden willen. Een beter begrip van de oorzaken van depressie zou kunnen helpen om betere behandelingen te ontwikkelen, en om de behandeling beter te laten aansluiten op wat een depressieve patiënt nodig heeft. In mijn proefschrift heb ik op twee manieren onderzoek gedaan naar de oorzaken van depressie.

In het eerste deel heb ik gekeken naar risicofactoren voor depressie. Risicofactoren zijn persoonlijke kenmerken (bijvoorbeeld leeftijd, persoonlijkheid en opleidingsniveau) en dingen die iemand heeft meegemaakt (bijvoorbeeld jeugdtrauma's of het overlijden van een naaste) die kunnen lijden tot een grotere kans op depressie. Ik heb eerst verschillende risicofactoren met elkaar vergeleken om te kijken welke factoren de grootste rol spelen bij het ontstaan van depressies, en daarna heb ik gedetailleerd onderzoek gedaan naar de effecten van sekse en leeftijd.

In het tweede deel heb ik onderzocht of er misschien subtypen van depressie bestaan, dat zijn groepen patiënten waarbij de depressie op verschillende manieren veroorzaakt wordt. Ik heb daarbij ook geprobeerd te bekijken hoe de methoden die gebruikt worden voor dit soort onderzoek verbeterd kunnen worden.

Hieronder zal ik deze twee benaderingen verder uitleggen. Ik zal daarbij de resultaten van mijn onderzoek en de interpretatie hiervan voor het eerste en het tweede deel apart beschrijven.

Deel 1. Risicofactoren van depressie

Achtergrond

Omdat de huidige behandelingen voor depressie niet genoeg helpen zijn er misschien andere benaderingen nodig om de effecten van depressie op de volksgezondheid aan te pakken. Het zou onder andere kunnen helpen om het ontstaan van nieuwe depressies te voorkomen door middel van preventieve behandelingen. Het gaat dan bijvoorbeeld om een online cursus om meer grip te krijgen op depressieve klachten die aangeboden wordt voordat deze klachten zich ontwikkelen tot een volledige depressie. Om te onderzoeken op welke risicogroepen en/of op welke ziektemechanismen deze preventieve behandelingen zich zouden moeten richten is het belangrijk om te weten welke risicofactoren de grootste rol spelen bij het ontstaan van depressie. Helaas bestaan er weinig datasets die dit soort

onderzoek mogelijk maken. Daarnaast zijn de meest gebruikte onderzoeksmethoden vaak niet in staat om risicofactoren met elkaar te vergelijken. Deze methoden zijn ook niet geschikt om complexe verbanden tussen een risicofactor zoals bijvoorbeeld leeftijd en/of sekse, en het ontstaan van depressies te onderzoeken. Complexere modellen zouden een meer gedetailleerd inzicht kunnen geven in dit verband, waardoor het misschien mogelijk wordt om groepen vinden die een hoger risico hebben op het krijgen van een depressie (bijvoorbeeld vrouwen rond de overgang), wat mogelijkheden oplevert voor preventieve behandelingen.

Doel van het onderzoek

Het eerste deel van dit proefschrift was bedoeld om meer inzicht te krijgen in risicofactoren van depressie door gebruik te maken van Lifelines (een dataset met heel veel verschillende gegevens over meer dan 160.00 deelnemers) en complexe modellen die (1) het belang van risicofactoren kunnen vergelijken en (2) complexe verbanden tussen geslacht en leeftijd en het voorkomen van depressie kunnen onderzoeken.

Resultaten

In **Hoofdstuk 2** werd gekeken naar de belangrijkste risicofactoren voor het ontstaan van depressies. Ik heb daarbij onderzocht of de belangrijkste risicofactoren voor een nieuwe depressieve periode verschillen tussen mensen die al eerder een depressie hebben gehad, en mensen die nog nooit depressief zijn geweest. Ik heb 21 risicofactoren bekeken, waarvan eerder is vastgesteld dat ze verband houden met depressie, waaronder leeftijd, geslacht, neuroticisme, familiegeschiedenis van depressie en angst, stress, jeugdtrauma, leefstijl, en algemene gezondheidstoestand.

