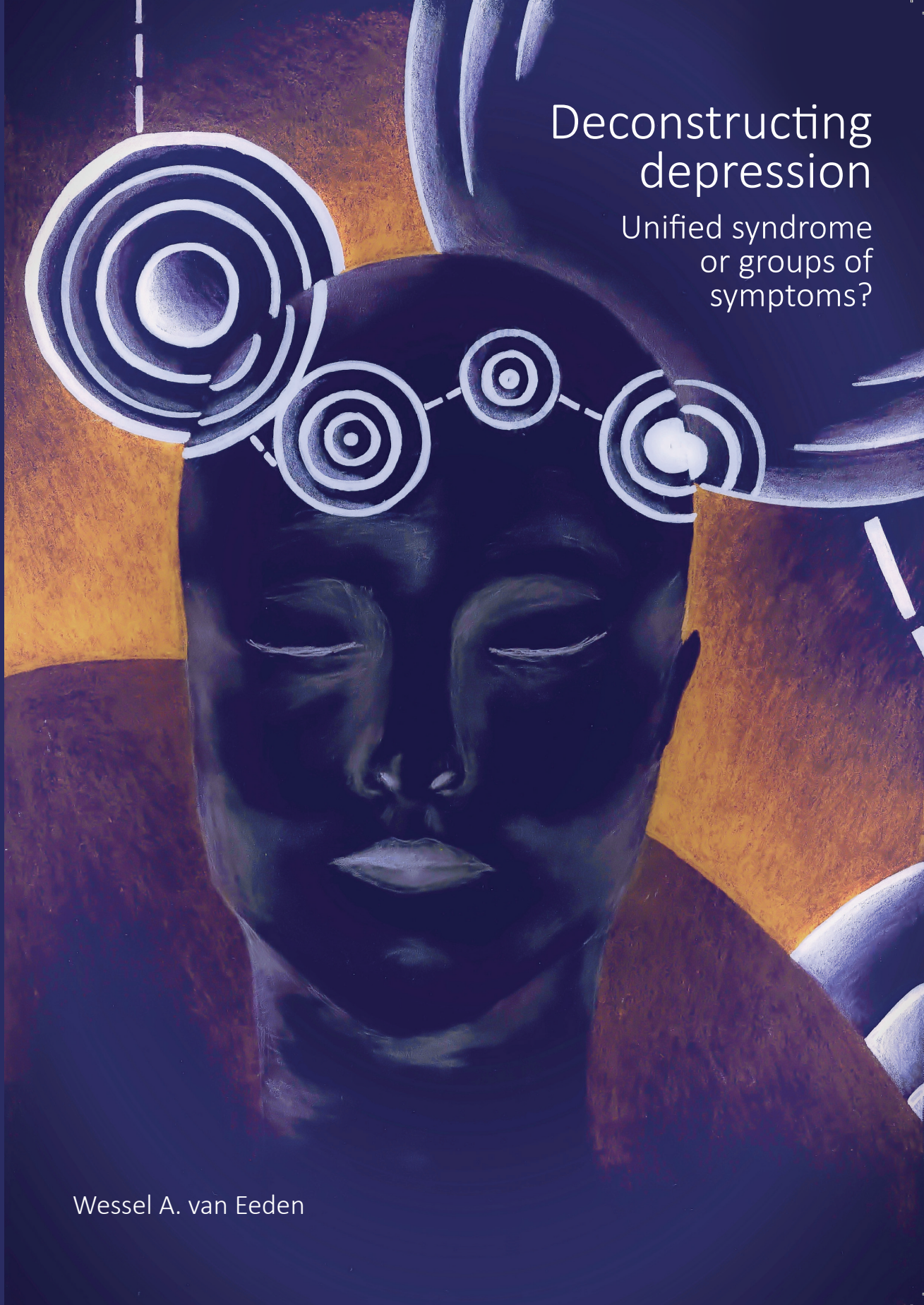
Universiteit Leiden
The Netherlands

Deconstructing depression: unified syndrome or groups of symptoms?

Eeden, W.A. van



Deconstructing depression: Unified syndrome or groups of symptoms?
Wessel A. van Eeden

Stellingen behorend bij het proefschrift

Deconstructing depression

Unified syndrome or groups of symptoms?

Wessel A. van Eeden

1. Somscores van depressieschalen doen geen recht aan de heterogeniteit van de symptomatologie van depressie en het beloop daarvan (dit proefschrift).
2. Inflammatie is niet gerelateerd aan depressie maar aan specifieke symptomen van depressie die gerelateerd zijn aan sickness-behaviour (dit proefschrift).
3. Het voorspellen van de prognose van depressie en angst middels hedendaagse machine learning biedt geen verbetering ten opzicht van de conventionele statistische methodieken (dit proefschrift).
4. Persoonlijkheidspathologie en persisterende depressie vertonen grote overlap bij het voorspellen van symptoombeloop (dit proefschrift).
5. Soms onbewust richten behandelaren en hun behandelingen zich meer op individuele symptomen dan op syndromen.
6. Een systematische review zou een onderdeel moeten zijn van elk promotietraject.
7. Meer wetenschappelijk onderzoek is urgent nodig om tot oplossingen te komen voor de lange wachtlijsten binnen de huidige GGZ.
8. Het symptoomprofiel kan bijdragen aan de selectie van patiënten die baat hebben bij anti-inflammatoire behandeling.
9. Geen diermodel kan ons helpen met betrekking tot de fenomenologie van depressie, want "If a rat is a good model for your emotional life, you're in big trouble." (Robert M. Sapolsky; Stress, Neurodegeneration and Individual Differences, 2001)
10. Door de complexiteit, ogenschijnlijke willekeur en chaos van menselijke emoties, motieven en gedrag zal machine learning dit onvoldoende kunnen voorspellen.
11. De emotionele rollercoaster van een PhD traject is verre van adaptief.

Deconstructing depression

Unified syndrome or groups of symptoms?

Wessel A. van Eeden

Deconstructing depression: Unified syndrome or groups of symptoms?

Wessel A. van Eeden
PhD thesis, Leiden Univertisy Medical Center, the Netherlands, 2022

Cover design: Lisa van der Windt
Printed by: ProefschriftMaken

Deconstructing depression

Unified syndrome or groups of symptoms?

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 29 september 2022
klokke 15:00 uur

door

Wessel Adrianus van Eeden
geboren te Delft
in 1992

Promotor	Prof. Dr. A.M. van Hemert
Co-promotoren	Dr. E.J. Giltay
	Dr. I.V.E. Carlier
Promotiecommissie	Prof. Dr. N.J.A. van der Wee
	Prof. Dr. F.P.M.L. Peeters (Maastricht University)
	Prof. Dr. B.G. Tiemens (Radboud Universiteit)
	Prof. Dr. A. Beekman (Amsterdam UMC, locatie VUmc)

Chapter 1	General introduction	p. 7
Chapter 2	Severity, course trajectory, and within-person variability of individual symptoms in patients with major depressive disorder – *Acta Psychiatrica Scandinavica*. 2019; 139(2): 194-205.	p. 29
Chapter 3	Neuroticism and chronicity as predictors of 9-year course of individual depressive symptoms – *Journal of Affective Disorders*. 2019; 252: 484-492.	p. 63
Chapter 4	Prognostic Value of Pathological Personality Traits for Treatment Outcome in Anxiety and Depressive Disorders: The Leiden Routine Outcome Monitoring Study – *The Journal of Nervous and Mental Disease.* 2022; 10-1097.	p. 89
Chapter 5	Basal and LPS-stimulated inflammatory markers and the course of individual symptoms of depression – *Translational Psychiatry*. 2020; 10(1): 235.	p. 123
Chapter 6	Basal and LPS-stimulated inflammatory markers and the course of anxiety symptoms – *Brain Behavior and Immunity*. 2021; 98: 378-387.	p. 171
Chapter 7	Predicting the 9-year course of mood and anxiety disorders with automated machine learning: a comparison between Auto-sklearn, naïve Bayes classifier, and traditional logistic regression – *Psychiatry Research*. 2021; 299: 113823.	p. 203
Chapter 8	Summary and general discussion	p. 239
Addendum	Nederlandse samenvatting	p. 261
	Dankwoord	p. 266
	List of publications	p. 267
	Curriculum Vitae	p. 268

Chapter 1

General Introduction

1.1 Preface

Depression has known many definitions over the course of time. Depression first appeared in writing in Mesopotamia in circa 1792 BC and was considered to be a consequence of demon possession causing distress, abnormal behavior and suicide [1]. Since then, our thoughts about cause and consequences of depression have drastically changed and depression is recognized today as a common, debilitating medical illness that affects how one feels, thinks and acts. Depression is described as a disorder in which the patient experiences feelings of sadness and/or a loss of interest in activities once enjoyed. It can lead to a variety of emotional and physical problems and impairs a person's ability to function at work and at home.

Depression is a substantial public health problem with a heavy burden for the patient, his/her caregivers and society [2-4]. It remains a disease with one of the highest burden, expressed in number of Disability Adjusted Life Years (DALY's; [5, 6]). A staggering 18.7% of the Dutch population has had a depression before the age of 65 [7]. Its economic impact is immense; the estimated cost to global economy is about 1 trillion euro in lost productivity per year [6, 8].

Over the past decades, much research has been conducted regarding its etiology, clinical characteristics, treatment, and course [9, 10]. Evidence-based guidelines for depression treatment have been developed consisting of psychotherapy (e.g. Cognitive Behavioural Therapy [CBT], interpersonal therapy [IPT]), pharmacotherapy (e.g. tricyclic antidepressants [TCA], selective serotonin reuptake inhibitor [SSRI]), or both [11]. These treatments are sufficient for the majority of patients, 50% of depressed patients recovers within 6 months and 76% recovers within 12 months. However, a substantial 20% suffers a chronic course and does not recover within 24 months, and a recurrent course is common [12].

One of the main challenges when researching or treating depression, has been the vast heterogeneity in etiology, symptomatology, and course of depression [13]. Two patients who both meet criteria for major depressive disorder (MDD) may have only few overlapping symptoms. For example, patient "A" is a 70 year old female who experienced psychotic symptoms of guilt, feelings of worthlessness, and a depressed mood. These symptoms started a few months after the loss of a loved one. She experienced exceeding psychological distress, but fortunately recovered with medical treatment within three months. Patient B is a 40 year old male who experienced less severe symptoms, mostly consisting of low energy,

psychomotor retardation and anhedonia. His symptoms were related to somatic problems, which despite treatment lasted for more then two years.

Given this heterogeneity, one may wonder whether depression actually constitutes a single unified disorder. Alternatively, it may be considered that not depression as a syndrome should be the topic of research but rather depression as a constellation of separate individual symptoms that together form the disease state of depression. Individual symptoms may have different risk factors and course trajectories, which remain unnoticed when depression is being researched as a unified syndrome. This heterogeneity formed the starting point for the present dissertation, which focuses on the separate symptoms of depression.

1.2 Diagnosing mood disorders and measuring symptom severity

Already in 1959, Karl Gustav Hempel wrote about the need for a progression from descriptive towards an explanatory classification in psychiatry [14]. Despite major research efforts during the past 70 years, this has not yet been achieved. Instead, disorders are "diagnosed", or rather classified, based on criteria that can be checked off by clinicians and researchers. Based on the Diagnostic and Statistical Manual of Mental Disorders, currently in its Fifth Edition (DSM-5), Major depressive disorder is a mood disorder that consists of at least five out of nine symptoms as presented in Table 1, of which a *depressed mood* or *anhedonia* (diminished interest or pleasure) must be present for at least two weeks. Classifications are mainly descriptive and with a few notable exceptions, such as post-traumatic stress disorder (PTSD), causal frameworks are excluded from the DSM classification [15]. Using this approach, disorders are classified regardless of underlying causality.

Table 1. DSM-5 criteria
Major depressive disorder (MDD)
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.
• Depressed most of the day, nearly every day as indicated by subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful) • Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by subjective account or observation) • Significant weight loss when not dieting or weight gain (e.g., change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day • Insomnia or hypersomnia nearly every day • Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down) • Fatigue or loss of energy nearly every day • Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick). • Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others) • Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide
The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
The episode is not attributable to the physiological effects of a substance or to another medical condition.
The occurrence of the major depressive episode is not better explained by schizoaffective disorder, schizophrenia, schizophreniform disorder, delusional disorder, or other specified and unspecified schizophrenia spectrum and other psychotic disorders.
There has never been a manic episode or a hypomanic episode.

Many psychiatric disorders cannot be sharply distinguished from each other, and disorders often seem to overlap, suggesting a shared etiology. DSM-5 classification is based on clinical consensus and does not assume that its categories represent distinct clinical entities with absolute borders [15]. Especially anxiety and depressive disorders often co-occur. There is a lifetime comorbidity between the two disorders of 43.6% for men and 55.8% for women [16].

These disorders can be treated with the same pharmacological interventions and have overlapping risk factors [17]. One can speculate if it is reasonable to approach mood- and anxiety disorders as two distinct entities. During the development of the DSM-5 and International Classification of Diseases, 11th version [ICD-11; 18], Goldberg, Krueger [19] stated that "*Mixed presentations of mood- and anxiety disorders may be the norm*".

Despite these difficulties, the DSM-5 could be considered a dictionary of mental disorders. It is of value that researchers and clinicians at least share a common international language, and that patients are classified by clinicians according to the same diagnostic criteria [20]. However, field trials demonstrated questionable (interrater) test-retest reliability of MDD with a kappa of 0.25 (95% confidence interval [CI]: 0.13–0.36). This was substantially lower than the other classifications that were assessed, such as borderline personality disorder (kappa: 0.34; 95% CI: 0.18–0.51) and alcohol use disorder (kappa: 0.40, 95% CI: 0.27–0.54) [21].

Another way of measuring depression and anxiety is by using self-report symptom measures. Self-report questionnaires are frequently used in mental healthcare to measure the severity and course of psychiatric disorders. The general format of self-report questionnaires is that frequency and severity of individual symptoms are scored with ordinal items, which are equally weighted and summed. These sum scores are thought to reflect the severity of the psychopathology and can be used as a tool for quantifying the patient's experienced symptoms. Hereby, self-reported measures can assist a clinician in the initial evaluation of patients [22], and when administered repeatedly, to monitor the effect of treatment [23]. Self-report questionnaires can be used to measure the general psychological distress/psychopathology [e.g., 48-item Symptom Questionnaire; SQ-48; 24] or symptoms of specific disorders such as depression [e.g., Inventory of Depressive Symptomatology self-report; IDS-SR; 25] or anxiety/fear [e.g., Beck's Anxiety Inventory; BAI; 26; Fear Questionnaire; FQ , 27].

1.3 Symptoms versus syndromes

A syndrome is defined by the Cambridge dictionary as *"a combination of medical problems that shows the existence of a particular disease or mental condition"* [28]. Most research thus far has focused on depression as a syndrome, assuming a single underlying disorder or construct [e.g. 26, 29, 30]. In this line of thought, individual symptoms are reflections of the

underlying latent depression construct or disease. When the disease severity declines, one assumes that the severity of its individual symptoms would also decline more or less together. The line of reasoning is similar to any medical disease, where it is assumed that symptoms improve as the underlying causal disease entity resolves. This model works rather well for other diseases, such as COVID-19 where symptoms of fever, coughing, stuffy nose and loss of smell and taste are a direct reflection of the underlying viral infection. Subsequently, when the SARS-CoV-2 virus infection has successfully been defeated by the immune system, COVID-19 symptoms generally resolve, although sometimes more permanent damage may continue. However, these assumptions may be problematic when it comes to depression. Although the SARS-CoV-2 virus can directly be measured as the pathological agent, this is not the case for depression. Moreover, although the nine main symptoms of depression correlate with each other, depression has highly different presentations among individuals and it seems unlikely that depression symptoms are interchangeable measurements of one latent construct of depression [31].

Depression symptomatology show substantial variability among individual patients, but also within the same person over time. Studies that did assess symptom-specific course trajectories have shown important average differences among the trajectories of individual symptoms. Symptoms differ in their response to treatment, with sleep problems, energy loss, and cognitive problems most often reported as showing a residual (and potentially pre-existing) course[32-37]. This is something that one would not expect if these symptoms are (equally) related to one underlying depression syndrome. If these symptoms would reflect one underlying latent disorder, they would demonstrate a more or less synchronized decrease when the patient would recover from the underlying disorder. Given the heterogeneous nature of depression it can be doubted whether depressive disorder, or at the same token anxiety disorders, do in fact exist as unified latent syndromes.

Multiple attempts for constructing more homogeneous subtypes of depression have failed. The proposed subtypes have been based on symptom profiles (melancholic depression, psychotic depression etc.[38]); etiologically-based (early trauma depression, organic depression, drug induced depression, etc.); time of onset-based (early and late onset depression, seasonal depression, etc.); treatment resistant depression subtype; symptom severity (mild, moderate, or severe); or several biology driven subtypes [39, 40]. In fact, a

meta-review of 754 reviews published between 2000 and 2011 identified 15 commonly used subtypes of major depressive disorder. Despite the vast amount of literature about this subject, no clear differences in clinical presentation or long-term outcomes between the different subtypes can consistently be replicated [39, 41]. Moreover, although the different subtypes attempt to overcome the non-specificity of major depressive disorder, no differential impacts of causes and treatments are found [40]. Rather, focusing on (the complex relationships among) individual symptoms may be our way forward [42-44].

Individual symptoms can be conceived as separate entities that may each have their own (genetic) etiology [45, 46]. This idea has been supported by evidence of a population based twin registry (3084 pairs), which found that major depressive disorder did not reflect a single dimension of genetic liability [47]. Instead, three underlying dimensions were found that index genetic vulnerability for cognitive/psychomotor, mood, and neurovegetative symptoms. Though replication is needed, this suggests that individual symptoms (or symptom domains) may have separate genetic etiology.

Although related to, but not entirely within the scope of the present dissertation, there can also be causal relationships between individual symptoms. Within the field of psychiatry and clinical psychology, this is also known as the *network-theory*. For example, sleeping problems may cause low energy, which in turn may cause concentration problems. Or feelings of worthlessness and guilt cause a depressed mood which causes anhedonia which eventually causes suicidal ideations. This explains why depression as a syndrome can have different etiologies among patients, although individual symptoms may be correlated to each other. Taken together, it can be fruitful to study depression not as a distinct unified syndrome or latent construct, but rather as a heterogeneous group of loosely related symptoms.

1.4 Predicting the course of depression and anxiety

Improving the ability to predict the onset and course of mood and anxiety disorders is of clinical relevance for prevention, early detection, staging, and personalized treatments [48]. However, despite a large body of epidemiological research, the course and onset of mood and anxiety disorders remain difficult to predict.

Several studies demonstrated that individual symptoms also may have different risk factors [49-52]. In a comprehensive study by Lux and Kendler [53], the relation between 25 risk

factors (including demographics, psychiatric history, personality and life events) and the nine DSM symptoms of depression were analysed. The relationships proved surprisingly complex. Some risk factors proved to have a specific relation to certain symptoms, other "risk" factors could be positively associated to some symptoms but negatively to others. If all depressive symptoms are caused by an underlying disorder, symptoms are expected to have similar risk factors because risk factors are supposed to influence the liability to develop depression, not specific symptoms [54]. However this is not the case, suggesting that research on depression risk could benefit from a symptom-specific approach instead of focusing on depression as a syndrome.

Several predictive variables have been established for predicting the onset and course of depression and anxiety [e.g., 55, 56]. These variables include demographic characteristics (e.g. gender, age, and socioeconomic status), clinical characteristics (e.g. symptomatology and preceding course), personality characteristics (e.g. neuroticism and personality psychopathology), and biological variables (e.g. somatic disease and inflammation). In the present dissertation, we researched the predictive values of preceding chronicity, personality traits, inflammation and how these relate to the individual symptoms of depression. Moreover, we assessed if more advanced statistical methods could improve the accuracy for predicting the onset and course of depression and anxiety.

1.4.1 Psychiatric history: preceding chronicity

The preceding course is one of the most important predictive factors for depression. Acknowledging the Importance of a preceding depressive course led to the addition of persistent depressive disorder (i.e. a combination of dysthymia *and* chronic depression) in the DSM-5 [57]. Patients meet the criteria of this disorder when they experience a depressed mood for most of the day, for more days than not, for at least 2 years. When patients experience a depression for two years or more, they are likely to have an unfavourable course in the future. In this way, chronic depression forms one of the strongest prognostic predictors [58]. A previous analysis in the Netherlands Study of Depression and Anxiety (NESDA) demonstrated that depression persisted over the course of 4 years in 53% of the patients with chronic depression at baseline versus 27.8% of patients with nonchronic depression at baseline; which is consistent with findings from others [55, 58-62]. It is currently unknown

whether chronicity affects the course of all symptoms equally, or rather individual symptoms at varying magnitudes.

1.4.2 Personality: traits as prognostic factor

Personality pathology (PP) is another well-established risk factor for onset and unfavourable course of depression. The best established dimensional measure of personality as a risk factor for depression is the construct of Neuroticism. Neuroticism is one of the five major dimensions of personality (Five Factor Model; FFM; Costa and McCrae, 1992). This trait reflects the disposition to experience negative affect, including anger, anxiety, irritability, emotional instability, low mood, and self-consciousness [63]. Persons with high levels of neuroticism respond poorly to psychosocial stress, interpret ordinary situations as threatening, and they can experience minor frustrations overwhelmingly as hopeless. A substantial body of research to support its heritability, childhood antecedents, temporal stability across the life span, and universal presence [64, 65].

Other personality psychopathology constructs have been established as well, such as the tendency to be paranoid, avoidant or dissocial behavior [for example measured with the Dimensional Assessment of Dimensional Psychopathology Self-Report; DAPP-SF; 66]. Personality pathology has been strongly linked to psychiatric disorders, such as depressive and anxiety disorders [67-69]. The risk of comorbid personality disorders for major depressive disorder has been estimated at 45% [70], and ranges from 35% to 52% for anxiety disorders. High personality pathology, increases the risk of depression, its unfavourable course, and a higher relapse rate [42, 43, 68, 71]. However, the effects of personality pathology on treatment outcome may be substantially lower when taking baseline symptom level into account, usually interpreted as severity. Surprisingly, the likely intermediary effects (either as a mediator variables, or a moderator variables) of baseline depression severity on the relationship between personality pathology and treatment outcome have received little attention in the current literature [72-74].

Given the heterogeneous nature of depression, focusing on individual symptoms may yield novel insights into the relationship with personality traits, and the course of depression [42-44]. Although one study has demonstrated that neuroticism was related to all nine depression symptoms [51], another found specific associations with appetite/weight and sleeping

problems [75]. Assessing the link between personality and individual symptoms might increase our ability to target (psychotherapeutic) treatment strategies earlier and, more specifically, on certain symptom patterns.

1.4.3 Somatic factors: low-grade inflammation and mood states

Most somatic diseases or infections cause physical symptoms, but also mental- and behavioural changes. These behavioural changes are defined as "*sickness behaviour*". Sickness behaviour is generally regarded as an organized group of reward-oriented behavioural and motivational changes that accompany inflammation and infections [76-79]. Sickness behavior symptoms show considerable overlap with depressive symptoms (e.g., anhedonia, anorexia, low concentration, low energy, low libido, psychomotor slowness, and irritability). Therefore, researchers have hypothesized that depression is a maladaptive or exacerbated form of sickness behavior in some patients with chronic low-grade inflammation [76-78, 80-82]. Besides their reward-sensitivity related symptoms, recent studies suggest that also trauma- and anxiety-related symptoms are related to inflammatory markers, resulting in a mix of overlapping symptoms of mood, anxiety, and post-traumatic stress disorder [83-86]. Researchers have theorized that sickness behaviour holds some evolutionary advantages and has protective mechanisms for the individual, because it preserves energy resources needed for healing infection or other diseases and it may help prevent the transmission of its potential infectious agent [78, 79]. During an inflammatory response, the innate and adaptive immune systems are activated and pro-inflammatory cytokines are produced. The causal chain may involve somatic triggers inducing an inflammatory response, followed by sickness behavior. Sickness behavior in turn overlaps with- and induces depression, with additional positive feedback loops between (neuro) inflammation and (neuro) degenerative processes [78, 79, 82]. Evidence from meta-analyzes suggests that depressed patients have higher circulating concentrations of acute-phase proteins and pro-inflammatory cytokines compared to healthy subjects [87-92]. Also inflammatory markers after *ex vivo* induction of lipopolysaccharide (the cell membrane of Gram-negative bacteria), demonstrated significant associations with depression. However, other studies in this context have not found significant associations [91, 92]. A causal pathway in which inflammation causes symptoms of anxiety has been less established [93-95].

Inflammatory markers and depression have been linked, but effect sizes were generally small with limited clinical relevance for the individual patient [79, 96]. Low-grade inflammation may only be strongly linked to a subset of depressive symptoms that overlap with sickness behaviour [97, 98], and therefore may only be involved in the pathogenesis of a subset of depressed patients. The effect size between inflammation and individual symptoms may be higher than depression as a syndrome. Identifying associations between pro-inflammatory markers and specific depressive symptoms could be important for the advance of personalized medicine [96]. Nevertheless, few clinical studies have analyzed whether inflammatory markers are associated with specific depressive symptoms [96, 97, 99, 100]. Moreover, prospective studies regarding anxiety symptom severity remain scarce.

1.4.4 Methods: improving accuracy with innovative statistical methods

One way to increase the predictive value of these variables may be to use multivariate statistical models. Most clinical data thus far have been analyzed by using data modelling methods (such as regression analysis) and selecting only a collection of selected predictors. It is possible that more complex (including nonlinear and higher dimensional) patterns exist in the data, which can efficiently be detected when analyzing all available data simultaneously using machine learning approaches [101, 102]. These approaches are able to examine huge numbers of potential predictors, such as current individual symptoms, in an unbiased manner while preventing overfitting [103]. Machine learning may be more time efficient, better suited for large and complex datasets, and better able to detect complex patterns in the data than current data-modelling [104]. These advanced methods may be better suited to handle depression heterogeneity. Moreover, machine learning incorporates less human decision making than traditional methods, and could be suited for full automatization [104-106].

Over the past decade more modern techniques of machine learning have also been applied in the field of psychiatry. Thus far, machine learning studies in the field of psychiatry have been promising. A recent meta-analysis, which included 20 studies that predicted the therapeutic outcome of depression using machine learning algorithms, found an overall accuracy of .82 (95% confidence interval [CI] .77–.87; Lee et al., 2018). However, recently published papers have demonstrated only limited added value of machine learning over traditional regression analyzes [107, 108]. Additionally, other studies found that when predicting suicide, machine learning did not outperform regression analysis and resulted in

positive predictive values below 0.01, thus limiting the practical utility of these predictions [109, 110]. Despite the increasing number of publications in this field, machine learning has yet to move towards clinical application [111].

1.5 Aim, research questions and hypothesis of this dissertation

The present dissertation aims to expand our knowledge of depression by researching the symptom-specific longitudinal characteristics, risk-factors, and methods of analyzing depression severity in which individual symptoms are taken into account. This dissertation will mainly focus on depression, although anxiety has been studied as well, as anxiety is highly prevalent in patients with depression and share a common etiology. This brings us to the following research questions:

Main research question:

Can major depressive disorder be characterized as a unified syndrome?

This can be divided in the following sub-questions and hypotheses:

1. Is the course of individual depressive symptoms uniform over time?

We hypothesized that depression is a disorder with substantial within-person heterogeneity between symptoms in terms of intercepts, slopes, and variability.

2. Are individual symptoms of depression related to the same risk factors?

We hypothesized that risk factors are associated with the course of specific symptoms, rather than depression as a homogeneous construct, with similar associations for each symptom. We hypothesized that low-grade inflammation inflammatory markers demonstrate the strongest associations with symptoms that overlap with sickness behaviour.

3. Are advanced statistical methods more adequate to handle depression heterogeneity?

We hypothesized that machine learning techniques are better in detecting complex patterns in the data and would outdo traditional regression analysis techniques and achieve higher levels of accuracy when predicting the course and onset of depression and anxiety. Moreover, we hypothesized that machine learning would be particularly effective when symptom-specific features of current depression and anxiety are included to predict future disorders.

1.6 Cohorts used in this dissertation

This thesis is built on pre-existing data from the Netherlands Study of Depression and Anxiety (NESDA; Chapters 2, 3, 5, 6, 7) and the Leiden Routine Outcome Monitoring Study (ROM; Chapters 4).

NESDA is an ongoing multi-site naturalistic longitudinal cohort study, which aims to investigate the course and consequences of depressive and anxiety disorders. The first wave (baseline) lasted from 2004 to 2007, and the sixth wave of measurement at the 9-year follow up finished in 2016. NESDA is a cohort study that recruited from the community (*n* = 564; 18.9%), general practice (*n* = 1,610; 54.0%), and secondary mental healthcare [n=807; 27.1%; 112]. It includes patients with a current or lifetime depressive or anxiety disorder as well as healthy controls. By applying only few exclusion criteria, NESDA aimed for a cohort that is representative for diverse populations of healthy controls and patients with depression and anxiety [112].

The Leiden Routine Outcome Monitoring Study (ROM) is an ongoing prospective cohort study of the Leiden University Medical Center (Department of Psychiatry, in cooperation with mental health care provider GGZ Rivierduinen) which was carried out to assess treatment progress/outcome for patients with mood, anxiety, and/or somatoform disorders in a naturalistic setting [113]. The first assessment occurred during an intake procedure; research nurses interviewed patients using the Mini International Neuropsychiatric Interview-Plus [MINI-Plus; 114]. Additionally, patients completed a number of self-report questionnaires which were repeated during treatment. ROM data are collected systematically to assess treatment effectiveness in everyday clinical practice, to inform clinicians and patients about treatment progress (Carlier et al., 2012b; Lambert, 2017; Lambert et al., 2018).

1.7 Dissertation outline

In **chapter 2**, we assessed the longitudinal symptom-specific course trajectories and within-person variability of major depressive disorder over a 9-year period (NESDA data). More specifically, we aimed to answer which symptoms have clinically favourable characteristics and which show a more persistent course. We addressed some of the methodological gaps in earlier studies, by assessment of within-person variability over time, in which repeated measures are nested within persons.

In **chapter 3**, we examined whether preceding chronic depression, defined as being depressed for at least 2 years (during a patient's past 4 years before baseline) and level of neuroticism could predict the 9-year trajectory of individual depressive symptoms (NESDA data). In particular, the focus was on the symptom-specific differences.

In **chapter 4**, the prognostic value of a broad range of dimensional personality pathology on treatment outcome among patients with depression and/or anxiety disorders is investigated (ROM data). We hereby assessed the potential mediating and/or moderating effects of baseline symptom level.

In **chapter 5,** we examined the associations between basal levels and LPS-induced inflammatory markers and individual depressive symptoms over the course of 9 years. Inflammation has been repeatedly linked to depression, presumably as a consequence of sickness-behavior. However, not all depression symptoms may be related to sickness behavior equally.

In **chapter 6**, we extended our research on inflammation by also examining whether basal as well as LPS-induced inflammatory markers determined at baseline are associated with the course of domains of anxiety symptomatology.

Chapter 7 has a more methodological approach. In this chapter, we assessed whether using more complex statistical methods would be better suited for dealing with the complexity of depression heterogeneity. We compared the performance of three methods: traditional multinominal logistic regression, a basic probabilistic machine learning algorithm (naïve Bayesian classifier [115], and a more advanced automated machine learning method (Auto-sklearn [106]) to predict DSM-IV-TR psychiatric diagnoses at 2-, 4-, 6-, and 9-year follow up with different sets of predictors, including current individual symptoms.

Finally, in **chapter 8**, we summarized the general conclusions of this dissertation and discuss the strengths and limitations of our studies. In addition, we discuss the clinical implications derived from our studies and recommendations for further research.

References

1. Reynolds, E.H. and J.V.K. Wilson, *Depression and anxiety in Babylon.* Journal of the Royal Society of Medicine, 2013. **106**(12): p. 478-481.
2. Buist-Bouwman, M., et al., *Functional disability of mental disorders and comparison with physical disorders: a study among the general population of six European countries.* Acta Psychiatrica Scandinavica, 2006. **113**(6): p. 492-500.
3. Spijker, J., et al., *Functional disability and depression in the general population. Results from the Netherlands Mental Health Survey and Incidence Study (NEMESIS).* Acta Psychiatrica Scandinavica, 2004. **110**(3): p. 208-214.
4. Üstün, T.B., et al., *Global burden of depressive disorders in the year 2000.* The British journal of psychiatry, 2004. **184**(5): p. 386-392.
5. Murray, C.J., et al., *Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010.* The lancet, 2012. **380**(9859): p. 2197-2223.
6. Trautmann, S., J. Rehm, and H.U. Wittchen, *The economic costs of mental disorders: Do our societies react appropriately to the burden of mental disorders?* EMBO reports, 2016. **17**(9): p. 1245-1249.
7. de Graaf, R., et al., *Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2.* Social psychiatry and psychiatric epidemiology, 2012. **47**(2): p. 203-213.
8. Chisholm, D., et al., *Scaling-up treatment of depression and anxiety: a global return on investment analysis.* The Lancet Psychiatry, 2016. **3**(5): p. 415-424.
9. Barrera, A.Z., L.D. Torres, and R.F. Munoz, *Prevention of depression: the state of the science at the beginning of the 21st Century.* International Review of Psychiatry, 2007. **19**(6): p. 655-670.
10. Cuijpers, P., *Four decades of outcome research on psychotherapies for adult depression: An overview of a series of meta-analyses.* Canadian Psychology/psychologie canadienne, 2017. **58**(1): p. 7.
11. Middleton, H., et al., *NICE guidelines for the management of depression*. 2005, British Medical Journal Publishing Group.
12. Spijker, J., et al., *Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS).* The British journal of psychiatry, 2002. **181**(3): p. 208-213.
13. Parker, G., *Beyond major depression.* Psychological medicine, 2005. **35**(4): p. 467.
14. Hempel, C.G., *The philosophy of Carl G. Hempel: studies in science, explanation, and rationality*. 2001: Oxford University Press.
15. American-Psychiatric-Association, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. 2013: American Psychiatric Pub.
16. De Graaf, R., et al., *Temporal sequencing of lifetime mood disorders in relation to comorbid anxiety and substance use disorders.* Social psychiatry and psychiatric epidemiology, 2003. **38**(1): p. 1-11.
17. Mathew, A., et al., *Co-morbidity between major depressive disorder and anxiety disorders: shared etiology or direct causation?* Psychological medicine, 2011. **41**(10): p. 2023.
18. First, M.B., et al., *The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders.* World Psychiatry, 2015. **14**(1): p. 82-90.
19. Goldberg, D., et al., *Emotional disorders: Cluster 4 of the proposed meta-structure for DSM-V and ICD-11.* 2009.
20. Brown, T.A., et al., *Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders.* Journal of abnormal psychology, 2001. **110**(1): p. 49.
21. Regier, D.A., et al., *DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses.* American journal of psychiatry, 2013. **170**(1): p. 59-70.
22. Demetriou, C., B.U. Ozer, and C.A. Essau, *Self-report questionnaires.* The encyclopedia of clinical psychology, 2014: p. 1-6.
23. Carlier, I.V., et al., *Routine outcome monitoring and feedback on physical or mental health status: evidence and theory.* Journal of Evaluation in Clinical Practice, 2012. **18**(1): p. 104-110.
24. Carlier, I., et al., *Development and validation of the 48-item Symptom Questionnaire (SQ-48) in patients with depressive, anxiety and somatoform disorders.* Psychiatry research, 2012. **200**(2): p. 904-910.

25. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): Psychometric properties.* Psychol. Med., 1996. **26**(3): p. 477-486.
26. Beck, A.T., R.A. Steer, and M.G. Carbin, *Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation.* Clinical psychology review, 1988. **8**(1): p. 77-100.
27. Marks, I.M. and A.M. Mathews, *Brief standard self-rating for phobic patients.* Behaviour research and therapy, 1979. **17**(3): p. 263-267.
28. *Cambridge dictionary*. Available from: https://dictionary.cambridge.org/.
29. Trivedi, M.H., et al., *The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation.* Psychological Medicine, 2004. **34**(1): p. 73-82.
30. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): Psychometric properties.* Psychological Medicine, 1996. **26**(3): p. 477-486.
31. Fried, E.I. and R.M. Nesse, *The Impact of Individual Depressive Symptoms on Impairment of Psychosocial Functioning.* PLoS One, 2014. **9**(2).
32. Madhoo, M. and S.Z. Levine, *Initial Severity Effects on Residual Symptoms in Response and Remission: A STAR* D Study During and After Failed Citalopram Treatment.* Journal of clinical psychopharmacology, 2015. **35**(4): p. 450-453.
33. Menza, M., H. Marin, and R.S. Opper, *Residual symptoms in depression: can treatment be symptom-specific?* Journal of Clinical Psychiatry, 2003. **64**(5): p. 516-523.
34. Nil, R., S. Lütolf, and E. Seifritz, *Residual symptoms and functionality in depressed outpatients: A one-year observational study in Switzerland with escitalopram.* Journal of Affective Disorders, 2016. **197**: p. 245-250.
35. Conradi, H.J., J. Ormel, and J.P. De, *Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study.* Psychological Medicine, 2011. **41**(6): p. 1165-1174.
36. Conradi, H., J. Ormel, and P. de Jonge, *Symptom profiles of the DSM-IV-defined remission, recoverym relapse, and recurrence of depression: the role of the core symptoms.* Depression and Anxiety, 2012. **29**(7): p. 638-645.
37. Romera, I., et al., *Residual symptoms and functioning in depression, does the type of residual symptom matter? A post-hoc analysis.* BMC Psychiatry, 2013. **13**: p. 51.
38. Martino, D.J., et al., *Melancholia: an attempt at definition based on a review of empirical data.* The Journal of nervous and mental disease, 2019. **207**(9): p. 792-798.
39. Beijers, L., et al., *Data-driven biological subtypes of depression: systematic review of biological approaches to depression subtyping.* Molecular psychiatry, 2019. **24**(6): p. 888-900.
40. Harald, B. and P. Gordon, *Meta-review of depressive subtyping models.* Journal of affective disorders, 2012. **139**(2): p. 126-140.
41. Kessing, L., *Epidemiology of subtypes of depression.* Acta Psychiatrica Scandinavica, 2007. **115**: p. 85-89.
42. Kotov, R., et al., *Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis*. 2010, American Psychological Association.
43. Malouff, J.M., E.B. Thorsteinsson, and N.S. Schutte, *The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis.* Journal of Psychopathology and Behavioral Assessment, 2005. **27**(2): p. 101-114.
44. Fried, Nesse, and Randolph, *Depression sum-scores don't add up: why analyzing specific depression symptoms is essential.* BMC medicine, 2015. **13**(1): p. 72.
45. Borsboom, D., et al., *The small world of psychopathology.* PLoS One, 2011. **6**(11): p. e27407.
46. Schmittmann, V.D., et al., *Deconstructing the construct: A network perspective on psychological phenomena.* New ideas in psychology, 2013. **31**(1): p. 43-53.
47. Kendler, K.S., S.H. Aggen, and M.C. Neale, *Evidence for multiple genetic factors underlying DSM-IV criteria for major depression.* JAMA psychiatry, 2013. **70**(6): p. 599-607.
48. McGorry, P.D., *Risk syndromes, clinical staging and DSM V: new diagnostic infrastructure for early intervention in psychiatry.* Schizophrenia research, 2010. **120**(1): p. 49-53.
49. Keller, M.C., M.C. Neale, and K.S. Kendler, *Association of different adverse life events with distinct patterns of depressive symptoms.* Am J Psychiatry, 2007. **164**(10): p. 1521-1529.
50. Fried, E.I., et al., *The differential influence of life stress on individual symptoms of depression.* Acta Psychiatrica Scandinavica, 2015. **131**(6): p. 465-471.

51. Fried, et al., *Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors.* Psychol. Med., 2014. **44**(10): p. 2067-2076.
52. Kendler, K.S. and S.H. Aggen, *Symptoms of major depression: Their stability, familiality, and prediction by genetic, temperamental, and childhood environmental risk factors.* Depress Anxiety, 2017. **34**(2): p. 171-177.
53. Lux, V. and K. Kendler, *Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria.* Psychological medicine, 2010. **40**(10): p. 1679.
54. Fried, E.I., et al., *Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors.* Psychol. Med., 2014. **44**(10): p. 2067-2076.
55. Colman, I., et al., *Predictors of long-term prognosis of depression.* Canadian Medical Association Journal, 2011. **183**(17): p. 1969-1976.
56. Kelly, K.M. and B. Mezuk, *Predictors of remission from generalized anxiety disorder and major depressive disorder.* J Affect Disord, 2017. **208**: p. 467-474.
57. Association, A.P., *Diagnostic and statistical manual of mental disorders (DSM-5®)*. 2013: American Psychiatric Pub.
58. Boschloo, L., et al., *The Four-Year Course of Major Depressive Disorder: The Role of Staging and Risk Factor Determination.* Psychotherapy and Psychosomatics, 2014. **83**(5): p. 279-288.
59. Coryell, W., J. Endicott, and M. Keller, *Outcome of patients with chronic affective disorder: a five-year follow-up.* The American journal of psychiatry, 1990. **147**(12): p. 1627.
60. Mynors-Wallis, L. and D. Gath, *Predictors of treatment outcome for major depression in primary care.* Psychological medicine, 1997. **27**(3): p. 731-736.
61. Stegenga, B.T., et al., *The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study.* Social psychiatry and psychiatric epidemiology, 2012. **47**(1): p. 87-95.
62. Trivedi, M.H., et al., *Evaluation of outcomes with citalopram for depression using measurement-based care in STAR* D: implications for clinical practice.* American journal of Psychiatry, 2006. **163**(1): p. 28-40.
63. Leary, M.R. and R.H. Hoyle, *Handbook of individual differences in social behavior*. 2009: Guilford Press.
64. Widiger, T.A. and J.R. Oltmanns, *Neuroticism is a fundamental domain of personality with enormous public health implications.* World Psychiatry, 2017. **16**(2): p. 144.
65. Widiger and Costa, *Integrating normal and abnormal personality structure: the five-factor model.* Journal of Personality, 2012. **80**(6): p. 1471-1506.
66. van Kampen, D., E. de Beurs, and H. Andrea, *A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF.* Psychiatry Research, 2008. **160**(1): p. 115-128.
67. Bienvenu, O.J., et al., *Normal personality traits and comorbidity among phobic, panic and major depressive disorders.* Psychiatry research, 2001. **102**(1): p. 73-85.
68. Shea, M.T. and S. Yen, *Personality traits/disorders and depression: A summary of conceptual and empirical findings.* 2005.
69. Friborg, O., et al., *Comorbidity of personality disorders in anxiety disorders: A meta-analysis of 30 years of research.* Journal of affective disorders, 2013. **145**(2): p. 143-155.
70. Friborg, O., et al., *Comorbidity of personality disorders in mood disorders: a meta-analytic review of 122 studies from 1988 to 2010.* Journal of affective disorders, 2014. **152**: p. 1-11.
71. Schmitz, N., J. Kugler, and J. Rollnik, *On the relation between neuroticism, self-esteem, and depression: results from the National Comorbidity Survey.* Comprehensive psychiatry, 2003. **44**(3): p. 169-176.
72. Bos, E.H., et al., *Effectiveness of systems training for emotional predictability and problem solving (STEPPS) for borderline personality problems in a 'real-world'sample: Moderation by diagnosis or severity?* Psychotherapy and psychosomatics, 2011. **80**(3): p. 173-181.
73. Ball, S.A., et al., *Reliability of personality disorder symptoms and personality traits in substance-dependent inpatients.* Journal of abnormal psychology, 2001. **110**(2): p. 341.
74. Sahin, Z., et al., *Clinical severity as a moderator of outcome in psychodynamic and dialectical behavior therapies for borderline personality disorder.* Personality Disorders: Theory, Research, and Treatment, 2018. **9**(5): p. 437.
75. Lux, V. and K.S. Kendler, *Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria.* Psychol Med, 2010. **40**(10): p. 1679-90.
76. Shattuck, E.C. and M.P. Muehlenbein, *Human sickness behavior: Ultimate and proximate explanations.* American Journal of Physical Anthropology, 2015. **157**(1): p. 1-18.
77. Miller, A.H., L. Capuron, and C.L. Raison, *Immunologic influences on emotion regulation.* Clinical Neuroscience Research, 2005. **4**(5-6): p. 325-333.