Het voorkomen van angst en depressie in de familie, jeugdtrauma, hoger neuroticisme, vrouwelijk geslacht, jongere leeftijd, chronische stress, lagere fysieke kwaliteit van leven en het al hebben van één of meer angststoornissen waren allemaal belangrijke voorspellers van het ontstaan van depressie, maar familiegeschiedenis had het grootste effect. De meeste risicofactoren voor het ontstaan van nieuwe depressies (d.w.z., bij mensen die nog niet eerder depressief waren geweest) voorspelden ook het terugkeren van depressie (d.w.z., bij mensen met een voorgeschiedenis van depressie). Een lager opleidingsniveau was bijvoorbeeld niet van groot belang bij het voorspellen van nieuwe depressies, maar was de sterkste voorspeller van het terugkeren van depressie. Het risico op het terugkeren van depressies bij mensen die al eens depressief zijn geweest was het hoogst bij mannen.

Deze resultaten lieten zien welke risicofactoren mogelijk relevant zouden kunnen zijn voor bevolkingsonderzoek, om op die manier personen te identificeren die een hoger

risico hebben op het ontstaan of terugkeren van depressies. Screening op depressie onder familieleden van depressieve personen zou bijvoorbeeld eerder ingrijpen bij mensen die depressieve klachten ontwikkelen mogelijk kunnen maken.

In **Hoofdstuk 3** werden complexe wiskundige modellen gebruikt om te kijken hoe angststoornissen en depressie (en hun symptomen) variëren gedurende het leven van de deelnemers van Lifelines. Daarbij werd ook gekeken naar sekseverschillen. Vrouwen hadden vaker een angststoornis of depressie dan mannen, maar het relatieve verschil tussen mannen en vrouwen bleef gelijk: vrouwen hadden op elke leeftijd ongeveer 1,7 keer meer risico op een angststoornis of depressie. Er waren kleine verschillen tussen de gemeten stoornissen, maar de hoeveelheid patiënten met een angststoornis of depressie nam over het algemeen toe tussen 18-30 jaar, stabiliseerde tussen 30-50 en nam af na de leeftijd van 50 jaar.

Deze resultaten lieten zien dat veranderingen in geslachtshormonen rond de menopauze waarschijnlijk geen significante invloed hebben op het gemiddelde risico van vrouwen op angststoornissen en depressie.

De resultaten van deze twee hoofdstukken samen impliceren dat vrouwen gemiddeld genomen vaker depressief zijn omdat zij op jonge leeftijd meer risico hebben op het krijgen van een depressie. Dit is in overeenstemming met eerdere studies die aantoonden dat de genderkloof in het voorkomen van depressie ontstaat in de puberteit. Dit betekent dat bij onderzoek naar behandelingen die het ontstaan van depressie proberen te voorkomen specifiek aandacht moet worden besteed aan adolescente vrouwen.

Deel 2. Subtypen van depressie

Achtergrond

Een andere benadering die zou kunnen helpen om de ziektelast van depressie te verminderen is het zoeken naar groepen patiënten die beter reageren op een specifieke behandeling. Er zijn grote verschillen tussen depressieve patiënten ten aanzien van symptomen, reactie op behandeling, en waarschijnlijk ook in de oorzaken van de depressie. Dit maakt het moeilijk om te voorspellen hoe goed een individuele patiënt zal reageren op een behandeling op basis van het gemiddelde van de hele groep. Het identificeren van groepen patiënten die meer op elkaar lijken (d.w.z., subtypes) zou deze kwaliteit van deze voorspelling kunnen verbeteren.

Er is in het verleden al redelijk wat onderzoek gedaan naar mogelijke subtypes van depressie ontwikkeld, maar dit heeft tot nu toe nog niet veel geholpen. Het zou kunnen dat dit komt doordat deze mogelijke subtypes gebaseerd zijn op klinische observatie, d.w.z., het onderzoeken van depressieve patiënten door psychiaters in de praktijk. Data-gedreven

benaderingen zijn wiskundige technieken die gebruikt kunnen worden om subtypen te identificeren op basis van een grote hoeveelheid gegevens. Omdat de patroonherkenning in geval door een computer wordt gedaan in plaats van door mensen zouden dit soort benaderingen kunnen helpen subtypen te identificeren die mogelijk over het hoofd worden gezien wanneer alleen op klinische observatie wordt vertrouwd.

Hoewel data-gedreven benaderingen al lang worden gebruikt binnen het psychiatrisch onderzoek zijn ze de laatste jaren steeds populairder geworden. Er zijn echter een aantal problemen met dit soort onderzoek waardoor het nog steeds onduidelijk is hoeveel verbetering er verwacht kan worden van data-gedreven subtypen. Het is bijvoorbeeld nog onduidelijk in hoeverre de resultaten van dit soort onderzoek bepaald worden door de gebruikte methoden. Als de resultaten kunnen veranderen als er iets aangepast wordt in de methode moeten we misschien voorzichtiger zijn met het interpreteren van de resultaten van dit soort onderzoek. Aan de andere kant, als we subtypen kunnen vinden die niet sterk afhangen van de gebruikte methode zou dit een belangrijke aanwijzing zijn dat deze resultaten mogelijk nuttig kunnen zijn bij het diagnosticeren van patiënten met depressie.