78. Dantzer, R. and K.W. Kelley, *Twenty years of research on cytokine-induced sickness behavior.* Brain, behavior, and immunity, 2007. **21**(2): p. 153-160.
79. Miller, A.H. and C.L. Raison, *The role of inflammation in depression: from evolutionary imperative to modern treatment target.* Nat Rev Immunol, 2016. **16**(1): p. 22-34.
80. Vollmer-Conna, U., et al., *Production of pro-inflammatory cytokines correlates with the symptoms of acute sickness behaviour in humans.* Psychological medicine, 2004. **34**(7): p. 1289-1297.
81. Smith, A.P., *Effects of the common cold on mood, psychomotor performance, the encoding of new information, speed of working memory and semantic processing.* Brain, behavior, and immunity, 2012. **26**(7): p. 1072-1076.
82. Maes, M., et al., *Depression and sickness behavior are Janus-faced responses to shared inflammatory pathways.* BMC Med, 2012. **10**(1): p. 66.
83. Felger, J.C., *Imaging the Role of Inflammation in Mood and Anxiety-related Disorders.* Curr Neuropharmacol, 2018. **16**(5): p. 533-558.
84. Naude, P.J.W., et al., *Anxiety disorders and CRP in a population cohort study with 54,326 participants: The LifeLines study.* World J Biol Psychiatry, 2018. **19**(6): p. 461-470.
85. Vogelzangs, N., et al., *Cytokine production capacity in depression and anxiety.* Translational psychiatry, 2016. **6**(5): p. e825.
86. Renna, M.E., et al., *The association between anxiety, traumatic stress, and obsessive–compulsive disorders and chronic inflammation: A systematic review and meta-analysis.* Depression and anxiety, 2018. **35**(11): p. 1081-1094.
87. Dowlati, Y., et al., *A meta-analysis of cytokines in major depression.* Biological psychiatry, 2010. **67**(5): p. 446-457.
88. Howren, M.B., D.M. Lamkin, and J. Suls, *Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis.* Psychosomatic medicine, 2009. **71**(2): p. 171-186.
89. Liu, Y., R.C.-M. Ho, and A. Mak, *Interleukin (IL)-6, tumour necrosis factor alpha (TNF-α) and soluble interleukin-2 receptors (sIL-2R) are elevated in patients with major depressive disorder: a meta-analysis and meta-regression.* Journal of affective disorders, 2012. **139**(3): p. 230-239.
90. 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, 2014. **71**(12): p. 1381-1391.
91. Valkanova, V., K.P. Ebmeier, and C.L. Allan, *CRP, IL-6 and depression: a systematic review and meta-analysis of longitudinal studies.* Journal of affective disorders, 2013. **150**(3): p. 736-744.
92. Haapakoski, R., et al., *Cumulative meta-analysis of interleukins 6 and 1β, tumour necrosis factor α and C-reactive protein in patients with major depressive disorder.* Brain, behavior, and immunity, 2015. **49**: p. 206-215.
93. Bierhaus, A., et al., *A mechanism converting psychosocial stress into mononuclear cell activation.* Proceedings of the National Academy of Sciences, 2003. **100**(4): p. 1920-1925.
94. Pace, T.W., et al., *Increased stress-induced inflammatory responses in male patients with major depression and increased early life stress.* Am J Psychiatry, 2006. **163**(9): p. 1630-3.
95. Glaus, et al., *The bidirectional relationship between anxiety disorders and circulating levels of inflammatory markers: Results from a large longitudinal population-based study.* Depression and anxiety, 2018. **35**(4): p. 360-371.
96. Köhler-Forsberg, O., et al., *Association between C-reactive protein (CRP) with depression symptom severity and specific depressive symptoms in major depression.* Brain, Behavior, and Immunity, 2017. **62**: p. 344-350.
97. Lamers, et al., *Metabolic and inflammatory markers: associations with individual depressive symptoms.* Psychological Medicine, 2017: p. 1-11.
98. Fried, E.I. and J.M. Haslbeck, *Using network analysis to examine links between individual depression symptoms, inflammatory markers, and covariates.* Psychological Medicine, 2019: p. 1-9.
99. Jokela, M., et al., *Inflammation and specific symptoms of depression.* JAMA psychiatry, 2016. **73**(1): p. 87-88.
100. White, J., et al., *Association of inflammation with specific symptoms of depression in a general population of older people: The English Longitudinal Study of Ageing.* Brain, behavior, and immunity, 2017. **61**: p. 27-30.
101. Hahn, T., A. Nierenberg, and S. Whitfield-Gabrieli, *Predictive analytics in mental health: applications, guidelines, challenges and perspectives.* Molecular psychiatry, 2016. **22**(1): p. 37.

102. Chekroud, A.M., et al., *Cross-trial prediction of treatment outcome in depression: a machine learning approach.* The Lancet Psychiatry, 2016. **3**(3): p. 243-250.
103. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics*. 2009, Springer New York.
104. Wang, Y., L. Kung, and T.A. Byrd, *Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.* Technological Forecasting and Social Change, 2018. **126**: p. 3-13.
105. Iniesta, R., D. Stahl, and P. McGuffin, *Machine learning, statistical learning and the future of biological research in psychiatry.* Psychological medicine, 2016. **46**(12): p. 2455-2465.
106. Feurer, M., et al., *Auto-sklearn: Efficient and Robust Automated Machine Learning*, in *Automated Machine Learning*. 2019, Springer. p. 113-134.
107. Christodoulou, E., et al., *A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models.* Journal of clinical epidemiology, 2019.
108. Mens, K.v., et al., *Predicting Future Suicidal Behaviour with Different Machine Learning Techniques: A Population-Based Longitudinal Study.* OSF, 2019. **April 4.**
109. Belsher, B.E., et al., *Prediction models for suicide attempts and deaths: a systematic review and simulation.* JAMA psychiatry, 2019. **76**(6): p. 642-651.
110. Kessler, R.C., et al., *Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration.* International journal of methods in psychiatric research, 2017. **26**(3): p. e1575.
111. Tran, B.X., et al., *The current research landscape on the artificial intelligence application in the management of depressive disorders: A bibliometric analysis.* International journal of environmental research and public health, 2019. **16**(12): p. 2150.
112. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* Int J Methods Psychiatr Res, 2008. **17**(3): p. 121-40.
113. de Beurs, et al., *Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice.* Clinical psychology & psychotherapy, 2011. **18**(1): p. 1-12.
114. Van Vliet, I. and E. De Beurs, *The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders.* Tijdschrift voor psychiatrie, 2007. **49**(6): p. 393.
115. Jayant, A. and a.O.R.M.C. Safari, *Data Science and Machine Learning Series: Naive Bayes Classifier Advanced Concepts*. 2020: Technics Publications.

[1] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands;
[2] Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, VU University Medical Center, and GGZ inGeest, Amsterdam, The Netherlands;

Chapter 2

Severity, course trajectory, and within-person variability of individual symptoms in patients with major depressive disorder

van Eeden, W. A.[1], van Hemert, A. M.[1], Carlier, I. V. E., Penninx, B. W.[2], & Giltay, E. J.[1]

(2019). Acta psychiatrica scandinavica, 139(2), 194-205.

Abstract

Background: Depression shows a large heterogeneity of symptoms between and within persons over time. However, most outcome studies have assessed depression as a single underlying latent construct, using the sum score on psychometric scales as an indicator for severity. This study assesses longitudinal symptom-specific trajectories and within-person variability of major depressive disorder over a 9-year period.

Methods: Data were derived from the Netherlands Study of Depression and Anxiety (NESDA). This study included 783 participants with a current major depressive disorder at baseline. The Inventory Depressive Symptomatology-Self-Report (IDS-SR) was used to analyze 28 depressive symptoms at up to six time points during the 9-year follow up.

Results: The highest baseline severity scores were found for the items regarding energy and mood states. The core symptoms depressed mood and anhedonia had the most favorable course, whereas sleeping problems and (psycho-)somatic symptoms were more persistent over 9-years follow-up. Within-person variability was highest for symptoms related to energy and lowest for suicidal ideation.

Conclusions: The severity, course, and within-person variability differed markedly between depressive symptoms. Our findings strengthen the idea that employing a symptom-focused approach in both clinical care and research is of value.

Significant findings:

1. Depressive symptoms have heterogenetic longitudinal characteristics.

2. Somatic/vegetative symptoms are less present at baseline but often exhibit a more persistent course trajectory.

3. Mood and cognitive symptoms are more severe at baseline but show favorable course trajectories.

Limitations:

1. The first part of the symptom trajectories were subject to a "regression to the mean" because patients were selected based on the criteria for MDD.

2. Outcomes were based on analysis with single items.

3. Because NESDA was an observational cohort study, other variables may have confounded our findings.

2.1 Introduction

Major depressive disorder (MDD) is a heterogeneous disease featuring large between-person differences in symptomatology and highly variable course trajectories [1, 2]. Most outcome research has focused on depression as a latent variable construct, representing a single underlying disorder, whereby the level of severity is measured as a sum score on self-report questionnaires of symptoms [1]. Given the possible unique combinations of the nine symptoms of which some are composite symptoms in the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5; e.g., anhedonia consists of two dimensions namely "loss of interest" and "inability to experience joy"), 227 different symptom combinations can be distinguished—all of which meet the requirements for a diagnosis of MDD (1). However, each individual symptom may have a separate severity, course trajectory, and variability over time, of which the potential importance is buried within the unified entity approach used in most outcome research [1]. Moreover, these sum-score-based methods do not maintain the hierarchical structure of the DSM-5 criteria of MDD, such as depressed mood or anhedonia, as a required core symptom.

Studies that did assess symptom-specific course trajectories have shown important differences between individual symptoms. Of the 12 studies that, to some extent, took symptom-specific courses within the adult population into account [2-13], sample sizes ranged from 51 [4] to 3,278 participants [13]. There were substantial differences in the methods and instruments that were used to assess individual symptoms. Studies used self-report measures [3, 4, 6, 10], clinician-rated measures [5, 11, 14], and structured interviews [2, 7, 9]. Therefore, comparing these studies should be done with caution. Most studies featured a prospective design with the duration of follow up ranging from 2 weeks [3] to 3 years [9]. Researchers often focused on identifying residual symptoms, and only three studies specifically reported on relatively fast remitting symptoms [2, 3, 12]. Four studies found that the two core symptoms, depressed mood and anhedonia, tended to persist as residual symptoms [2, 4, 5, 11], but sleep problems, energy loss, and cognitive problems were more often reported as residual symptoms [2, 5, 6, 12, 13, 15]. Fast remitting symptoms were negative self-view and psychomotor problems [3, 7, 12]. Some studies found no differences between individual symptoms [7, 8, 10].

The within-person variability of individual MDD symptoms over time has rarely been investigated. Patients with MDD tend to show a recurring and chronic disease course, with fluctuating levels of severity [2, 16]. Two studies found that a high variability of sum scores for severity was associated with an increased risk of relapse [17, 18], whereas another did not [7]. Some depressive symptoms tend to show large changes over time in a single patient, whereas other symptoms tend to remain stable or are in steady decline. Based on the mean range of the Hamilton Depression Scale item scores [19], energy loss, loss of libido, and sleep problems showed considerable levels of variability during the 3-year follow up of 114 patients with MDD [17]. On the other hand, suicidal thoughts and psychomotor retardation have demonstrated a more stable course [17].

The present study assessed the longitudinal symptom-specific characteristics of MDD in a large cohort over a 9-year period. To gain more insight into the heterogeneity of MDD, it is important to know which symptoms feature clinically favorable characteristics and which show a more persistent course. Despite the common use of aggregate sum scores in most research, we hypothesized that MDD is a disorder with substantial heterogeneity between symptoms in terms of severity, within-person slopes, and variability. A primary aim was to address some of the methodological gaps in earlier studies by assessing within-person variability over time in which repeated measures are nested within persons [20]. Therefore, we assessed baseline severity, course trajectory, and within-person variability of individual symptoms of depression over a 9-year period in a large sample of patients initially suffering from a current MDD.

2.2 Methods

2.2.1 Study sample and procedure

Participants were selected from the Netherlands Study of Depression and Anxiety (NESDA) cohort. A detailed description of the NESDA design and sampling procedures are published elsewhere [21]. The aim of the NESDA is to investigate the course and consequences of depressive and anxiety disorders. The first wave (baseline) started in 2004 and ended in September 2007. The sixth wave of measurement at the 9-year follow up finished in October 2016. The baseline measurement (*n* = 2,981) consisted of demographic and personal characteristics, standardized diagnostic psychiatric interviews, and medical assessments (e.g., BMI, blood sampling, etc.). The 1-year follow up consisted of a self-report questionnaire and was completed by 2,445 participants (82.0%). A face-to-face follow up assessment was conducted at 2 years (*n* = 2,596; 87.1%), 4 years (*n* = 2,256; 80.6%), 6 years (*n* = 2,256; 75.7%), and at 9 years postbaseline (*n* = 2,069; 69.4% of the baseline sample).

The cohort was recruited from the community (*n* = 564; 18.9%), general practice (*n* = 1,610; 54.0%), and secondary mental health care (*n* = 807; [27.1%; 21]. For the present analysis we only included patients with an 1-month diagnosis of MDD —the excluded participants did not have a mood disorder at the time of baseline assessment (67.3%), had dysthymia without MDD (2.1%) or a minor depression (2.9%). This resulted in a final study sample of 783 participants.

2.2.2 Measures

We used the Composite International Diagnostic Interview (CIDI; WHO version 2.1) to assess the presence of depressive disorders according to the DSM-IV. The CIDI is a fully standardized diagnostic interview with extensively validated psychometric characteristics [21, 22].

Chronic depression and chronic somatic disease at baseline were measured for the purpose of post hoc sensitivity analyzes. Depression history was assessed using the Life Chart Interview method—a standardized interview designed to retrospectively assess the course of psychopathology (22). The Life Chart Interview uses age- and calendar-linked life events that occurred over the course of a patient's past 4 years and then assesses the presence and the severity of symptoms during this period. Participants who were depressed for 24 months or

more during this period of 48 months (i.e., > 50% of the time) were defined as being chronically depressed (22).

Patients were asked if they exhibited the following chronic somatic diseases: asthma, chronic bronchitis or pulmonary emphysema, heart disease, diabetes, stroke or CVA, osteoarthritis, cancer, stomach or intestinal ulcers, intestinal disorders, liver disease, epilepsy, or thyroid gland disease. Patients were also asked if they had other chronic somatic diseases that caused substantial disability, was being treated by a clinician or was treated with medication.

The individual items of the IDS-SR [23] were used as the outcome measures. The scale concerns all symptoms of depression, including melancholic, atypical, and anxious symptoms. Moreover, several additional symptoms have been added, such as sympathetic arousal, pessimism, and interest in sex. The IDS-SR consists of 30 equally weighted items rated on a 4-point ordinal scale ranging from 0 to 3. On the IDS-SR, a sum score of 14–25 is considered mild depression, 26–38 severe, and 39–49 very severe depression [23, 24].

The psychometric characteristics of the IDS-SR have been assessed in samples which included MDD outpatients, chronic MDD outpatients, and euthymic subjects [23, 25]. The IDS-SR demonstrated adequate internal consistency, with Cronbach's alphas ranging from 0.92–0.94. The IDS-SR sum score significantly discriminated between symptomatic and nonsymptomatic patients ($p < 0.0001$) and was highly related to the 17-item Hamilton Rating Scale for Depression [correlation: 0.88; 26]and Beck's Depression Inventory [correlation: 0.93; 27]. Analysis of sensitivity to change in symptom severity showed that the IDS-SR sum score dropped at about the same rate as the Hamilton Rating Scale for Depression [23]. At item level, effect sizes of change were larger for the IDS-SR as compared to the Hamilton Rating Scale for Depression [25].

In our study sample, the Cronbach's alphas were 0.83, 0.89, 0.89, 0.90, 0.90, and 0.90 for the six time points, respectively, from baseline to 9 years. Because Items 11 and 12 (increased/decreased appetite) and Items 13 and 14 (weight gain/weight loss) contained opposite features, these item pairs were combined into one ordinal item in order to maintain psychometric similarity between the items, which yielded 28 items for the present analyzes [23]. In order to enhance interpretability, we grouped the symptoms by symptom clusters, which were previously identified across various studies in the figures (28). The symptom

clusters had no role in computing our outcome variables—only in how they were grouped in the figures. The clusters include 10 mood symptoms (capacity for pleasure, general interest, quality of mood, reactivity of mood, feeling anxious or tense, feeling irritable, feeling sad, interpersonal sensitivity, leaden paralysis, panic/phobic symptoms), 14 somatic/vegetative symptoms (aches and pains, constipation/diarrhea, mood in time of the day, waking up early, low energy, sympathetic arousal, problems falling asleep, sleep during the night, psychomotor agitation, psychomotor retardation, interest in sex, sleeping too much, weight gain/loss, increased/decreased appetite), and four cognitive problems (concentration/decision-making, view of my future, view of myself, suicidal thoughts; [28])

2.2.3 Statistical analysis

Multiple steps were taken to assess the longitudinal MDD symptom characteristics. The outcome measures (baseline item score, slope, and fraction of variance unexplained) were summarized and presented with a 95% CI (represented by error bars) in forest plots, which were sorted by the size of each mean effect estimate. All analyzes were computed using R, version 3.4.1, with main packages mixor (29), mirt (30), tidyverse (31), ggplot2 (32), and ggrepel (33).

Baseline item scores

The changes for each of the IDS-SR item scores over time were examined by calculating the mean at each time point (baseline, year 1, year 2, year 4, year 6, and year 9) and by visualizing trajectories of the means in a line graph. The baseline mean score for each IDS-SR item represents baseline severity for each symptom. In order to test the psychometrics and whether or not the IDS-SR items measured a unidimensional latent construct, we conducted polytomous item response theory analyzes (IRT) on all IDS items at baseline in 783 MDD patients. This was done once for all 28 items of the IDS-SR, and once for a selection of 6 items that suggested to represent a unidimensional melancholia construct in earlier studies, i.e. item 5 “feeling sad”, item 7 “anxious or tense”, item 16 “view of myself”, item 19 “general interest, item 20 “energy level”, and item 23 “psychomotor retardation” (34, 35).

Slopes

We analyzed the course trajectories for each of the 28 items using a cumulative link ordinal response mixed effects model (29). This model takes the ordinal outcome and longitudinal nature of our data into account; models are fitted by using an adaptive quadrature and an ordered probit link (29). Equal intervals between the ordinal scores (0–1, 1–2, 2–3) were not assumed (36). The model returns estimated parameters like the slope and intercept (29).

Because most recovery occurred within the first year, the slopes were calculated separately over this period. To analyze which symptoms remitted relatively faster, or were relatively more persistent than others over the course of 9 years, the 9-year slope was estimated, while adjusting for the sum score at each time point. This yielded the symptom trajectory relative to the overall decrease of the sum score. Thus, a negative value indicates that that item has a larger decrease than the overall decrease of the sum score, and a positive value indicates the opposite.

To compare each of the mean slopes, baseline severity must be taken into account. A baseline item-score of 0, has only room for change towards the higher scores. On the opposite, a ordinal score of 3 is the highest level measured in the IDS-SR and no values above that point are possible. Baseline severity was taken into account by letting the random intercept and random slope correlate with each other when computing the ordinal mixed model.

Fraction of variance unexplained

We calculated the fraction of variance unexplained (FVU) per item as a measure of within-person variability. A high FVU represents a variable course with more fluctuation throughout the follow up years. A low FVU represents a stable course, that is, symptoms with a steady decline or a stable persistent course or symptoms that, if not present, are not likely to be present in the future. FVU was calculated using a simple linear regression analysis. We computed the regression analyzes per person and per item, resulting in a total of 15,624 modeled regression lines (i.e., n. of participants * no. of items). As the steep slope within the first year at follow up had disproportional large impact on the FVU measure and we were interested in the FVU as a function of within-person variability over time, and not as a function of recovery, we decided to exclude the baseline measurements when calculating the FVU. When patients did not fulfill all five follow up IDS assessments, regression analyzes were

computed based on the remaining time points (at least three). This approach of modeling course variability per individual has been used in other fields of medical research, for example, blood pressure variability (37).

Sensitivity analyzes

To test the robustness of the baseline mean item score and FVU, several sensitivity analyzes were done in subsamples that excluded chronically depressed patients (at baseline), patients with chronic somatic diseases, and antidepressant users. In addition, we tested the robustness of the 1-year and 9-year slopes in ordinal response mixed effects models, for which we additionally adjusted for four variables: a history of chronic depression at baseline, chronic somatic diseases, age, and the use of antidepressants.

2.3. Results

2.3.1 Demographics

Characteristics of the study population are presented in Table 1. Age at baseline ranged from 18–64 (*M* = 41.75, *SD* = 12.0) years, and 362 (66.3%) participants were women. The mean sum IDS-SR score of the study sample was 35.6 (*SD* = 11.3), indicating severe depression at baseline. A large portion of the sample had one or more chronic somatic diseases (see Table 1).

Table 1. Sociodemographic and clinical characteristics

	Cohort (*N* = 783)
Age in years (mean, *SD*)	41.75 (12.0)
Female (%)	66.28
North-European ethnicity (%)	92.72
Years of education (mean, *SD*)	11.0 (3.1)
Chronic depressed[1] (%)	39.1
One or more chronic somatic deseases[2] (%)	62.1
Treatment setting[3]	
Primary care (%)	45.9
Secondary care (%)	54.1
Antidepressants	
TCA (%)	3.7
SSRI (%)	30.3
Other (%)	10.5
no AD (%)	56.6
total baseline score IDS-SR (mean, *SD*)	35.56 (11.24)

Note. TCA = tricyclic antidepressants. SSRI = selective serotonin reuptake inhibitors. AD = antidepressants. IDS-SR = The Inventory Depressive Symptomatology-Self-Report

[1] Depressed for 24 months or more at baseline

[2] The following diseases were asked: asthma, chronic bronchitis or pulmonary emphysema, heart disease, diabetes, stroke or CVA, osteoarthritis, cancer, stomach or intestinal ulcers, intestinal disorders, liver disease, epilepsy, or thyroid gland disease, and other chronic disease

[3] Refers to mental health care

2.3.2 Mean values over time

After 2 years, 30% of our original study population fulfilled the DSM-IV criteria for MDD, implying that a large part of the sample met the criteria for (partial) remission of MDD. The

number of patients fulfilling criteria for MDD was cross-sectionally assessed at each later wave of follow up. The percentage of patients fulfilling criteria of MDD further declined to 25.6% at the 4-year follow up, 22.1% at the 6-year follow up, and 17.1% at the 9-year follow up. The unadjusted means of the individual symptoms at all six time points are presented in Table 2 and Figure 1. Despite a large variation in the mean scores at baseline and the magnitude of decrease over the years of follow-up, a similar pattern was found for all symptoms, that is, for each of the items, the largest decline in the mean scores occurred between baseline and the 1-year follow up, and the decline was much less in later years. Three items remained remarkably high after the 9-year follow up: Item 30 "leaden paralysis," Item 2 "sleep during the night," and Item 25 "aches and pains."

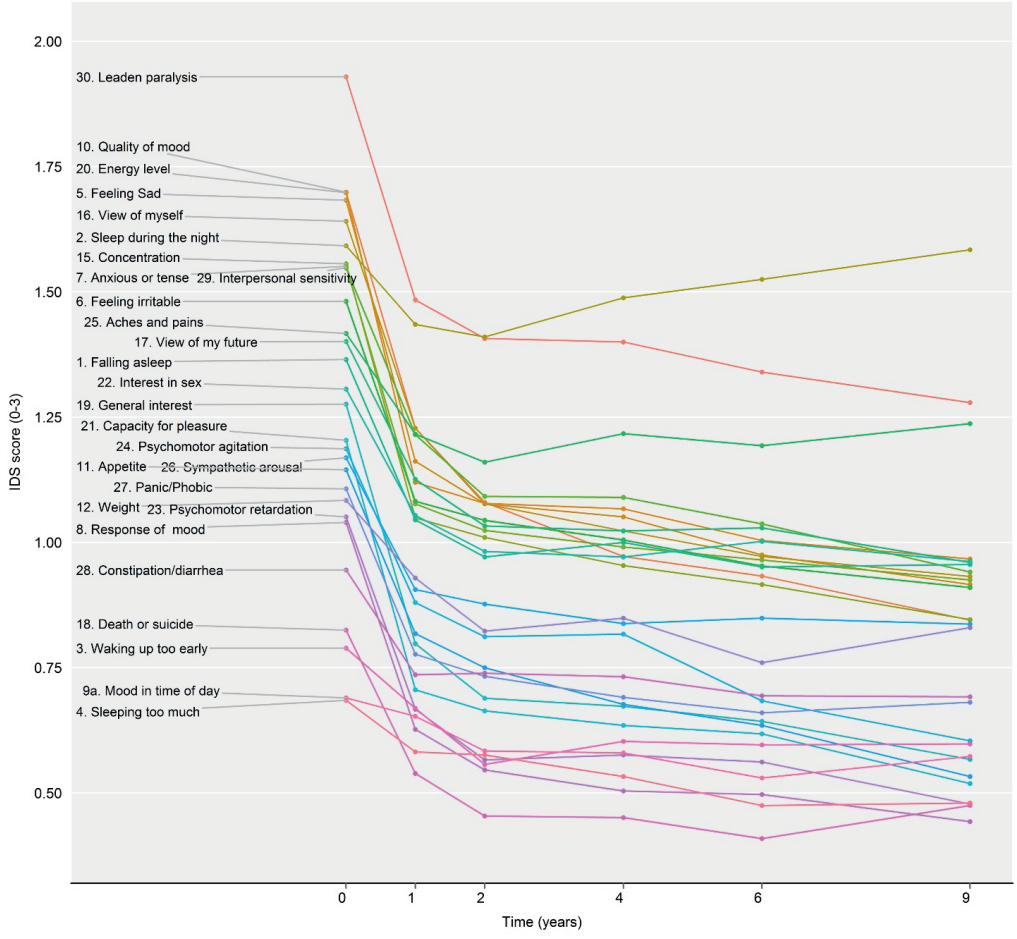


Fig. 1. Group-level mean item scores over the course of 9 years.

2.3.3. Baseline severity

The baseline mean (with standard error) is presented in Table 2 and Figure 1. Baseline mean with the 95% CI and baseline mean in relation to the 1-year slope are presented in the supplementary materials (see Figure 1 and Figure 2). The baseline mean of all items combined was 1.29 and ranged from 0.69 (Item 4 “sleeping too much”) to 1.93 (Item 30 “leaden paralysis”). The highest baseline severity was found for items concerning energy and depressed mood (Items 30, 10, 20, 5), followed by “low self-esteem” (Item 16), “sleep during the night” (Item 2), “concentration” (Item 15), “feeling anxious or tense” (Item 7), and “sensitivity” (Item 29). Interestingly, the mean of Items 20 “energy level” and 2 “sleep during the night” showed a much higher baseline mean level compared to most other symptoms within the somatic/vegetative cluster. Other items within the somatic/vegetative domain were less severe at baseline. The lowest mean baseline values were found for Item 4 “sleeping too much,” Item 9a “mood in time of the day,” Item 3 “waking up too early,” and “thoughts of death or suicide.”.

The results regarding the IRT analysis suggested that the IDS-SR was not unidimensional, i.e. items did not measure a single latent construct as principle component loadings varied widely. The component loadings ranged from 0.059 (item 9a “mood in time of day”) to 0.689 (item 21 “capacity for pleasure”). Of the 28 items in the IDS-SR, 18 items had a component loading below < 0.400. The discrimination values were rather weak (a’s), for example item 1 “falling asleep” (a = 0.164) and item 9a “mood in time of day” (a = 0.101). Only 5 items had discrimination values higher than 1, notably item 5 “feeling sad” (a = 1.350) and item 21 “capacity for pleasure” (a = 1.616).

When assessing the six items that in previous studies were found to represent a unidimensional melancholia construct (34, 35), component loadings ranged from 0.244 (item 16 “view of myself”) to 0.605 (item 5 “feeling sad”). Of the six items, 2 items had rather weak component loadings below 0.400, i.e. item 16 “view of myself” (loading = 0.244), and item 23 “psychomotor retardation” (loading = 0.338). Only item 5 “feeling sad” (a = 1.293) had a discrimination parameter above 1 and three items had partial credit model parameters that were not ordered in accordance with the item scales, i.e. item 16 “View of myself”, item 19 “General interest”, and item 23 “Psychomotor retardation”. More detailed results regarding component loadings, the discriminative properties and item specific partial credit model

parameter estimates thresholds can be found in table 1 of the supplementary material and figure 3 of the supplementary material. In sum, our findings from the IRT analyzes are not in support of the idea of a single coherent latent construct of depression.

2.3.4 Slope during the first year

The symptom-specific slope during the first year is presented in Table 2 and Figure 2A. The overall mean slope of all items combined was −0.566, ranging from −0.061 (Item 9a "mood in time of the day") to −0.993 (Item 20 "energy level"). Many slopes of items within the somatic/vegetative symptom cluster were close to 0 (horizontal slopes). Exceptions were Item 20 ("energy level"), Item 11 ("change in appetite"), and items assessing psychomotor retardation and agitation (Item 23, Item 24). The symptoms with the smallest decrease (mean slopes close to 0) were found for items concerning quality of sleep, diurnal variation in mood (Item 9a "mood in time of the day"), and somatic complaints (e.g., sympathetic arousal, headache, and back pain). Larger slopes (steeper declines) were found for the mood symptoms (e.g., both core symptoms; depressed mood and anhedonia), concentration, anxious and anger symptoms ("anxious or tense," "feeling irritable"), and energy (i.e., energy level).

The symptom course (slope) in relation to baseline severity (mean item score) is shown in Figure 2 of the supplementary material. Items with a high baseline mean tended to show a stronger decrease over time. Two items with steep slopes fell within the 95% CI: Item 30 ("leaden paralysis") and Item 10 ("quality of mood"). Two items with slopes close to zero also had a small mean baseline item score (within the 95% CI): Item 4 ("sleeping too much") and Item 28 ("constipation/diarrhea). The regression line with a 95% CI provides insight into the association between baseline severity and slope and symptoms that do not fulfill this association.

2.3.5 Slope adjusted for IDS sum scores

The adjusted slopes over 9 years are presented in Table 2 and Figure 2B. Twelve symptoms had a slope that was significantly different from 0, indicating a larger or smaller decrease than the overall sum score. Of the six items with a relatively larger decline, three were in the mood symptom cluster: Item 5 ("feeling sad"), Item 8 ("response of mood"), and Item 10 ("quality of mood"); one was a cognitive symptom (Item 15 "concentration"); and two items were in the somatic/vegetative symptom cluster: Item 11 ("appetite") and Item 24 ("psychomotor

agitation"). Four symptoms with a smaller decrease than the overall sum score were in the somatic/vegetative symptom cluster: Item 2 ("sleep during the night"), Item 3 ("waking up too early"), Item 25 ("aches and pains"), Item 22 ("interest in sex"), and Item 26 ("sympathetic arousal"). One item with a small decrease fell within the cognitive symptom cluster: Item 17 ("view of my future").

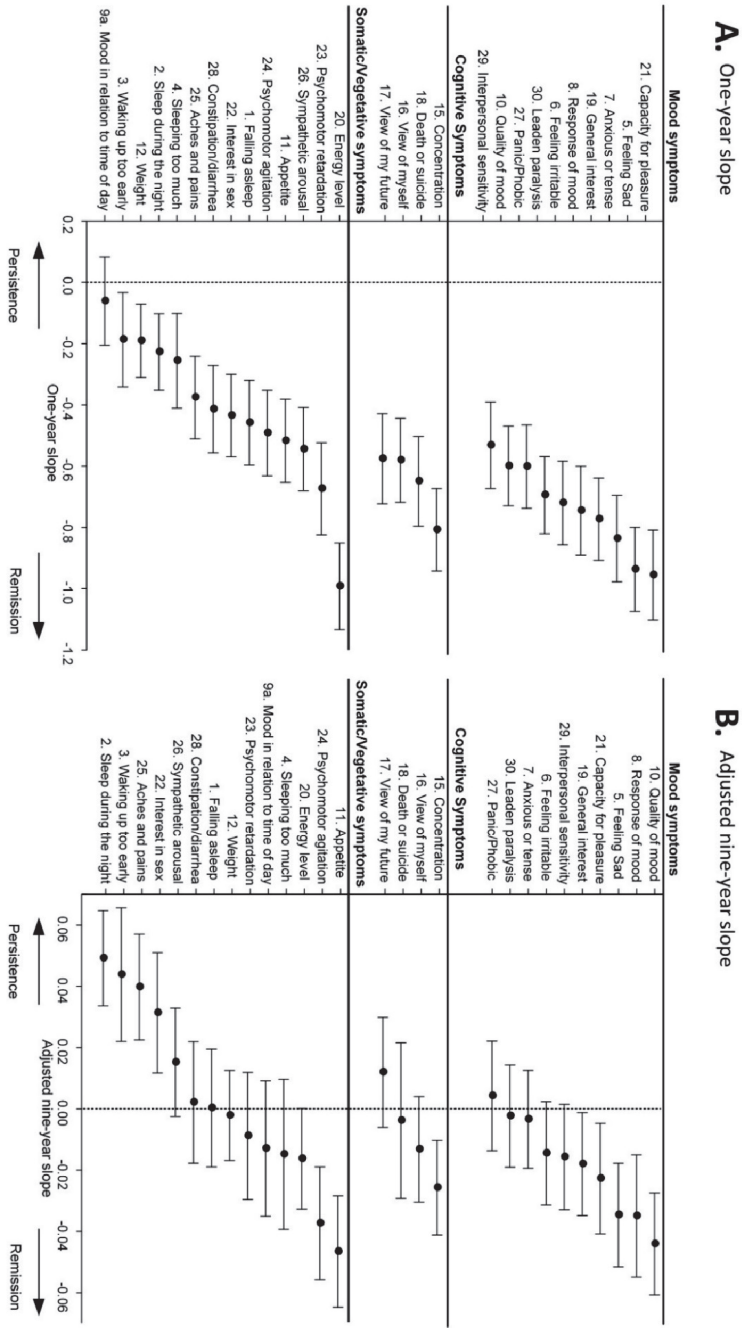


Fig. 2. Unadjusted 1-year slope represents the decrease in symptom severity after the first year of follow-up. Negative values represent a steeper decline. Sum score adjusted 9-year slope represents the decline in symptom severity in relation to the sum score. Negative values mean that the symptom had a steeper decline compared to the overall sum score.

2.3.6 Variability

Patients with four or more IDS-SR assessments were included for the FVU analysis, which resulted in a sample size of n = 498. Excluded patients (with less than four assessments; n = 244) were less likely to be of northern European heritage (86.9% vs. 95.3%; $p < 0.006$) and had lower mean number years of education (M = 10.4 vs. M = 11.7 years; $p < 0.004$). We found similar characteristics between included and excluded patients for the remaining variables mentioned in Table 1, such as gender (64.7% female), antidepressants (3.2% TCA; 29.9% SSRI; 10.9% other antidepressants; 56.8% no antidepressants), chronic depression (36%), chronic somatic disease (61.2%), and IDS-SR sum score (34.7; SD = 11.3).

The within-person FVU for each symptom is presented in Table 2 and Figure 3. The overall FVU of all items combined was 0.498, ranging from 0.339 (Item 18 “death or suicide”) to 0.591 (Item 30 “leaden paralysis”). Among the items with high within-person variability, all three symptom clusters were equally represented. Item 30 (“leaden paralysis”) was the most unstable followed by Item 11 (“weight”), Item 7 (“anxious or tense), Item 25 (“aches and pains), Item 29 (“interpersonal sensitivity”), and Item 20 (“low energy level”). The most stable items fell within the somatic/vegetative symptom cluster, with the exception of Item 18 (“thinking of death or suicide”). Other particularly stable items were Item 3 (“waking up too early”), Item 23 (“psychomotor retardation”), and Item 4 (“sleeping too much”). Note that many of the stable symptoms had low baseline severity. This means that when symptoms were not present at baseline, they were often unlikely to be present at the follow up, except for Item 1 “falling asleep.”

2.3.7 Sensitivity analysis

We conducted several post hoc sensitivity analyzes in which we assessed the effects on baseline severity, slope, and FVU. These results are presented in Table 2 of the supplementary material. We assessed baseline severity (i.e., the mean baseline item score) in the subgroup of patients with a history of chronic depression, chronic somatic disease, and antidepressant users. Overall, the mean baseline severity was slightly lower when we excluded chronic depressed patients (delta = −0.09; M = 1.20), patients with chronic somatic diseases (delta = −0.05; M = 1.25), and antidepressant users (delta = −0.08; M = 1.21). When taking individual items into account, no meaningful differences were found because only two items had a delta

(i.e., unadjusted mean minus the adjusted mean) larger than −0.20: Item 28 “constipation/diarrhea” adjusted for chronic somatic diseases (−0.23) and Item 25 “aches and pains” adjusted for chronic somatic diseases (−0.25). When symptoms were sorted according to the level of severity, the overall order remained almost similar.

For the next sensitivity analyzes, 1-year slope and 9-year slope findings were tested for robustness. Therefore, models were adjusted for a history of chronic depression, chronic somatic diseases, the use of antidepressants, and age. This again resulted in similar findings. Sorting on effect sizes did not change the order (see Table 2 in the supplementary material).

Finally, sensitivity analyzes for the FVU hardly affected our findings. Only three items showed slight changes when patients with a history of chronic depression were excluded (Item 15 “concentration,” Item 17 “view of my future,” and Item 25 “aches and pains”) and three items when antidepressant users were excluded (Item 2 “sleep during the night,” Item 15 “concentration,” and Item 17 “view of my future”; see Table 2 in the supplementary material).