Het is ook nog niet duidelijk welke soort gegevens de beste resultaten zal opleveren als het gaat om data-gedreven subtypering. Onderzoek naar subtypes van depressie heeft tot nu toe voornamelijk gebruik gemaakt van symptomen, maar er is weinig bewijs dat dit de beste optie is. Daarom moet er onderzoek worden gedaan naar subtypering op basis van andere gegevens (bijvoorbeeld biologische gegevens zoals genetische informatie, breinactiviteit of de hoeveelheid stresshormoon in het bloed). Dit zou misschien kunnen helpen om groepen patiënten te vinden waarbij de oorzaken van depressie hetzelfde zijn, wat betekent dat ze meer kans hebben om goed te reageren op dezelfde behandelingen. Het doel van het tweede deel van dit proefschrift was om te onderzoeken of, en hoe goed, data-gedreven subtypering op basis van biologische gegevens de ontdekking van subtypes van depressie mogelijk maken.

Resultaten

Hoofdstuk 4 beschreef de literatuur die beschikbaar is over subtypes van depressie die gebaseerd zijn op biologische gegevens. Hierbij werd gekeken naar onderzoeken die (1) biologische gegevens gebruikten als basis voor hun onderzoek, of (2) symptomen gebruikten als basis voor hun onderzoek, maar de subtypes vervolgens vergeleken op biologische kenmerken. Dit hoofdstuk heeft gebruik gemaakt van negenentwintig artikelen met een totaal van ongeveer 4000 proefpersonen. Vijf onderzoeken gaven aan dat er mogelijk subtypes van depressie zijn met en zonder verstoorde neurotransmitterniveaus, dat zijn signaalstoffen die impulsen overbrengen in de zenuwen in de hersenen en het lichaam. Zeven op symptomen

gebaseerde onderzoeken identificeerden subtypen die voornamelijk werden bepaald door de ernst van de depressie, en door gewichtstoename versus gewichtsverlies. Deze subtypen verschilden van elkaar op het gebied van stofwisseling, ontsteking, hersenactiviteit, en een aantal andere biologische kenmerken. Er waren ook studies die onderzoek deden naar subtypes op basis van genetische informatie of hersenactiviteit. Hoewel alle studies uit dit hoofdstuk interessante aanknopingspunten boden voor toekomstig onderzoek kon ik helaas nog geen definitieve conclusies trekken over het bestaan van biologische subtypes van depressie, vanwege de verschillen tussen de onderzoeken.

Hoofdstuk 5 liet zien dat het mogelijk is om de subtypes te vinden op basis van biochemische gegevens zoals cholesterol, leptine en tryptofaan met dezelfde technieken die vaak worden gebruikt in studies die gebruik maken van symptomen als basis voor hun subtypes. De gevonden groepen werden voornamelijk gekenmerkt door verschillende niveaus van metabole gezondheid (d.w.z., stofwisseling) en kregen de labels 'mager' (21,6%), 'gemiddeld' (62,2%) en 'overgewicht' (16,2%). De resultaten van dit hoofdstuk suggereren dat er mogelijk een subtype van depressie is met een slechte metabole gezondheid, wat in overeenstemming is met eerdere subtyperings-studies gebaseerd op symptomen.

Hoofdstuk 6 beschreef een analyse die bekeek of het aantal gevonden groepen hetzelfde bleef, of juist niet, als er kleine veranderingen werden aangebracht in de gebruikte methode. Daarvoor werden de resultaten van 1.200 unieke analyses bekeken. Deze methoden zochten allemaal naar subtypen in een groep depressieve patiënten, maar ze gebruikten allemaal net een beetje andere wiskunde. De aantallen groepen die gevonden werden verschilden sterk, wat wil zeggen dat de invloed van de methode op de resultaten groot was.

Dezelfde analyse werd ook gedaan met een aantal datasets ik zelf had gemaakt, waarvan ik dus van tevoren wist hoeveel groepen er gevonden zouden moeten worden. Het merendeel van de 1200 methoden vond het juiste aantal groepen in deze datasets, maar naar mate het aantal groepen in de dataset groter werd ging dit steeds minder goed. Ik had ook verschillende datasets gemaakt met een steeds hoger niveau van ruis, wat het vinden van het juiste aantal groepen verder bemoeilijkte.