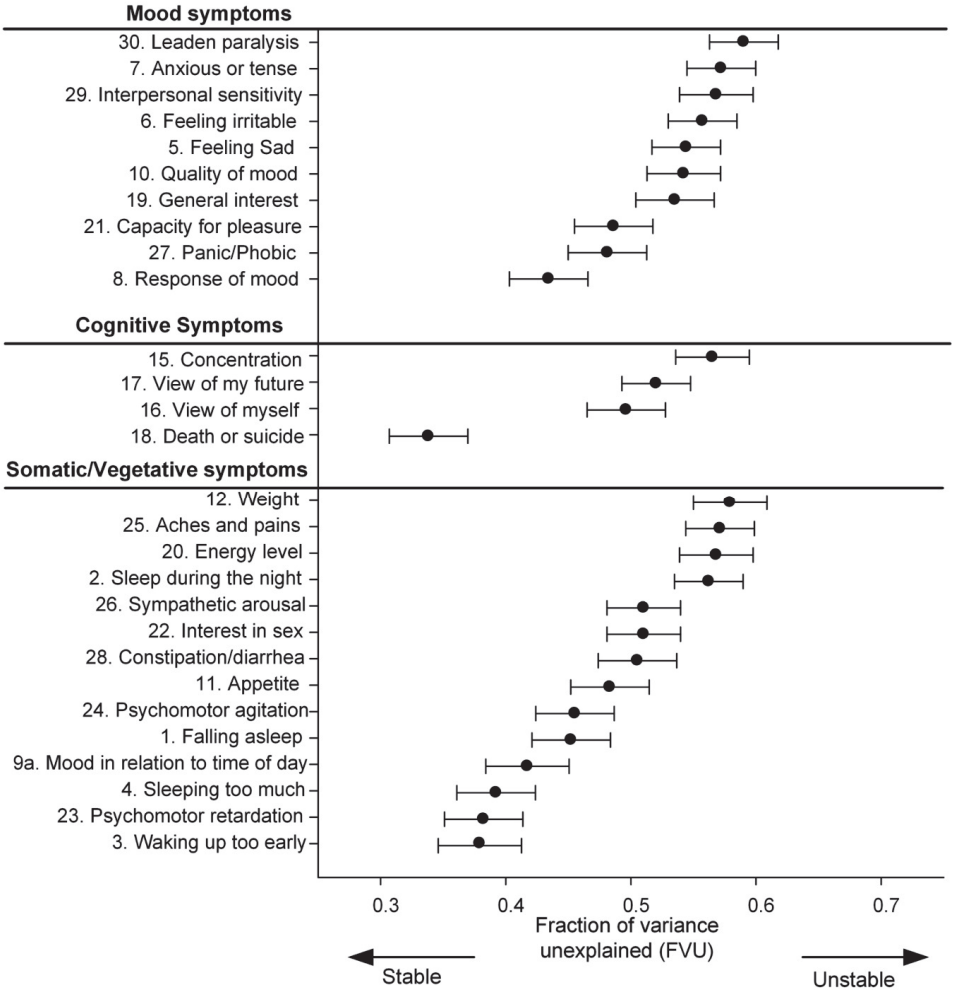


Fig. 3. Within-person variability based on 8 years follow-up (baseline excluded)

Table 2. IDS symptoms during 9 years follow up

Item	Baseline	Year 1	Year 2	Year 4	Year 6	Year 9	1-year slope	Adjusted 9-year slope	FVU
1. Falling asleep	1.37 (0.04)	1.05 (0.05)	0.97 (0.04)	1.00 (0.05)	0.95 (0.05)	0.96 (0.05)	-0.458 (0.071)	<0.001 (0.010)	0.453 (0.016)
2. Sleep during the night	1.59 (0.04)	1.44 (0.04)	1.41 (0.04)	1.49 (0.04)	1.53 (0.05)	1.58 (0.05)	-0.227 (0.064)	0.049 (0.008)	0.563 (0.014)
3. Waking up too early	0.79 (0.04)	0.67 (0.04)	0.56 (0.04)	0.60 (0.04)	0.60 (0.04)	0.60 (0.05)	-0.187 (0.078)	0.043 (0.011)	0.380 (0.017)
4. Sleeping too much	0.69 (0.03)	0.58 (0.03)	0.58 (0.03)	0.53 (0.03)	0.48 (0.03)	0.48 (0.03)	-0.256 (0.079)	-0.015 (0.013)	0.393 (0.016)
5. Feeling sad	1.68 (0.03)	1.16 (0.03)	1.08 (0.03)	1.05 (0.04)	0.98 (0.04)	0.92 (0.04)	-0.937 (0.070)	-0.035 (0.009)	0.545 (0.014)
6. Feeling irritable	1.48 (0.03)	1.08 (0.03)	1.04 (0.03)	1.01 (0.03)	0.95 (0.04)	0.91 (0.04)	-0.720 (0.070)	-0.015 (0.009)	0.558 (0.014)
7. Anxious or tense	1.55 (0.03)	1.08 (0.04)	1.02 (0.03)	0.99 (0.03)	0.97 (0.04)	0.93 (0.04)	-0.837 (0.072)	-0.004 (0.008)	0.573 (0.014)
8. Response of mood	1.04 (0.03)	0.63 (0.03)	0.55 (0.03)	0.50 (0.03)	0.50 (0.04)	0.44 (0.04)	-0.746 (0.074)	-0.035 (0.010)	0.435 (0.016)
9a. Mood in time of day	0.69 (0.04)	0.65 (0.04)	0.58 (0.04)	0.58 (0.04)	0.53 (0.04)	0.57 (0.05)	-0.061 (0.073)	-0.013 (0.011)	0.418 (0.017)
10. Quality of mood	1.70 (0.03)	1.23 (0.05)	1.08 (0.04)	0.97 (0.05)	0.93 (0.05)	0.85 (0.05)	-0.600 (0.066)	-0.045 (0.009)	0.543 (0.015)
11. Appetite	1.15 (0.04)	0.82 (0.04)	0.75 (0.04)	0.68 (0.04)	0.64 (0.04)	0.53 (0.04)	-0.518 (0.069)	-0.047 (0.009)	0.484 (0.016)
12. Weight	1.08 (0.04)	0.93 (0.04)	0.82 (0.04)	0.85 (0.04)	0.76 (0.04)	0.83 (0.05)	-0.191 (0.061)	-0.003 (0.008)	0.580 (0.015)
15. Concentration	1.56 (0.03)	1.05 (0.04)	1.01 (0.03)	0.95 (0.04)	0.92 (0.04)	0.85 (0.04)	-0.808 (0.069)	-0.026 (0.008)	0.566 (0.015)
16. View of myself	1.64 (0.04)	1.23 (0.05)	1.08 (0.05)	1.02 (0.05)	0.97 (0.05)	0.93 (0.06)	-0.581 (0.070)	-0.014 (0.009)	0.497 (0.016)
17. View of my future	1.40 (0.03)	1.13 (0.03)	1.03 (0.03)	1.02 (0.04)	1.03 (0.04)	0.96 (0.04)	-0.576 (0.075)	0.012 (0.009)	0.521 (0.014)
18. Death or suicide	0.83 (0.03)	0.54 (0.03)	0.45 (0.03)	0.45 (0.03)	0.41 (0.03)	0.48 (0.03)	-0.649 (0.075)	-0.004 (0.013)	0.339 (0.016)
19. General interest	1.28 (0.03)	0.80 (0.04)	0.69 (0.03)	0.67 (0.04)	0.64 (0.04)	0.57 (0.04)	-0.773 (0.069)	-0.018 (0.009)	0.536 (0.016)
20. Energy level	1.70 (0.03)	1.12 (0.04)	1.08 (0.04)	1.07 (0.04)	1.00 (0.04)	0.97 (0.04)	-0.993 (0.072)	-0.017 (0.008)	0.569 (0.015)
21. Capacity for pleasure	1.20 (0.03)	0.71 (0.03)	0.66 (0.03)	0.64 (0.03)	0.62 (0.03)	0.52 (0.03)	-0.956 (0.075)	-0.023 (0.009)	0.487 (0.016)
22. Interest in sex	1.31 (0.04)	1.05 (0.04)	0.98 (0.04)	0.97 (0.04)	1.00 (0.05)	0.96 (0.05)	-0.435 (0.069)	0.031 (0.010)	0.511 (0.015)
23. Psychomotor retardation	1.05 (0.03)	0.67 (0.04)	0.57 (0.04)	0.58 (0.04)	0.56 (0.04)	0.48 (0.04)	-0.674 (0.077)	-0.009 (0.011)	0.383 (0.016)
24. Psychomotor agitation	1.19 (0.03)	0.88 (0.04)	0.81 (0.04)	0.82 (0.04)	0.68 (0.04)	0.60 (0.04)	-0.492 (0.071)	-0.038 (0.009)	0.456 (0.016)
25. Aches and pains	1.42 (0.03)	1.22 (0.03)	1.16 (0.03)	1.22 (0.04)	1.19 (0.04)	1.24 (0.04)	-0.376 (0.069)	0.039 (0.009)	0.572 (0.014)
26. Sympathetic arousal	1.17 (0.03)	0.91 (0.03)	0.88 (0.03)	0.84 (0.03)	0.85 (0.03)	0.84 (0.04)	-0.545 (0.069)	0.015 (0.009)	0.511 (0.015)
27. Panic/phobic	1.11 (0.04)	0.78 (0.04)	0.73 (0.03)	0.69 (0.03)	0.66 (0.04)	0.68 (0.04)	-0.601 (0.069)	0.004 (0.009)	0.482 (0.016)
28. Constipation/diarrhea	0.95 (0.03)	0.74 (0.04)	0.74 (0.03)	0.73 (0.04)	0.69 (0.04)	0.69 (0.04)	-0.414 (0.073)	0.002 (0.010)	0.506 (0.016)
29. Interpersonal sensitivity	1.55 (0.04)	1.22 (0.04)	1.09 (0.04)	1.09 (0.04)	1.04 (0.04)	0.94 (0.04)	-0.532 (0.072)	-0.016 (0.009)	0.569 (0.015)
30. Leaden paralysis	1.93 (0.03)	1.48 (0.04)	1.41 (0.04)	1.40 (0.04)	1.34 (0.04)	1.28 (0.04)	-0.695 (0.064)	-0.003 (0.009)	0.591 (0.014)

Note. Mean values, standard error (in parentheses), 1-year slope, sum score adjusted 9-year slope, and fraction of variance unexplained (FVU; $\Sigma(y_i - \hat{y}_i)^2/\Sigma(y_i - \bar{y})^2$) for each of the individual symptoms of the Inventory of Depressive Symptomatology- Self-Report (IDS-SR).

2.4. Discussion

Our study confirms the existence of substantial heterogeneity between depressive symptoms in terms of symptom severity at baseline, slopes over time, and within-person variability over time. Furthermore, results of the IRT analysis suggested that the individual symptoms measured with the IDS-SR do not unidimensionally assess one latent construct, for example high scores on “feeling sad” and “capacity for pleasure”, may be much more meaningful for the severity of depression than high scores on “falling asleep” and “mood in time of day”. Mood symptoms (e.g., core symptoms depressed mood and anhedonia) were (on average) more severe at baseline and showed a relatively favorable course. Somatic/vegetative symptoms (e.g., sleep and somatic complaints) showed (on average) less severity at baseline and their characteristics often followed a more persistent course. These results persisted after adjusting for a history of chronic depression, chronic somatic diseases, age, and the use of antidepressants. Additionally, energy symptoms showed a higher variability within patients than did suicidal thoughts. This diversity in longitudinal symptom characteristics raises the question as to whether using a sum score of 28 items addresses the heterogeneity between symptoms.

For all items in our study, the largest (mean) recovery took place within the first study year. When the diagnostic criteria for MDD were assessed 2 years postbaseline, 70% of the patients had recovered from MDD. However, other studies report that, although 50–90% recovered within the first year, many patients still experienced residual symptoms or relapsed after initial remission [16, 28].

Research on the symptom-specific characteristics during and directly following a depressive episode is scarce. In our group of MDD patients, a depressed mood and low energy level were among the most severe symptoms at baseline, which is in line with most other reports [7, 12]. In our population, in contrast to others [2, 5, 11], the mood symptoms (e.g., depressed mood and anhedonia) showed a more favorable course. Somatic/vegetative symptoms, such as sleep and somatic complaints, often had more persistent course trajectories. The persistent course of insomnia is in line with most other studies [5, 29-31], with two exceptions [7, 13]. The generally low severity at baseline, but persistent nature of multiple somatic symptoms associated with depression, has been documented in earlier studies [32, 33]. These studies

suggested that patients who experience these symptoms may represent a separate subgroup of MDD [32, 33]

We found significant differences between symptoms regarding within-person variability. Suicidal ideation tended to be stable and showed less fluctuation within patients over time. If patients had suicidal ideations, they were likely to keep on having these ideations during the subsequent years of follow up. If patients did not have suicidal ideations during their depressive episode at baseline, they were unlikely to experience them in the future. Suicidality is described in the literature as being related to a specific cognitive response pattern of hopelessness; this pattern is continually present throughout an individual's life [34]. From a psychometric perspective, we could argue that the latent thresholds for scoring 0, 1, 2, or 3 on the item "energy level" are much lower than those on the item "suicidal ideation" [35]. A 1-point change in an unstable item, such as "energy level," is clinically of less importance than a 1-point change in a stable item, such as "suicidal ideation." Our results on variability are in line with those of Karp et al. (16) who found energy loss to be an unstable symptom and suicidal thoughts to be a stable symptom among 114 patients with MDD (aged 21–65 years) during a follow up lasting 3 years. More research is needed on the topic of within-person variability. Beside group-level changes of individual items, the within-person variability may have additional predictive and/or clinical value.

Drawing inferences about changes in depression severity is an imperfect process because severity cannot be measured directly [20]. Outcome measurements are generally based on a questionnaire sum score in which the same weight is given to each item. This method would be valid if MDD was a unified construct and all its symptoms contributed equally to its latent construct [1, 16]. However, MDD is unlikely to be a distinct illness that causes all of its symptoms [1, 9, 36]. Instead, MDD is more like a complex system in which symptoms are connected by a dynamic network of causality [37-39]. The symptom-specific diversity in mean item scores, slopes, and variability shows that symptoms are not diagnostically equivalent and are not interchangeable [40]. The persistent use of merely a sum score to estimate depression severity may obscure insight into both patient and symptom-specific characteristics and can lead to misinterpretations regarding depressive severity over time [1, 41]. For example, a patient who recovers by feeling less depressed will show a similar change in the depressive severity measure as a patient whose recovery takes place in another

symptom domain, such as sleep. Even when there is a significant change in the sum score, a clinically important change might be obscured by more trivial changes on other items. It is therefore advised to assess individual symptoms in addition to sum scores when testing a patient's (longitudinal) depressive characteristics.

Research on personalized medicine in mental health care [15, 42, 43] and treatment of specific (residual) symptoms has highlighted that a symptom-specific approach may be beneficial [44, 45]. In general, depression treatment focuses mainly on the core symptoms of depression. However, other symptoms (e.g., sleeping problems) are more persistent and can indicate a risk factor for relapse; therefore, these symptoms deserve particular attention as a focus for treatment [29, 30]. Moreover, because a causal relationship exists between symptoms on group level (47-49), targeting the key symptoms (i.e., more central in the causal network of depressive symptoms) in clinical care may benefit a patient's recovery. Symptom-specific treatment of, for example, sleeping problems are widely available [31, 46]. For instance, cognitive behavioral therapy and pharmacological treatment for insomnia appear to have a positive effect on depression [46, 47]. It seems that our currently applied treatments warrant a more symptom-specific approach in order to also take the persistent (somatic/vegetative) symptoms into account.

The present study has several strengths. A large sample of MDD patients was included and followed for up to 9 years, whereas many earlier studies featured shorter follow-up periods or cross-sectional designs. Using a per-person, per-item method allowed us to compute a measure for within-person variability. Although the use of this method is relatively rare in the field of psychiatric research, it is often used in other fields of medicine [48].

The study also has some limitations. First, because all patients were initially selected to fulfill criteria for MDD, the first part of the symptom trajectories were subject to a "regression to the mean" effect [49]. Therefore, baseline severity needed to be taken into account when interpreting the slope measures. Furthermore, because the steep decline within the first year had a large effect on the variance within patients, we calculated the FVU and excluded the baseline measure. Second, the FVU measure may be affected by the design of the IDS-SR with severity measured on a nominal scale. When participants scored a baseline severity of 0 on a particular item, there would only be room for change towards the higher scores. On the other hand, a baseline ordinal score of 3 is the highest score and scores above that point cannot be

measured, this again limits the ability of the instrument to detect variability. Third, assessing individual symptoms based on single items presents psychometric difficulties. Single items are more strongly affected by random error than sum scores of items, which may have particularly affected our FVU measures. Moreover, we did not assess the reliable change indices because the focus of our study was not on the clinical impact of a one-point ordinal scale change of each item. Finally, because the NESDA was an observational cohort study, several variables may have confounded our findings. We performed multiple sensitivity analyzes to test other variables, such as pharmacological treatment (e.g., antidepressants). We found that our results remained robust and that only minimal changes occurred after adjusting for other variables.

In this study, we examined within-person trajectories over time of different depressive symptoms measured using the IDS-SR. The severity, course, and variability differed markedly between the depressive symptoms and between patients, which further supports the idea that MDD is a heterogeneous disease, rather than a singular construct, when studied over time [1, 40]. We recommend the advancement of symptom-specific and personalized approaches for both interventional and observational research. The sum scores of symptom questionnaires might obscure too much information potentially yielded by the individual symptoms. Moreover, a symptom-specific study approach may help the development of symptom-specific treatment strategies.

Author Statement

Acknowledgments and financial support

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organization for Health Research and Development (ZonMw, grant number 10-000-1002) and through the financial contributions of participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum).

Conflicts of interest

None.

References

1. Fried, E.I. and R.M. Nesse, *Depression sum-scores don't add up: why analyzing specific depression symptoms is essential.* BMC medicine, 2015. **13**(1): p. 72.
2. Conradi, H., J. Ormel, and P. de Jonge, *Symptom profiles of the DSM-IV-defined remission, recoverym relapse, and recurrence of depression: the role of the core symptoms.* Depression and Anxiety, 2012. **29**(7): p. 638-645.
3. Sakurai, H., et al., *Trajectories of individual symptoms in remitters versus non-remitters with depression.* Journal of Affective Disorders, 2013. **151**(2): p. 506-513.
4. Young, M.A., et al., *The temporal onset of individual symptoms in winter depression: differentiating underlying mechanisms.* Journal of affective disorders, 1991. **22**(4): p. 191.
5. Romera, I., et al., *Residual symptoms and functioning in depression, does the type of residual symptom matter? A post-hoc analysis.* BMC Psychiatry, 2013. **13**: p. 51.
6. Madhoo, M. and S.Z. Levine, *Initial Severity Effects on Residual Symptoms in Response and Remission: A STAR* D Study During and After Failed Citalopram Treatment.* Journal of clinical psychopharmacology, 2015. **35**(4): p. 450-453.
7. Minor, K.L., J.E. Champion, and I.H. Gotlib, *Stability of DSM-IV criterion symptoms for major depressive disorder.* Journal of Psychiatric Research, 2005. **39**(4): p. 415-420.
8. Paykel, E.S., et al., *Residual symptoms after partial remission: an important outcome in depression.* Psychological medicine, 1995. **25**(6): p. 1171-80.
9. Wardenaar, K.J., et al., *Symptom-specific course trajectories and their determinants in primary care patients with Major Depressive Disorder: Evidence for two etiologically distinct prototypes.* Journal of Affective Disorders, 2015. **179**: p. 38-46.
10. Rabin, A.S., N.J. Kaslow, and L.P. Rehm, *Changes in symptoms of depression during the course of therapy.* Cognitive Therapy and Research, 1984. **8**(5): p. 479-487.
11. Harada, E., et al., *Residual symptoms in patients with partial versus complete remission of a major depressive disorder episode: patterns of painful physical symptoms in depression.* Neuropsychiatric Disease and Treatment, 2016. **12**: p. 1599-1607.
12. Conradi, H.J., J. Ormel, and J.P. De, *Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study.* Psychological Medicine, 2011. **41**(6): p. 1165-1174.
13. Nil, R., S. Lütolf, and E. Seifritz, *Residual symptoms and functionality in depressed outpatients: A one-year observational study in Switzerland with escitalopram.* Journal of Affective Disorders, 2016. **197**: p. 245-250.
14. Paykel, E.S., G.L. Klerman, and B.A. Prusoff, *Treatment setting and clinical depression.* Archives of General Psychiatry, 1970. **22**(1): p. 11-21.
15. Menza, M., H. Marin, and R.S. Opper, *Residual symptoms in depression: can treatment be symptom-specific?* Journal of Clinical Psychiatry, 2003. **64**(5): p. 516-523.
16. Musliner, K.L., et al., *Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes.* Journal of Affective Disorders, 2016. **192**: p. 199-211.
17. Karp, J.F., et al., *Relationship of variability in residual symptoms with recurrence of major depressive disorder during maintenance treatment.* American Journal of Psychiatry, 2004. **161**(10): p. 1877-1884.
18. Torres, L.D., *Understanding the variability of depression symptoms in recovery: Life stress as context and consequence in the course of recurrent depression.* Dissertation abstracts international. B, The sciences and engineering, 2008. **68**(12-B): p. 8417-8417.
19. Hamilton, M., *Hamilton depression scale.* Internationale Skalen für Psychiatrie. Beltz Test, Göttingen, 1996: p. 93-96.
20. Curran, P.J. and D.J. Bauer, *The disaggregation of within-person and between-person effects in longitudinal models of change.* Annual review of psychology, 2011. **62**: p. 583-619.
21. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* International Journal of Methods in Psychiatric Research, 2008. **17**(3): p. 121-40.
22. Wittchen, H.-U., *Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review.* Journal of Psychiatric Research, 1994. **28**(1): p. 57-84.

23. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): Psychometric properties.* Psychological Medicine, 1996. **26**(3): p. 477-486.
24. Rush, A.J., et al., *The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression.* Biological Psychiatry, 2003. **54**(5): p. 573-583.
25. Rush, A.J., et al., *Self-reported depressive symptom measures: sensitivity to detecting change in a randomized, controlled trial of chronically depressed, nonpsychotic outpatients.* Neuropsychopharmacology, 2005. **30**(2): p. 405-416.
26. Hamilton, M., *Hamilton Depression Scale.* Group, 2004. **1**: p. 4.
27. Beck, A.T., R.A. Steer, and G.K. Brown, *Beck depression inventory-II.* San Antonio, 1996. **78**(2): p. 490-8.
28. Rhebergen, D., et al., *The three-year naturalistic course of major depressive disorder, dysthymic disorder and double depression.* Journal of affective disorders, 2009. **115**(3): p. 450-459.
29. Buysse, D.J., et al., *Prevalence, course, and comorbidity of insomnia and depression in young adults.* SLEEP, 2008. **31**(4): p. 473.
30. Lustberg, L. and C.F. Reynolds, *Depression and insomnia: questions of cause and effect.* Sleep medicine reviews, 2000. **4**(3): p. 253-262.
31. Baldwin, D.S. and G.I. Papakostas, *Symptoms of fatigue and sleepiness in major depressive disorder.* Journal of Clinical Psychiatry, 2006. **67**(SUPPL. 6): p. 9-15.
32. Wanders, R.B.K., et al., *Data-driven atypical profiles of depressive symptoms: Identification and validation in a large cohort.* Journal of affective disorders, 2015. **180**: p. 36.
33. Monden, R., et al., *Simultaneous decomposition of depression heterogeneity on the person-, symptom- and time-level: the use of three-mode principal component analysis.* PloS one, 2015. **10**(7): p. e0132765.
34. Williams, J., et al., *Cognitive reactivity, suicidal ideation and future fluency: preliminary investigation of a differential activation theory of hopelessness/suicidality.* Cognitive Therapy and Research, 2008. **32**(1): p. 83-104.
35. Oort, F.J., M.R. Visser, and M.A. Sprangers, *Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift.* Journal of clinical epidemiology, 2009. **62**(11): p. 1126-1137.
36. Sjöholm, L., C. Lavebratt, and Y. Forsell, *A multifactorial developmental model for the etiology of major depression in a population-based sample.* Journal of affective disorders, 2009. **113**(1): p. 66-76.
37. Beard, C., et al., *Network analysis of depression and anxiety symptom relationships in a psychiatric sample.* Psychological Medicine, 2016. **46**(16): p. 3359-3369.
38. Borkulo, C.v., et al., *Association of symptom network structure with the course of longitudinal depression.(Report).* JAMA Psychiatry, 2015. **72**(12): p. 1219.
39. Fried, E.I., et al., *What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis.* Journal of Affective Disorders, 2016. **189**: p. 314-320.
40. Cramer, A.O.J., et al., *Major depression as a complex dynamic system.* PLoS ONE, 2016. **11**(12).
41. Snaith, P., *What do depression rating scales measure?* The British Journal of Psychiatry, 1993. **163**(3): p. 293-298.
42. Arnow, B.A., et al., *Depression Subtypes in Predicting Antidepressant Response: A Report From the iSPOT-D Trial.* American Journal of Psychiatry, 2015. **172**(8): p. 743-750.
43. Chekroud, A.M., et al., *Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach.* JAMA psychiatry, 2017. **74**(4): p. 370-378.
44. Fava, M., *Pharmacological approaches to the treatment of residual symptoms.* Journal of Psychopharmacology, 2006. **20**(3 Suppl): p. 29-34.
45. Rafanelli, C., S.K. Park, and G.A. Fava, *New psychotherapeutic approaches to residual symptoms and relapse prevention in unipolar depression.* Clinical Psychology & Psychotherapy, 1999. **6**(3): p. 194-201.
46. Britton, W.B., et al., *Mindfulness-based cognitive therapy improves polysomnographic and subjective sleep profiles in antidepressant users with sleep complaints.* Psychotherapy and Psychosomatics, 2012. **81**(5): p. 296-304.
47. Manber, R., et al., *Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disor-der and insomnia.* Sleep, 2008. **31**(4): p. 489-495.
48. Webb, A.J., et al., *Effects of antihypertensive-drug class on interindividual variation in blood pressure and risk of stroke: a systematic review and meta-analysis.* The Lancet, 2010. **375**(9718): p. 906-915.
49. Streiner, D.v.L., *Regression toward the mean: its etiology, diagnosis, and treatment.* The Canadian Journal of Psychiatry, 2001. **46**(1): p. 72-76.

Supplementary material

Supplementary material table 1. Item response theory analysis of the IDS-SR 30

Item	Loading	h^2	a	b1	b2	b3
1. Falling asleep	0.096	0.009	0.164	5.147	-2.621	-0.080
2. Sleep during the night	0.105	0.011	0.179	-4.045	2.994	-1.731
3. Waking up too early	0.123	0.015	0.211	6.359	0.695	0.859
4. Sleeping too much	0.091	0.008	0.155	3.790	6.191	5.334
5. Feeling sad	0.622	0.386	1.350	-2.210	-0.306	1.414
6. Feeling irritable	0.418	0.175	0.784	-1.996	0.065	2.061
7. Anxious or tense	0.538	0.289	1.086	-2.029	-0.121	1.870
8. Response of mood	0.549	0.302	1.119	-0.486	0.764	2.120
9a. Mood in time of day	0.059	0.004	0.101	6.395	18.799	-8.212
10. Quality of mood	0.230	0.053	0.402	-3.505	0.356	0.684
11. Appetite	0.201	0.040	0.349	-1.454	3.635	-0.042
12. Weight	0.100	0.010	0.172	0.514	3.376	2.243
15. Concentration	0.446	0.199	0.847	-2.448	0.179	1.570
16. View of myself	0.260	0.068	0.458	-1.836	3.478	-3.304
17. View of my future	0.555	0.308	1.136	-2.564	0.994	1.219
18. Death or suicide	0.348	0.121	0.632	0.111	1.409	3.602
19. General interest	0.506	0.256	0.999	-1.395	1.398	0.620
20. Energy level	0.477	0.228	0.925	-2.041	-0.816	1.957
21. Capacity for pleasure	0.689	0.474	1.616	-1.263	0.684	1.999
22. Interest in sex	0.271	0.074	0.480	-0.658	0.339	1.761
23. Psychomotor retardation	0.360	0.130	0.657	0.437	-0.177	3.277
24. Psychomotor agitation	0.184	0.034	0.319	-0.655	0.070	5.193
25. Aches and pains	0.269	0.072	0.475	-3.100	0.567	3.820
26. Sympathetic arousal	0.310	0.096	0.554	-2.184	1.125	4.956
27. Panic/phobic	0.233	0.055	0.409	0.047	0.484	3.117
28. Constipation/diarrhea	0.183	0.034	0.317	0.037	2.179	3.729
29. Interpersonal sensitivity	0.306	0.094	0.548	-2.238	-2.190	1.095
30. Leaden paralysis	0.411	0.169	0.767	-2.764	-0.737	0.620

Note. Loadings denote as component loadings. h^2 denotes as the variances accounted for in each item. a denotes item's discrimination parameter. b1, b2, and b3 denotes the generalized partial credit model parameter estimates.

Supplementary material table 2. Sensitivity analysis, values adjusted for a history of chronic depression, chronic somatic disease, age, or the use of antidepressants

Item	Baseline mean adjusted for chronic depression	Baseline mean adjusted for chronic somatic disease	Baseline Mean adjusted for the use of antidepressants	One-year slope adjusted for chronic depression	One-year slope adjusted for chronic somatic disease	One-year slope adjusted for age	One-year slope adjusted for the use of antidepressants	Nine-year slope adjusted for chronic depression	Nine-year slope adjusted for chronic somatic disease	Nine-year slope adjusted for age	Nine-year slope adjusted for the use of antidepressants	FVU adjusted for chronic depression	FVU adjusted for chronic somatic disease	FVU adjusted for the use of antidepressants
1. Falling asleep	1.26 (0.05)	1.34 (0.07)	1.31 (0.06)	-0.459 (0.07)	-0.459 (0.07)	-0.459 (0.04)	-0.458 (0.07)	-0.001 (0.010)	-0.001 (0.010)	0.000 (0.010)	0.000 (0.010)	0.464 (0.020)	0.473 (0.027)	0.475 (0.022)
2. Sleep during the night	1.57 (0.05)	1.53 (0.06)	1.60 (0.05)	-0.224 (0.06)	-0.227 (0.06)	-0.243 (0.04)	-0.227 (0.06)	0.049 (0.008)	0.048 (0.008)	0.047 (0.008)	0.049 (0.008)	0.620 (0.019)	0.597 (0.023)	0.618 (0.019)
3. Waking up too early	0.77 (0.05)	0.73 (0.06)	0.80 (0.05)	-0.180 (0.08)	-0.186 (0.08)	-0.202 (0.04)	-0.188 (0.08)	0.042 (0.011)	0.043 (0.011)	0.040 (0.011)	0.043 (0.011)	0.361 (0.021)	0.340 (0.027)	0.360 (0.022)
4. Sleeping too much	0.63 (0.04)	0.67 (0.05)	0.54 (0.04)	-0.248 (0.08)	-0.256 (0.08)	-0.246 (0.03)	-0.254 (0.08)	-0.015 (0.013)	-0.015 (0.013)	-0.019 (0.013)	-0.018 (0.012)	0.418 (0.022)	0.428 (0.027)	0.391 (0.022)
5. Feeling Sad	1.52 (0.04)	1.67 (0.05)	1.60 (0.04)	-0.926 (0.07)	-0.937 (0.07)	-0.938 (0.03)	-0.937 (0.07)	-0.035 (0.009)	-0.033 (0.009)	-0.035 (0.009)	-0.035 (0.009)	0.590 (0.018)	0.579 (0.024)	0.577 (0.020)
6. Feeling irritable	1.42 (0.04)	1.44 (0.05)	1.50 (0.04)	-0.711 (0.07)	-0.720 (0.07)	-0.716 (0.03)	-0.720 (0.07)	-0.014 (0.009)	-0.014 (0.009)	-0.013 (0.009)	-0.014 (0.009)	0.598 (0.019)	0.585 (0.023)	0.602 (0.019)
7. Anxious or tense	1.42 (0.04)	1.46 (0.05)	1.45 (0.04)	-0.826 (0.07)	-0.836 (0.07)	-0.836 (0.03)	-0.837 (0.07)	-0.002 (0.008)	-0.003 (0.008)	-0.003 (0.008)	-0.004 (0.008)	0.613 (0.019)	0.610 (0.024)	0.618 (0.019)
8. Response of mood	0.91 (0.04)	1.06 (0.05)	0.89 (0.04)	-0.742 (0.07)	-0.746 (0.07)	-0.750 (0.03)	-0.745 (0.07)	-0.035 (0.010)	-0.034 (0.010)	-0.036 (0.010)	-0.036 (0.010)	0.424 (0.021)	0.444 (0.027)	0.420 (0.022)
9a. Mood in time of day	0.70 (0.04)	0.79 (0.06)	0.61 (0.04)	-0.073 (0.07)	-0.061 (0.07)	-0.059 (0.04)	-0.060 (0.07)	-0.014 (0.011)	-0.011 (0.011)	-0.010 (0.011)	-0.013 (0.011)	0.433 (0.021)	0.435 (0.027)	0.400 (0.023)
10. Quality of mood	1.65 (0.05)	1.65 (0.06)	1.62 (0.05)	-0.602 (0.07)	-0.600 (0.07)	-0.596 (0.03)	-0.600 (0.07)	-0.044 (0.009)	-0.044 (0.009)	-0.044 (0.009)	-0.044 (0.009)	0.558 (0.020)	0.537 (0.026)	0.570 (0.021)
11. Appetite	1.12 (0.05)	1.09 (0.06)	1.07 (0.05)	-0.527 (0.07)	-0.518 (0.07)	-0.511 (0.04)	-0.516 (0.07)	-0.046 (0.009)	-0.047 (0.009)	-0.046 (0.009)	-0.046 (0.009)	0.497 (0.021)	0.484 (0.027)	0.504 (0.022)
12. Weight	1.09 (0.05)	1.04 (0.06)	1.04 (0.05)	-0.189 (0.06)	-0.191 (0.06)	-0.188 (0.04)	-0.190 (0.06)	-0.001 (0.008)	-0.003 (0.008)	-0.002 (0.008)	-0.002 (0.008)	0.581 (0.019)	0.561 (0.025)	0.602 (0.020)
15. Concentration	1.42 (0.04)	1.52 (0.05)	1.51 (0.04)	-0.802 (0.07)	-0.808 (0.07)	-0.809 (0.03)	-0.808 (0.07)	-0.025 (0.008)	-0.025 (0.008)	-0.026 (0.008)	-0.026 (0.008)	0.631 (0.020)	0.572 (0.025)	0.657 (0.019)
16. View of myself	1.48 (0.05)	1.75 (0.07)	1.53 (0.06)	-0.582 (0.07)	-0.581 (0.07)	-0.577 (0.04)	-0.580 (0.07)	-0.014 (0.009)	-0.012 (0.009)	-0.012 (0.009)	-0.014 (0.009)	0.519 (0.021)	0.507 (0.026)	0.520 (0.022)
17. View of my future	1.25 (0.04)	1.40 (0.05)	1.29 (0.04)	-0.568 (0.08)	-0.575 (0.08)	-0.582 (0.03)	-0.575 (0.08)	0.011 (0.009)	0.012 (0.009)	0.011 (0.009)	0.011 (0.009)	0.604 (0.019)	0.556 (0.026)	0.616 (0.020)
18. Death or suicide	0.74 (0.04)	0.83 (0.05)	0.70 (0.04)	-0.640 (0.08)	-0.649 (0.07)	-0.651 (0.03)	-0.647 (0.07)	-0.004 (0.013)	-0.004 (0.013)	-0.004 (0.013)	-0.005 (0.013)	0.313 (0.020)	0.317 (0.025)	0.345 (0.022)
19. General interest	1.17 (0.04)	1.25 (0.06)	1.18 (0.05)	-0.764 (0.07)	-0.773 (0.07)	-0.773 (0.03)	-0.773 (0.07)	-0.018 (0.009)	-0.018 (0.009)	-0.018 (0.009)	-0.018 (0.009)	0.550 (0.020)	0.511 (0.027)	0.537 (0.022)
20. Energy level	1.61 (0.04)	1.62 (0.05)	1.58 (0.04)	-0.986 (0.07)	-0.993 (0.07)	-0.995 (0.03)	-0.993 (0.07)	-0.017 (0.008)	-0.017 (0.008)	-0.017 (0.008)	-0.017 (0.008)	0.619 (0.019)	0.566 (0.025)	0.612 (0.019)
21. Capacity for pleasure	1.08 (0.03)	1.16 (0.05)	1.09 (0.04)	-0.952 (0.08)	-0.956 (0.07)	-0.960 (0.03)	-0.956 (0.07)	-0.022 (0.009)	-0.022 (0.009)	-0.024 (0.009)	-0.023 (0.009)	0.487 (0.021)	0.475 (0.027)	0.482 (0.022)
22. Interest in sex	1.22 (0.05)	1.24 (0.06)	1.14 (0.05)	-0.428 (0.07)	-0.434 (0.07)	-0.442 (0.04)	-0.433 (0.07)	0.030 (0.010)	0.030 (0.010)	0.030 (0.010)	0.030 (0.010)	0.536 (0.020)	0.522 (0.026)	0.540 (0.021)
23. Psychomotor retardation	0.96 (0.04)	0.96 (0.05)	0.89 (0.04)	-0.675 (0.08)	-0.674 (0.08)	-0.679 (0.03)	-0.675 (0.08)	-0.008 (0.011)	-0.009 (0.011)	-0.010 (0.011)	-0.010 (0.011)	0.363 (0.021)	0.368 (0.027)	0.376 (0.022)
24. Psychomotor agitation	1.13 (0.04)	1.16 (0.05)	1.19 (0.05)	-0.487 (0.07)	-0.492 (0.07)	-0.495 (0.03)	-0.492 (0.07)	-0.037 (0.009)	-0.038 (0.009)	-0.038 (0.009)	-0.038 (0.009)	0.446 (0.021)	0.463 (0.026)	0.474 (0.022)
25. Aches and pains	1.36 (0.04)	1.17 (0.05)	1.40 (0.04)	-0.372 (0.07)	-0.377 (0.07)	-0.383 (0.03)	-0.376 (0.07)	0.038 (0.009)	0.038 (0.009)	0.039 (0.009)	0.040 (0.009)	0.628 (0.019)	0.601 (0.024)	0.651 (0.019)
26. Sympathetic arousal	1.12 (0.03)	1.02 (0.04)	1.08 (0.04)	-0.536 (0.07)	-0.544 (0.07)	-0.555 (0.03)	-0.544 (0.07)	0.014 (0.009)	0.013 (0.009)	0.013 (0.009)	0.014 (0.009)	0.557 (0.020)	0.542 (0.025)	0.545 (0.021)
27. Panic/Phobic	1.03 (0.04)	1.09 (0.06)	1.00 (0.04)	-0.596 (0.07)	-0.601 (0.07)	-0.600 (0.04)	-0.600 (0.07)	0.004 (0.009)	0.004 (0.009)	0.005 (0.009)	0.003 (0.009)	0.489 (0.020)	0.495 (0.027)	0.498 (0.022)
28. Constipation/diarrhea	0.91 (0.04)	0.72 (0.05)	0.92 (0.04)	-0.408 (0.07)	-0.411 (0.07)	-0.416 (0.03)	-0.414 (0.07)	0.001 (0.010)	0.000 (0.010)	0.002 (0.010)	0.002 (0.010)	0.526 (0.020)	0.502 (0.025)	0.518 (0.021)
29. Interpersonal sensitivity	1.38 (0.05)	1.55 (0.06)	1.46 (0.05)	-0.528 (0.07)	-0.532 (0.07)	-0.524 (0.04)	-0.531 (0.07)	-0.017 (0.009)	-0.015 (0.009)	-0.013 (0.009)	-0.016 (0.009)	0.616 (0.019)	0.633 (0.023)	0.629 (0.020)
30. Leaden paralysis	1.79 (0.04)	1.80 (0.05)	1.78 (0.04)	-0.697 (0.06)	-0.694 (0.06)	-0.696 (0.03)	-0.693 (0.06)	-0.003 (0.009)	-0.004 (0.009)	-0.002 (0.009)	-0.003 (0.009)	0.636 (0.018)	0.634 (0.024)	0.633 (0.020)

Note. Mean values, standard error (in parentheses), 1-year slope, sum-score adjusted 9-year slope, and fraction of variance unexplained (FVU; $\Sigma(y_i - \hat{y}_i)^2/\Sigma(y_i - \bar{y})^2$) for each of the individual symptoms of the Inventory of Depressive Symptomatology- Self-Report (IDS-SR) adjusted for history of chronic depression (24 out of 48 months prior to baseline), chronic somatic disease, age (only slopes), or the use of antidepressants.

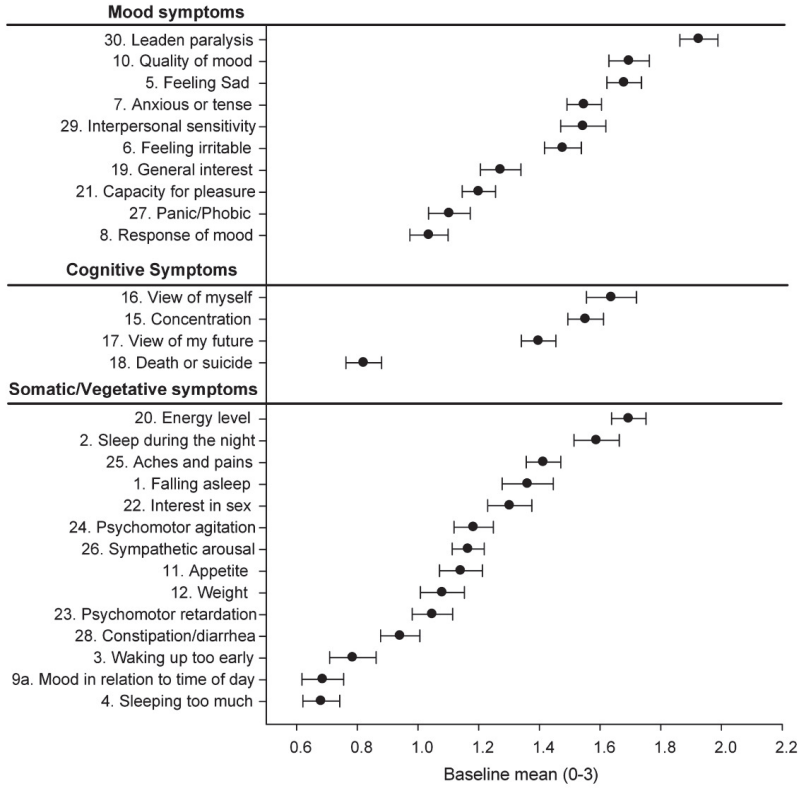


Figure S1. Group-level mean item scores over the course of 9 years.

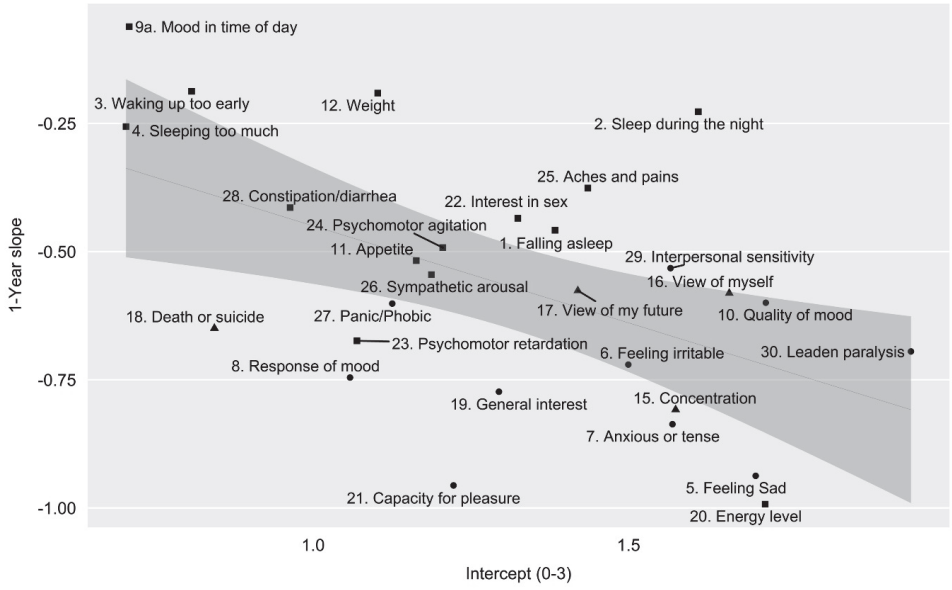
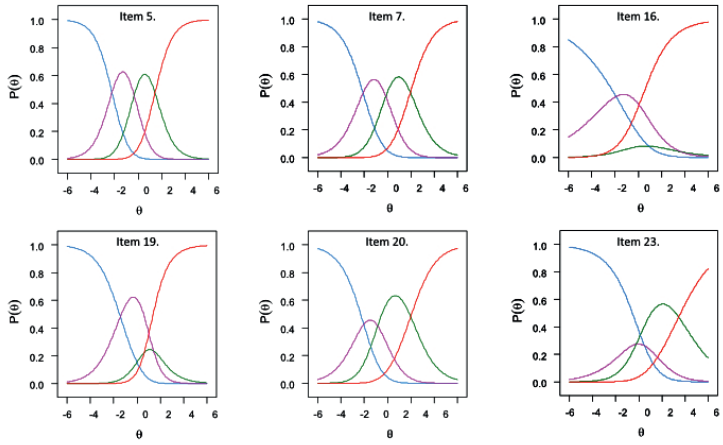


Figure S2. Supplementary material. One-year slope in relation to level of severity (intercept).