Dit hoofdstuk liet zien dat het doen van veel verschillende analyses die onderling een klein beetje verschillen zou kunnen helpen bij de ontwikkeling van subtypes van depressie, omdat het mogelijk is groepen te vinden die bestand zijn tegen dit soort veranderingen. Maar als dit soort onderzoek geen duidelijke uitkomsten geeft kan ook het gevolg zijn problemen met de dataset, dus een gebrek aan uitkomsten betekent niet altijd dat er geen subtypes aanwezig zijn.

De resultaten van deze drie hoofdstukken samen laten zien dat, hoewel subtypering op basis van biologische gegevens een veelbelovende richting is, dit onderzoek in sommige

opzichten nog in de kinderschoenen staat. Momenteel heeft het subtype met biologische ontregeling op het gebied van stofwisseling en ontsteking de grootste wetenschappelijke basis, maar er moet nog veel onderzoek worden gedaan voordat dit subtype van depressie gebruikt kan worden om mensen een diagnose te geven en te behandelen.

Sterke en zwakke punten van dit onderzoek

Dit proefschrift had verschillende sterke punten, zoals de grote aantallen deelnemers, het grote aantal gemeten gegevens en de geavanceerde wiskundige technieken. De resultaten moeten echter ook worden geïnterpreteerd in het licht van enkele beperkingen, zoals bijvoorbeeld de beperkte beschikbaarheid van biologische gegevens over de belangrijkste theorieën over de oorzaken van depressie (bijvoorbeeld bepaalde neurotransmitters, stresshormoon, of ontstekingswaardes). Omdat de gebruikte datasets net zoals andere grote datasets niet beschikken over deze gegevens is het niet mogelijk om te proberen subtypes te vinden waarbij een van deze dingen het hoofdprobleem is, bijvoorbeeld een subtype met een verhoogd stresshormoon en een subtype met een tekort aan bepaalde neurotransmitters. Een andere beperking is dat ik in dit proefschrift alleen onderzoek heb gedaan naar effecten van kleine variaties in de wiskundige methoden - er zijn nog vele bronnen van variatie in onderzoeksmethoden die ik niet heb onderzocht. Kleine veranderingen in de manier waarop de gebruikte gegevens bewerkt worden om ze klaar te maken voor de analyses zouden bijvoorbeeld een vergelijkbaar effect kunnen hebben op de uitkomsten van data-gedreven subtyperingsonderzoek.

Conclusie

Het eerste deel van dit proefschrift leverde een aantal aanknopingspunten op voor onderzoek naar mogelijke interventies om het ontstaan en terugkeren van depressie te voorkomen, en enkele interessante inzichten in de genderkloof van depressie. Het tweede deel van dit proefschrift toonde aan dat subtyperingsonderzoek een zeer complexe onderneming is, waarvoor geïnvesteerd moet worden in het verzamelen van uitgebreide datasets en de ontwikkeling van nieuwe en ingewikkelde methodes. Tot slot werden er in dit proefschrift een aantal interessante uitgangspunten voor toekomstig onderzoek naar data-gedreven subtypes van depressie geïdentificeerd, zoals verder onderzoek naar subtypes met biologische ontregeling op het gebied van stofwisseling en ontsteking of subtypes met verschillende niveaus van bepaalde stoffen in de hersenen. Een andere veelbelovende richting voor toekomstig onderzoek is subtypering op basis van verschillende soorten gegevens, zoals genetische gegevens, hersenactiviteit, bloedwaardes én symptomen.

Dankwoord

Dit proefschrift zou niet tot stand zijn gekomen zonder de support van mijn familie, vrienden en collega's. Ik wil iedereen die me gesteund heeft deze afgelopen jaren ontzettend bedanken, maar een paar mensen in het bijzonder.

Allereerst wil ik mijn begeleiders bedanken. Zoals ik keer op keer met collega's, vrienden en familie heb gedeeld, voel ik me zeer bevoorrecht dat ik met zo'n geweldig team heb mogen werken. Ik had me geen betere copromotoren kunnen wensen dan Klaas Wardenaar en Hanna van Loo, die beiden oprecht geïnteresseerd waren - niet alleen in mijn onderzoek, maar ook in hoe het met mij ging. Onze wekelijkse bijeenkomsten waren over het algemeen erg nuttig, maar ook altijd erg gezellig. Ook mijn promotor, Robert Schoevers, ben ik erg dankbaar. Ondanks het feit dat zijn agenda sneller vol lijkt te raken dan een normaal mens kan werken, kwam hij altijd met interessante suggesties en ondersteunde hij mijn persoonlijke ontwikkeling waar nodig.