Item	Loading	h^2	a	b1	b2	b3
5. Feeling sad	0.605	0.366	1.293	-2.223	-0.334	1.444
7. Anxious or tense	0.504	0.254	0.993	-2.104	-0.148	1.963
16. View of myself	0.244	0.060	0.428	-1.919	3.714	-3.577
19. General interest	0.493	0.243	0.964	-1.430	1.423	0.618
20. Energy level	0.469	0.220	0.903	-2.046	-0.848	1.989
23. Psychomotor retardation	0.338	0.114	0.611	0.489	-0.228	3.454

Note. Loadings denote as component loadings. h^2 denotes as the variances accounted for in each item. a denotes item's discrimination parameter. b1, b2, and b3 denotes the generalized partial credit model parameter estimates. Θ denotes the latent variable. PΘ denotes the probability of corresponding to latent variable.

Figure S3. Supplementary material. Item response theory analysis of the IDS-SR 6.

[1] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands;
[2] Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, VU University Medical Center, and GGZ inGeest, Amsterdam, The Netherlands;
[3] Institute of Psychology, Leiden University, Leiden, The Netherlands;

Chapter 3

Neuroticism and chronicity as predictors of 9-year course of individual depressive symptoms

van Eeden, W. A.[1], van Hemert, A. M.[1], Carlier, I. V., Penninx, B. W.[2], Spinhoven, P.[1,3], & Giltay, E. J.[1]

(2019). Journal of affective disorders, 252, 484-49

Abstract

Background: The large between-person differences in symptomatology suggest that Major Depressive Disorder (MDD) is a heterogeneous psychiatric disorder. However, symptom-specific prospective studies are scarce. We hypothesized that chronicity (i.e., being depressed for 24 months during a patient's preceding 48 months at baseline) and neuroticism at baseline would predict adverse course trajectories over 9 years of follow up with differential magnitudes for individual depressive symptoms.

Methods: In total, 560 patients with a current MDD were included from the Netherlands Study of Depression and Anxiety (NESDA-cohort). We used a multivariate linear mixed model with repeated measures, with a history of chronicity and neuroticism separately as main independent variables and with Inventory of Depressive Symptomatology self-report (IDS-SR) item scores as outcome variables. For each individual symptom, the model was adjusted for age, gender, and baseline depression severity.

Results: Patients were on average 42.7 (*SD* = 12.1) years old and 64.7% were women. Patients with chronic depression or high levels of neuroticism showed similar absolute rates of decline over time compared to their counterparts. However, because symptoms had higher starting points for mood, cognitive, and somatic/vegetative symptoms (in that order), symptom severity remained higher over time. Chronicity and neuroticism were especially linked to persistent low self-esteem and high interpersonal sensitivity.

Limitations: Neuroticism is partly state dependent and likely affected by depression severity.

Conclusions: Chronicity and neuroticism predict long-term persistence of diverse psychiatric symptoms, in particular low self-esteem and high interpersonal sensitivity.

Highlights

1. A history of chronic depression and level of neuroticism are associated to similar symptom profiles

2. Chronic depression is associated to mood and cognitive symptoms and to a lesser extent to somatic/vegetative symptoms

3. Neuroticism is associated to mood and cognitive symptoms but on average not to somatic vegetative symptoms

4. Patients with chronic depression and/or high levels of neuroticism report particularly on low self-esteem and high interpersonal sensitivity items. This may be of importance in the development of personalized treatment.

3.1 Introduction

Major Depressive Disorder (MDD) is a heterogeneous psychiatric illness with large between-person differences in both symptomatology and course trajectories [1, 2]. Although several predictive variables have been established for a more chronic course, most feature low predictive power [e.g., 3, 4]. Of these, 'preceding chronic depression' and a 'high level of neuroticism' are two of the stronger predictors [5, 6]; however, their predictive value diminishes after adjusting for baseline severity scores, which may serve as an intermediary factor [6-8]. It is currently unknown whether chronicity or neuroticism affect the course of symptoms equally, or affect a particular subset of symptoms, but not others. Moreover, the importance of symptom-specific research is beginning to emerge in the field of psychiatry (Fried and Nesse, 2015).

A previous analysis in the Netherlands Study of Depression and Anxiety (NESDA) demonstrated that MDD persisted over the course of 4 years in 53.0% of the patients with chronic MDD at baseline versus 27.8% of patients with nonchronic MDD at baseline; this is consistent with findings from others [3, 6, 9-12]. Acknowledging the importance of a preceding depressive course led to the addition of persistent depressive disorder (i.e. a combination of dysthymia *and* chronic depression) in the DSM-5 [13].

Neuroticism is one of the five major dimensions of personality (Five Factor Model; FFM; Costa and McCrae, 1992) and reflects the tendency to respond to distress by being moody, anxious, or sad. High neuroticism increases the risk of MDD, its unfavorable course, and a higher relapse rate [5, 14-22].

Chronic depression and neuroticism seem to be linked, i.e. chronically depressed patients generally show higher levels of neuroticism than patients with an episodic depressive course [5, 23-26]. Also, neuroticism represents a trait-like substrate in chronic depression and is more state-dependent when the depression has an episodic course, regardless of eventual depression remission [22]. Because of their associations with early onset, childhood maltreatment, Cluster C personality disorders, and genetics [27-30], neuroticism and chronic depression may share etiological factors [22] and thus represent partly overlapping constructs [22].

Most research has focused on MDD as a latent variable construct, representing a single underlying disorder, where the level of severity is measured as a sum score on self-report questionnaires [e.g. 31, 32, 33]. However, given the heterogeneous nature of MDD, focusing on individual symptoms (rather than on sum scores) may yield important new insights into the relationship between history of chronicity, personality traits, and the course of MDD [1, 5, 21].

Previous cross-sectional studies conducted in a nonclinical sample found that risk factors correlated with individual depressive symptoms with different strengths [34-37]. Two studies analyzed chronicity and neuroticism in relation to MDD symptom profiles in clinical samples. One cross-sectional study examined the symptom-specific associations with both neuroticism and chronicity among 1,015 MDD patients [38]. Fatigue and suicidal ideation were significantly associated with chronicity, and appetite/weight and sleeping problems were associated with neuroticism [38]. The second prospective study, with 20 years of follow-up, examined the symptom profiles of 450 MDD patients [39]. Patients with long-term depression more frequently reported symptoms of disturbed memory, low self-esteem, hopelessness, fear of everyday tasks, fear of being alone, and suicidal ideations [39]. We are not aware of previous studies that have analyzed the predictive value of neuroticism and chronicity for the course of individual depressive symptoms in patients with MDD. Such findings might increase our ability to target (psychotherapeutic) treatment strategies earlier and, more specifically, on certain symptom patterns.

The present study aimed to examine whether chronic MDD, defined as being depressed for at least 2 years (during a patient's past 4 years before baseline) and level of neuroticism could predict the 9-year trajectory of individual depressive symptoms. In particular, the focus was on the symptom-specific differences in this regard. It was hypothesized that chronic MDD and neuroticism at baseline would be associated with the course of some specific symptoms, rather than depression as a homogeneous construct, with similar associations for each symptom. Because previous studies suggested that chronic depression and neuroticism may represent overlapping constructs, we expected these variables to be associated with the same depressive symptoms over time. Further, we hypothesized that the average severity of mood and cognitive MDD symptoms would tend to remain at a higher average level in the presence of chronicity and neuroticism at baseline.

3.2 Materials and Methods

3.2.1 Study sample and procedures

Participants were selected from the Netherlands Study of Depression and Anxiety (NESDA) cohort. A detailed description of the NESDA design and sampling procedures is published elsewhere [40]. The first wave (baseline) started in 2004 and ended in 2007, and the sixth wave of measurement at the 9-year follow-up finished in 2016. The Composite International Diagnostic Interview (CIDI WHO, version 2.1) was used to assess the presence of depressive and anxiety disorders according to the Diagnostic and Statistical Manual of Mental Disorders-Fourth Edition (DSM-IV; Wittchen, 1994). The baseline sample (*n* = 2,981) consisted of patients with anxiety and/or depressive disorders and normal controls. Postbaseline, follow-up assessments were conducted at 1 year (*n* = 2,445; 82.0%), 2 years (*n* = 2,596; 87.1%), 4 years (*n* = 2,256; 80.6%), 6 years (*n* = 2,256; 75.7%), and 9 years (*n* = 2,069; 69.4%) (Penninx et al., 2008).

For the present study, we selected all patients who met DSM-IV criteria for MDD within one month prior to the baseline assessment. Furthermore, our patients needed to have completed the Inventory of Depressive Symptomatology-Self-Report (IDS-SR; see Measures) for at least four of the six time point assessments, which needed to include the baseline assessment. This resulted in a study sample of 558 participants. Because of some missing items, the two subsets differed in sample size: *n* = 550 participants for the analysis of chronicity and *n* = 553 for the analysis of neuroticism.

3.2.2 Measures

3.2.2.1 Independent variables: Chronic depression and neuroticism

Chronic depression at baseline was measured using the Life Chart Interview method [41], a standardized interview designed to retrospectively assess the course of psychopathology. The Life Chart Interview uses age- and calendar-linked life events over a patient's past 4 years and then assesses the presence and severity of symptoms during this period. When patients were depressed for ≥ 50% during and between these life events, they were defined as being chronically depressed [29]. This is similar to the DSM-5 criteria for persistent depressive disorder, which states that criteria for MDD should be met for at least 2 years with a maximum of 2 months without symptoms [13].

Neuroticism was assessed at baseline using the NEO five-factor inventory (NEO-FFI), i.e. the 60-item version of the longer 240-item NEO Personality Inventory Revised (NEO-PI-R). The NEO-FFI consists of five factors that measure the Big Five personality traits: neuroticism, extraversion, agreeableness, conscientiousness, and openness to experience. Neuroticism was assessed with 12 aggregated items on a 5-point scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The NEO-FFI scale has good internal and test-retest reliability [42] and is a valid instrument for both clinical and healthy samples [43, 44]. Although neuroticism is generally considered to be a trait throughout a person's life, it is also known to have state dependencies (Spinhoven, Does, Omel, Zitman, and Penninx, 2013). When assessing the interclass correlation coefficients (ICCs) of neuroticism between baseline, 2-year and 4-year follow-up, we found an ICC value of 0.672 ($F = 3.05$, $p < .001$), suggesting moderately high interclass correlations. Note that only the baseline level of neuroticism was used as predictor variable.

3.2.2.2 Dependent variables: IDS items

The individual item scores of the Inventory of Depressive Symptomatology Self-Report (IDS-SR) were used as the outcome measure for severity and course of depressive symptoms [31, 32]. The IDS-SR consists of 30 equally weighted items, rated on a 4-point Likert scale (0-3). The scale includes all symptoms of depression including melancholic, atypical, and anxious symptoms. Moreover, several additional symptoms have been added, for example, sympathetic arousal, pessimism, and interest in sex. The IDS-SR has adequate reliability, acceptable validity, good responsiveness, and good discriminative ability with Cronbach's alphas ranging from 0.92-0.94 [32]. In the present study sample, the Cronbach's alphas were 0.83, 0.89, 0.89, 0.90, 0.90, and 0.90 for the six time points from baseline to the 9-year follow up. Alphas were only slightly different for patient groups with chronic, nonchronic, high neuroticism (above median 43), or low neuroticism (below median 43), with Cronbach's alphas at baseline equaling 0.80, 0.85, 0.77, and 0.83, respectively. Note that these Cronbach's alpha values need to be interpreted with caution because they are not particularly well suited for ordinal data and tend to increase when questionnaires contain a larger number of items (Sijtsma, 2009). Items 11/12 ("increased/decreased appetite") and Items 13/14 ("weight gain/weight loss") contain opposite features. In order to maintain psychometric similarity between items, these item pairs were combined into one ordinal item, yielding 28

items for the current analyzes [32]. We applied previously identified symptom clusters in our results, including 10 mood symptoms, 14 somatic/vegetative symptoms, and four cognitive symptoms [see also Figures 1 and 2; 45].

3.2.3 Statistical analysis

Using chi-square analysis for categorical variables and independent t-tests for continuous variables, we evaluated and described baseline clinical characteristics and demographic variables across patients with chronic/nonchronic depression and high/low levels of neuroticism. For this purpose, neuroticism was dichotomized using a median split (median = 43); however, in all subsequent analyzes, neuroticism was analyzed as a continuous variable. A multivariate linear mixed model with repeated measures was used, with item score as the outcome variable and chronicity or neuroticism as the main independent variables. The models were adjusted for age, gender, and baseline depression severity. Because of positively skewed distributions, we $\log_e$ transformed the dependent variable scores and $\log_e$–$\log_e$ transformed time, which improved the fit of our linear models. These analyzes were repeated for each individual symptom separately, once with chronicity as the main independent variable and once with neuroticism as the main independent variable. Because this resulted in statistical tests for each of the 28 items, our outcomes were adjusted for multiple testing using the Bonferroni correction, which yielded a critical significance level of p = .002. Moreover, additional analyzes were computed per item in which we adjusted the effects of chronicity by adjusting for neuroticism and vice versa. This resulted in the predictive values of chronicity and neuroticism independent from each other. In order to yield beta coefficients that can be compared between symptoms, all outcome and independent variables were standardized (i.e., *z* scores). Our analyzes focused on the standardized difference of severity (SDS) as the outcome measure, which represents the difference (units of *SD*s) of each of the symptoms between patients with chronicity and patients with higher levels of neuroticism (continuous, in units of *SD*s) compared to their counterparts. Subsequently, forest plots with SDS values and error bars representing standardized errors (SE) were assigned to each individual symptom and sorted by the symptom cluster [45] and SDS value. Analyzes were performed using SPSS, version 23.

3.3 Results

3.3.1 Sociodemographic characteristics at baseline

Characteristics of the study sample are presented in Table 1. At baseline, age ranged from 18-64 (mean 42.7, *SD* = 12.1) years, and 64.7% were women (*n* = 362). Also, at baseline, the mean sum IDS-SR score was 34.6 (*SD* = 11.1), indicating a moderate depression severity in our patients. The level of neuroticism was 3.02 points (SE 0.57; standardized difference: 0.463) higher for patients with a chronic depression at baseline, resulting in a significant t-test for independent samples (t = 5.31, p < .001). Of the 558 included patients, 204 had a history of chronic MDD and 281 experienced high levels of neuroticism (> 43). In the whole sample, 21.5% had both a chronic depression and a neuroticism level above the median.

Table 1. Sociodemographic and clinical characteristics

	Chronic depression (N= 204)	Non-chronic depression (N=346)	t/chi-square	p	Neuroticism score >43 (N= 272)	Neuroticism score <43 (N=281)	t/chi-square	p
Age in years (mean, SD)	44.84 (11.79)	41.31 (12.18)	5.71	**0.017**	40.61 (11.88)	44.72 (12.12)	0.04	0.848
Female (%)	64.04	65.90	0.27	0.606	69.12	61.21	3.80	0.051
North-European ethnicity (%)	93.14	96.53	3.28	0.070	97.06	93.95	3.09	0.079
Education level (%)			3.15	0.208			8.99	**0.011**
Elementary or lower	9.31	5.49			4.63	4.63		
General intermediate or secondary	65.20	66.18			68.75	62.99		
College or university	25.49	28.32			22.43	32.38		
Psychotherapy*, yes (%)	76.56	75.46	0.01	0.920	80.68	69.77	4.99	**0.025**
Comorbid anxiety disorder, yes (%)	74.02	60.11	11.34	**0.003**	77.94	52.67	38.85	**<0.001**
Chronic depression, yes (%)	100	0			43.87	29.89	10.74	**0.001**
Total score IDS								
Baseline	37.99 (10.77)	32.40 (10.93)	5.99	**<0.001**	39.40 (9.91)	29.66 (10.31)	11.46	**<0.001**
year 1	30.36 (12.78)	23.43 (12.00)	6.03	**<0.001**	30.71 (12.86)	21.45 (10.81)	8.71	**<0.001**
year 2	28.61 (13.04)	21.93 (11.51)	6.20	**<0.001**	27.96 (12.60)	21.05 (11.46)	6.67	**<0.001**
year 4	27.28 (13.45)	22.01 (12.69)	4.50	**<0.001**	27.18 (13.04)	20.63 (12.51)	5.92	**<0.001**
year 6	27.59 (13.38)	20.94 (12.41)	5.55	**<0.001**	27.46 (13.03)	19.44 (12.08)	7.07	**<0.001**
year 9	26.99 (13.72)	19.75 (11.81)	5.77	**<0.001**	26.66 (12.94)	17.85 (11.32)	7.59	**<0.001**
Neo-FFI								
Neuroticism score (mean, SD)	44.64 (5.85)	41.62 (6.73)	5.71	**0.017**	48.04 (3.32)	37.64 (4.62)		
Extraversion score (mean, SD)	30.18 (6.32)	33.88 (6.44)	0.48	0.490	29.99 (6.30)	34.85 (6.10)	0.02	0.876
Openness (mean, SD)	37.48 (6.78)	38.37 (5.11)	3.46	0.063	37.88 (6.53)	38.22 (5.80)	3.96	**0.047**
Agreebleness (mean, SD)	42.25 (5.52)	43.10 (5.11)	1.82	0.178	42.08 (5.49)	34.54 (4.97)	1.59	0.208
Conscientiousness (mean, SD)	37.42 (7.21)	39.88 (6.26)	1.08	0.300	36.76 (6.46)	41.11 (6.24)	0.28	0.598
Antidepressants								
TCA (%)	9.62	2.31	2.72	0.099	4.04	2.49	1.06	0.304
SSRI (%)	37.26	26.01	7.70	**0.006**	33.09	27.40	2.12	0.145
Other (%)	10.11	12.26	0.60	0.437	13.96	8.87	2.25	0.133
no AD (%)	42.01	59.42	15.85	**<0.001**	48.94	61.24	6.59	**0.010**

Note. AD = antidepressants. TCA = tricyclic antidepressants. SSRI = selective serotonin reuptake inhibitors. AD = antidepressants. TCA = tricyclic antidepressants. Patients are selected twice, once for chronic/nonchronic depression and once for high/low neuroticism. Sample size is unequal due to missing items. Chi-square = Pearson Chi-square, two-sided.

* Due to missing values, assessed in 62.8% of sample

3.3.2 Chronic depression

Patients with a chronic depression at baseline were older, had higher neuroticism scores, and were more likely to use antidepressants than patients with no chronic depression at baseline (Table 1).

Chronic MDD was independently associated with a higher severity of depression over the course of 9 years (i.e., IDS scores adjusted for age, gender, and baseline depression severity; SDS = 0.131, t = 11.023, p < .001). This translated into a 0.131 *SD* higher average score for each of the 28 IDS items. In general, highly parallel courses were found for both chronically and nonchronically depressed patients. Although, the interaction terms between time and chronicity were significant, the effect sizes were small and deemed not clinically important (interaction = –0.022, t = 4.30, p = .012). Symptom-specific course trajectories are presented in Figure 1 of the supplementary material. When we adjusted the effect of a history of chronic depression for baseline neuroticism, this main effect remained significant (SDS = 0.078, t = 6.467, p < .001).

Next, we analyzed putative differential effects of chronicity on the 9-year course of 28 IDS items (Table 2 and Figure 1). Important differences emerged, i.e. the SDS ranged from –0.080 (“weight”) through 0.275 (“view of myself”). All individual symptoms showed on average a higher severity for patients with chronic MDD compared to nonchronic MDD, except for “weight” (i.e., Item 12). All mood and cognitive symptoms were more severe in patients with chronic depression, especially interpersonal sensitivity and low self-esteem. Chronicity was less related to somatic/vegetative symptoms, of which only about half the symptoms were more severe for chronic patients. When using the Bonferroni-corrected critical level of significance of .002, significant associations were found with four of four cognitive symptoms, six of 10 mood symptoms, and three of 14 somatic/vegetative symptoms. When adjusting the associations of chronicity with individual symptoms for level of neuroticism at baseline, the outcomes remained largely significant for most symptoms, except for Items 6 “feeling irritable,” 10 “quality of mood,” 19 “general interest,” and 27 “panic/phobia” (Table 2).