De afgelopen vier jaar zouden een stuk minder leuk zijn geweest zonder de vele collega's die ik op het ICPE heb leren kennen. Anna en Melissa, vind ik het jammer dat we onze laatste jaren niet samen op kantoor hebben doorgebracht, maar ik denk met veel plezier terug aan onze dagen op het ICPE.

Ik wil mijn familie bedanken voor hun onvoorwaardelijke liefde en steun. Ik heb de afgelopen vier jaar veel ups en downs meegemaakt, en ze waren er altijd voor me. Ik wil mijn ouders nogmaals bedanken voor het met mij delen van hun woonruimte, en voor het afstaan van hun slaapkamer tijdens kantooruren, zodat ik dit proefschrift kon afronden vanuit de geborgenheid van mijn ouderlijk huis.

Tot slot wil ik de Plum Village-gemeenschap bedanken, met name de Wake Up Sangha (Groningen en internationaal), en de Mind and Life/Mindful Researchers-gemeenschap, voor het gevoel van veiligheid en verbondenheid in deze moeilijke tijden. Dit netwerk is van onschatbare waarde geweest voor de realisatie van dit proefschrift en voor mijn persoonlijke groei.

This thesis is published within the **Research Institute SHARE** (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen. Further information regarding the institute and its research can be obtained from our website: http://www.share.umcg.nl/.

More recent theses can be found in the list below. Supervisors and co-supervisors are between brackets.

2021

Bochane MI
Uniform screening for atypical language development in Dutch child health care
(prof CP van der Schans, prof SA Reijneveld, dr MR Luinge)

Stegmann ME
Communication in oncology; positioning the general practitioner in shared decision making for patients with cancer
(prof MY Berger, dr AJ Berendsen)

Zhang Y
Toxic effects of heavy metals on the immune system; evidence from population and in vitro studies
(dr MM Faas, dr X Xu, dr X Hou, dr P de Vos)

Rienstra W
Beyond the joint; measurement and treatment of sensitisation in patients undergoing total knee or hip arthroplasty
(prof SK Bulstra, dr I van den Akker-Scheek, dr M Stevens)

Kuilman L
Antecedents of Ethical Decision Making by physician assistants and nurse practitioners: validation of instruments and their application
(prof P Roodbol, dr GJ Jansen, dr LB Mulder)

Waanders JB
Functional relevance of eccentric strength maintenance with age during walking
(prof T Hortobagyi, dr A Murgia, dr JR Franz)

Kok D
Memory metal in lumbar spinal fusion; biological, mechanical, clinical and radiological studies
(prof SK Bulstra, dr M Stevens, dr FH Wapstra)

Kristoffersen MB
Serious gaming for learning the intuitive, non-natural control of prosthetic hands
(prof CK van der Sluis, dr RM Bongers, dr A Murgia)

Noh JW
Key factors to improve maternal and child health in Sindh province, Pakistan
(prof J Stekelenburg, dr YM Kim, dr YD Kwon)

Fitria N
Pregnancy complications: health economics of screening and prevention
(prof MJ Postma, dr ADI van Asselt)

Zijlstra E
The impact of the hospital environment; understanding the experience of the patient journey
(prof CP van der Schans, prof M Hagedoorn, dr MP Mobach)

Feenstra M
Trajectories of health status in older people
(prof SEJA de Rooij, prof BC van Munster, dr N Smidt)

Roossien C
Sensors@work; towards monitoring physical workload for sustainable employability
(prof MF Reneman, prof GJ Verkerke)

Klerk R de
Development of a wheelchair propulsion laboratory
(prof LHV van der Woude, dr HEJ Veeger, dr RJK Vegter)

Zhou Y
Machine learning for identifying patterns in human gait; classification of age and clinical groups
(prof CJC Lamoth, prof T Hortobayi)

Brink MJ van den
Management of heavy menstrual bleeding
(prof MY Berger, prof MY Bongers, dr JH Dekker)

Brouwer MA
At the boundaries of life; suffering and decision-making in children with life-threatening conditions (1-12 years)
(prof AAE Verhagen, dr E Maeckelberghe)

Beune IM
Standardization in fetal growth restriction; progression by consensus
(prof JJ Erwich, dr SJ Gordijn, dr JW Ganzevoort)

Keizer MNJ
Neuromuscular control of knee laxity after an anterior cruciate ligament reconstruction
(prof E Otten, dr JM Hijmans, dr RW Brouwer)

For earlier theses visit our website.