Table 2. IDS symptoms in ralation to chronicity and neuroticism

	Baseline mean item score (SE)				Chronic			Neuroticism		
Item	Chronic	Non-chronic	High N	Low N	Crude SDS	p	Adjusted SDS	p	Crude SDS	p	Adjusted SDS	p
1. Falling asleep	1.515 (0.08)	1.234 (0.06)	1.433 (0.07)	1.208 (0.07)	0.121 (0.08)	0.118	0.090 (0.08)	0.258	0.035 (0.04)	0.424	0.034 (0.04)	0.448
2. Sleep during the night	1.683 (0.08)	1.634 (0.06)	1.663 (0.06)	1.475 (0.07)	0.007 (0.06)	0.906	-0.015 (0.06)	0.797	-0.025 (0.03)	0.448	-0.012 (0.03)	0.683
3. Waking up too early	0.744 (0.08)	1.516 (0.05)	0.780 (0.06)	0.668 (0.07)	0.021 (0.06)	0.748	-0.017 (0.07)	0.791	0.050 (0.04)	0.172	0.048 (0.04)	0.202
4. Sleeping too much	0.727 (0.06)	0.635 (0.04)	0.701 (0.05)	0.627 (0.05)	0.031 (0.06)	0.598	0.015 (0.06)	0.804	-0.011 (0.03)	0.750	-0.009 (0.03)	0.801
5. Feeling Sad	1.898 (0.06)	1.497 (0.04)	1.896 (0.05)	1.362 (0.05)	0.247 (0.05)	<0.001	0.179 (0.05)	<0.001	0.165 (0.03)	<0.001	0.155 (0.03)	<0.001
6. Feeling irritable	1.456 (0.06)	1.375 (0.05)	1.651 (0.05)	1.117 (0.05)	0.122 (0.06)	0.018	0.045 (0.05)	0.374	0.144 (0.03)	<0.001	0.143 (0.03)	<0.001
7. Anxious or tense	1.620 (0.06)	1.355 (0.04)	1.708 (0.04)	1.156 (0.05)	0.195 (0.05)	<0.001	0.113 (0.05)	0.021	0.183 (0.03)	<0.001	0.174 (0.03)	<0.001
8. Response of mood	1.186 (0.07)	0.893 (0.05)	1.192 (0.05)	0.781 (0.05)	0.176 (0.05)	<0.001	0.120 (0.05)	0.018	0.044 (0.03)	0.116	0.033 (0.03)	0.246
9a. Mood in time of day	0.654 (0.07)	0.717 (0.07)	0.755 (0.06)	0.638 (0.06)	0.041 (0.06)	0.499	0.010 (0.06)	0.868	0.041 (0.03)	0.229	0.042 (0.03)	0.226
10. Quality of mood	1.743 (0.06)	1.634 (0.05)	1.780 (0.05)	1.538 (0.07)	0.138 (0.06)	0.016	0.075 (0.06)	0.194	0.085 (0.03)	0.008	0.080 (0.03)	0.014
11. Appetite	1.093 (0.07)	1.067 (0.05)	1.274 (0.06)	0.832 (0.05)	0.044 (0.05)	0.420	-0.023 (0.05)	0.672	0.052 (0.03)	0.086	0.060 (0.03)	0.052
12. Weight	1.045 (0.07)	1.052 (0.06)	1.102 (0.06)	0.976 (0.06)	-0.080 (0.05)	0.134	-0.105 (0.05)	0.056	-0.011 (0.03)	0.726	0.002 (0.03)	0.958
15. Concentration	1.729 (0.06)	1.362 (0.04)	1.715 (0.05)	1.254 (0.05)	0.199 (0.05)	<0.001	0.130 (0.05)	0.010	0.137 (0.03)	<0.001	0.125 (0.03)	<0.001
16. View of myself	1.864 (0.08)	1.452 (0.06)	2.024 (0.06)	1.115 (0.07)	0.275 (0.07)	<0.001	0.144 (0.06)	0.022	0.328 (0.03)	<0.001	0.313 (0.04)	<0.001
17. View of my future	1.576 (0.06)	1.217 (0.04)	1.596 (0.05)	1.074 (0.04)	0.184 (0.05)	<0.001	0.106 (0.04)	0.016	0.174 (0.02)	<0.001	0.167 (0.02)	<0.001
18. Death or suicide	0.921 (0.06)	0.727 (0.05)	0.970 (0.05)	0.597 (0.05)	0.171 (0.05)	0.001	0.111 (0.05)	0.037	0.106 (0.03)	<0.001	0.100 (0.03)	<0.001
19. General interest	1.319 (0.07)	1.118 (0.05)	1.366 (0.06)	0.980 (0.06)	0.098 (0.05)	0.047	0.040 (0.05)	0.414	0.016 (0.03)	0.557	0.017 (0.03)	0.545
20. Energy level	1.804 (0.04)	1.542 (0.04)	1.771 (0.05)	1.490 (0.05)	0.170 (0.05)	<0.001	0.133 (0.05)	0.009	0.005 (0.03)	0.861	0.001 (0.03)	0.981
21. Capacity for pleasure	1.342)0.06)	1.055 (0.04)	1.329 (0.05)	0.961 (0.05)	0.227 (0.05)	<0.001	0.175 (0.05)	<0.001	0.028 (0.03)	0.302	0.016 (0.03)	0.556
22. Interest in sex	1.457 (0.08)	1.182 (0.05)	1.427 (0.06)	1.116 (0.06)	0.243 (0.06)	<0.001	0.194 (0.06)	0.002	0.034 (0.03)	0.330	0.017 (0.04)	0.627
23. Psychomotor retardation	1.275 (0.07)	0.897 (0.05)	1.177 (0.06)	0.798 (0.06)	0.221 (0.05)	<0.001	0.178 (0.05)	0.001	0.016 (0.03)	0.608	0.000 (0.03)	0.992
24. Psychomotor agitation	1.275 (0.07)	1.151 (0.05)	1.328 (0.05)	1.047 (0.06)	0.070 (0.06)	0.248	0.004 (0.06)	0.953	0.118 (0.03)	<0.001	0.119 (0.03)	<0.001
25. Aches and pains	1.515 (0.05)	1.313 (0.05)	1.391 (0.05)	1.371 (0.05)	0.012 (0.05)	0.824	-0.004 (0.05)	0.053	-0.062 (0.03)	0.039	-0.053 (0.03)	0.078
26. Sympathetic arousal	1.167 (0.05)	1.081 (0.04)	1.201 (0.04)	1.008 (0.05)	0.066 (0.05)	0.192	0.029 (0.05)	0.569	0.001 (0.03)	0.965	0.003 (0.03)	0.929
27. Panic/Phobic	1.078 (0.07)	0.945 (0.05)	1.168 (0.05)	0.794 (0.06)	0.143 (0.06)	0.020	0.088 (0.06)	0.062	0.097 (0.03)	0.006	0.090 (0.04)	0.011
28. Constipation/diarrhea	0.941 (0.06)	0.891 (0.05)	0.946 (0.05)	0.868 (0.06)	0.013 (0.06)	0.830	-0.015 (0.06)	0.812	-0.056 (0.03)	0.107	-0.052 (0.04)	0.140
29. Interpersonal sensitivity	1.809 (0.07)	1.342 (0.06)	1.963 (0.06)	1.008 (0.06)	0.261 (0.06)	<0.001	0.172 (0.05)	0.001	0.249 (0.03)	<0.001	0.233 (0.03)	<0.001
30. Leaden paralysis	2.132 (0.06)	1.744 (0.05)	2.020 (0.06)	1.727 (0.07)	0.244 (0.05)	<0.001	0.201 (0.05)	<0.001	0.032 (0.03)	0.254	0.022 (0.03)	0.440

Note. Mean values at baseline, standard deviation (in parentheses) for patients with a history of chronic depression (24 out of 48 months before baseline), high neuroticism levels (above median; >43) and their counterparts. Standardized difference in symptom severity (IDS-SR item scores) represent the beta-coefficients of chronic depression at baseline (dichotomous) and level of neuroticism (continuous z-score) assessed with a mixed model with repeated measures with standardized IDS-SR item-score as outcome variable over the course of 9 years follow-up assessed at 6 time-points. Crude SDS is adjusted for age, gender and baseline depression severity. Adjusted SDS is adjusted for age, gender, baseline depression severity and for either chronicity or neuroticism.

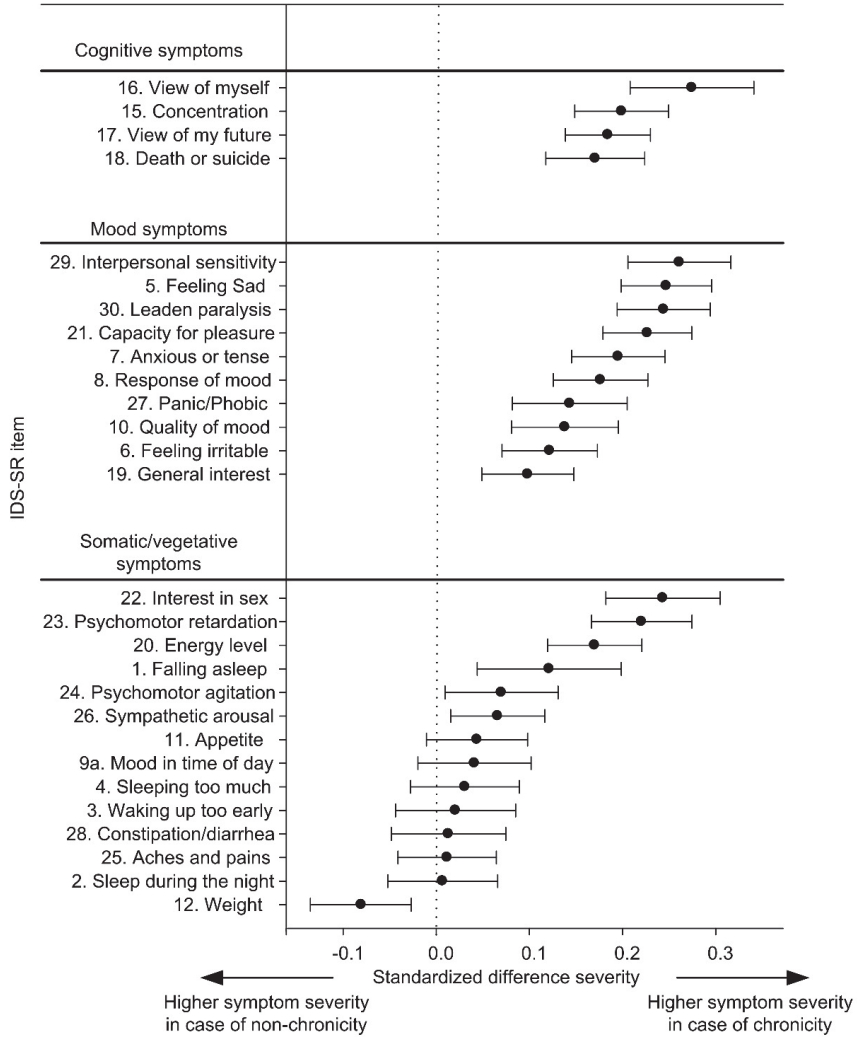


Figure 1. Standardized difference in symptom severity (IDS-SR item scores) according to a history of chronicity (depressed 24 of 48 months before baseline) during the 9-year follow-up. Assessed with linear mixed models with repeated measures and adjusted for gender, age, and baseline MDD severity.

3.3.3 Neuroticism

Patients with higher levels (> 43 score) of neuroticism had lower education levels, were more likely to be chronically depressed, were more likely to use antidepressants, and had higher baseline depression severity than patients with lower levels of neuroticism (Table 1).

High neuroticism was independently associated with a higher severity of depression over the course of 9 years (i.e., IDS scores, adjusted for age, gender, and baseline severity; SDS = 0.071, t = 10.509, p < .001). This translated into a 0.071 *SD* higher average score for each of the 28 IDS items. In general, symptoms in relation to neuroticism mostly followed parallel course trajectories. However, although interaction terms between time and neuroticism were significant, the effect sizes were very small and deemed not clinically important (interaction = 0.014, t = 5.35, p < .001). These parallel course trajectories per item are shown in Figure 2 of the supplementary material. When we adjusted the effect of neuroticism for baseline chronicity, this main effect remained significant (SDS = 0.066, t = 9.766, p < .001).

Next, we compared the effects of neuroticism on the 9-year course across the 28 IDS items (Table 2 and Figure 2). Important differences were found, i.e. the SDS ranged from –0.062 ("aches and pains") through 0.328 ("view of myself"). Most of the individual symptoms were on average more severe in patients with high levels of neuroticism compared to those with lower levels of neuroticism. This was not the case for the five items that were negatively associated with neuroticism (i.e., Items 4, 2, 12, 25, and 28), but only "aches and pains" was significantly different from 0 (Item 25; SDS = –0.062, t = –2.070, p = .039). This indicated that high levels of neuroticism were associated with fewer aches and pains. Neuroticism was strongly related to mood and cognitive symptoms and (to a much lesser extent) to somatic/vegetative symptoms. Particularly patients with high levels of neuroticism were likely to experience low self-esteem and high interpersonal sensitivity. When using the Bonferroni-corrected critical level of significance of .002, significant associations were found with four of four cognitive symptoms, three of 10 mood symptoms, and four of 14 somatic/vegetative symptoms. When we adjusted the associations between neuroticism and individual symptoms for chronicity at baseline, the outcomes again remained largely significant for most symptoms, except for Item 25, "aches and pains" (Table 2).

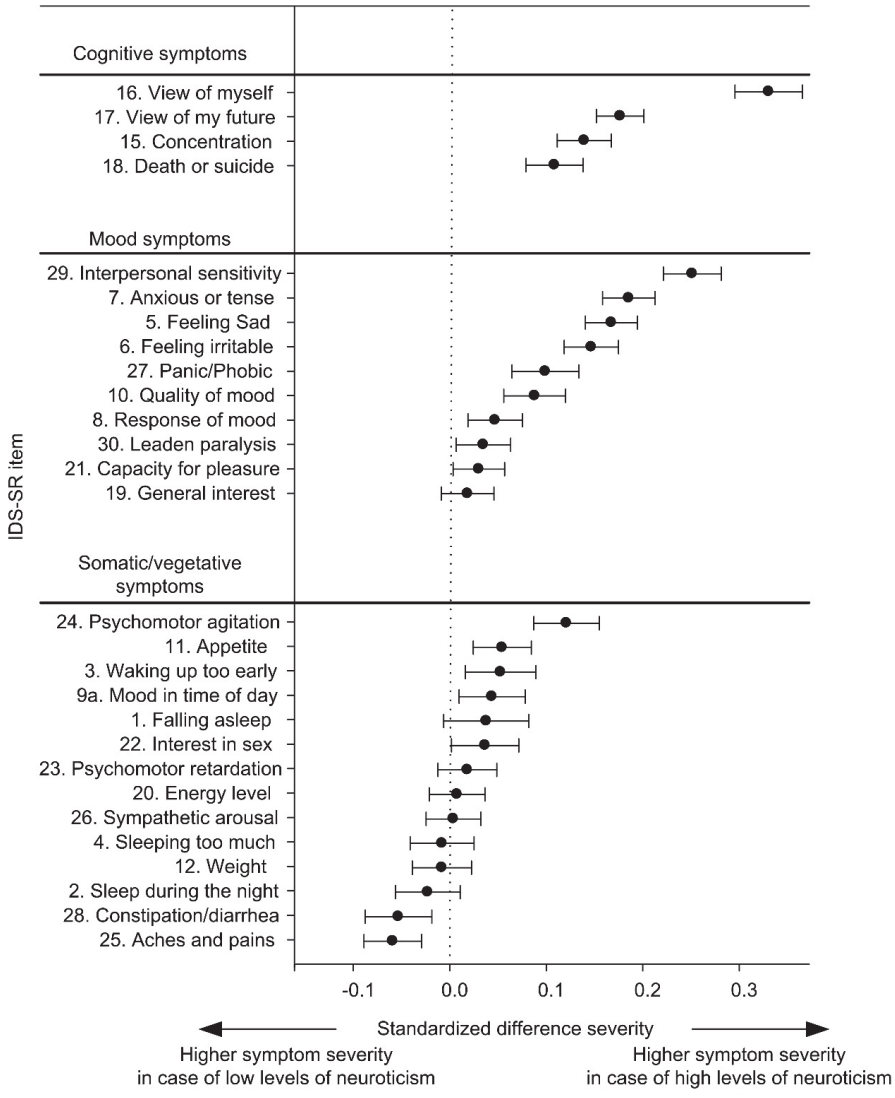


Figure 2. Standardized difference in symptom severity (IDS-SR item scores) according to a 1-*SD* increase in neuroticism at baseline during the 9-year follow-up. Assessed with linear mixed models with repeated measures and adjusted for gender, age, and baseline MDD severity.

3.4 Discussion

The present study found that a history of chronic depression and of neuroticism at baseline was a predictor for the severity of most individual symptoms during 9 years of follow-up, albeit of varying magnitudes. Although the improvements showed parallel trajectories over time, according to chronicity and neuroticism, IDS-SR scores remained higher for patients who initially had chronic depression or higher levels of neuroticism. Findings for the effects of chronicity and neuroticism were remarkably similar, even though only 21.5% of our sample had both a history of chronic depression and a neuroticism score above the median. Although the effects on five of 29 symptoms were no longer significant when adjusting the effects of chronicity for neuroticism and vice versa, both baseline variables independently predicted an adverse course of symptoms of the mood and cognitive symptom clusters, whereas the effects on the somatic/vegetative symptoms were smaller. Chronicity and neuroticism showed the strongest link to 'low self-esteem' (Item 16) and 'interpersonal sensitivity' (Item 29).

Epidemiological research on MDD generally focuses on MDD as a unified syndrome, using a questionnaire sum score as a measure for the level of severity. If depression is truly one unified latent construct, all risk factors would have affected the individual symptoms with similar effect sizes. However, previous cross-sectional studies also found that, at baseline, risk factors such as neuroticism and chronicity (among others) are associated with different individual depressive symptoms [36, 38]. We extended these findings by using a prospective design, which helped to show that the history of chronic depression and neuroticism affects the level of mood and cognitive symptoms, but not somatic/vegetative symptoms. This provides additional epidemiological support for the heterogeneity of individual depressive symptoms [36].

In predictive research, focusing on individual symptoms instead of syndromes may yield important new findings. In this regard, specific emphasis should be given to the strong relationship we found between both chronic depression and neuroticism, and self-esteem and interpersonal sensitivity. The similar results for chronicity and neuroticism in relation to these two symptoms seem to suggest that either these symptoms are core features of MDD or that a third dimension (e.g., general severity of MDD, chronic arousal and stress activation,

or social isolation) underlies the reported relationships or both. Although no longer in practice since the introduction of the DSM-III, our findings are relevant in light of a proposition to revive *neurotic depression*, a subtype of depression which is reactive to life events, persistent, and unlikely to benefit from antidepressants (Nassir Ghaemi, 2008). Our findings concerning low self-esteem and high interpersonal sensitivity may also indicate a possible comorbid avoidant personality disorder (i.e., preoccupation with being criticized or rejected in social situations and feeling socially inept) and dependent personality disorder (i.e., feeling inadequate to take care of oneself and seeking excessive support). Higher rates of Cluster C personality disorders have been reported in chronic versus nonchronic depression (Baldessarini et al., 2017; Russell et al., 2003). Moreover, patients may not meet the criteria for personality disorders after their depressive symptoms are in remission, suggesting an overlap in symptomatology and etiology (Costa et al., 2005; Fava et al., 2002). Low self-esteem and high levels of interpersonal sensitivity can play a role in the overall persistence and relapse of depression [46-49]. Also, a causal relationship may exist between the symptoms [50-52], and targeting key symptoms (i.e., symptoms more central in a causal network) may benefit a patient's recovery. Low self-esteem and interpersonal sensitivity could be key symptoms in patients with chronic depression and high levels of neuroticism.

Multiple evidence-based treatments are available for low self-esteem, such as Competitive Memory Training [COMET; 53, 54] and mindfulness-based cognitive behavioral therapy [55, 56]. Interpersonal sensitivity is an important treatment target in interpersonal therapy [57]. More research is needed to assess if these, or other treatments, could be implemented as symptom-specific treatment methods, and whether a symptom-specific treatment approach is indeed beneficial for the patient.

This study has several strengths. First, the heterogeneous nature of depression was examined in a substantial number of MDD patients by analyzing depression at symptom level over a follow-up period of 9 years. Moreover, the analyzes were adjusted for multiple covariates, including baseline severity [baseline IDS sum-score; 7]. Nevertheless, some limitations also need addressing. First, because NESDA is an observational cohort study, several variables may have confounded our findings. Some patients were exposed to different types of treatments, such as psycho- and pharmacotherapy. For example, patients with chronic depression or higher levels of neuroticism were more often treated with antidepressants than their

counterparts. Certain symptoms, such as reduced libido, can stem from medication side-effects rather than from depression as such and, as a result, may be more prevalent among chronically depressed patients than among nonchronically depressed patients (Baldwin, 2003; Rosse et al., 2007). However, most SDS values and the order of the symptom SDS did not change substantially after adjusting for the received treatment (results available upon request). Second, although chronic depression and neuroticism were interrelated and showed associations with similar symptom profiles, they were also (in part) independent constructs, since chronically ill patients had a neuroticism score that was only (mean) 3.02 (SE 0.57) points higher than that of non-chronically ill patients. More research is needed to unravel the underlying mechanisms that link chronic depression and neuroticism. Third, our definition of chronic depression (i.e., being depressed for ≥ 24 months during the last 48 months) differs from that used in other studies [e.g., 12]. Moreover, our chronic patients may not have experienced symptoms for ≥ 2 months over the course of 2 years and, thus, did not meet the criteria for persistent depressive disorder (according to the DSM-5). Fourth, assessing individual symptoms based on single items presents psychometric difficulties, because single items are more strongly affected by random error than the sum scores of items. However, there are also arguments in favor of single items, especially concerning practical use (Diamantopoulos et al., 2012). Finally, both individual symptoms and level of neuroticism were measured using self-report measures. Self-report measures require patients to possess a certain level of insight, which may be lacking when levels of psychopathology are high. Although the interclass correlation of the neuroticism score over three time points was of moderate strength (ICC = 0.672; F = 3.05; p = <.001), an earlier study using NESDA data reported that levels of neuroticism were affected by a patient's current depressive state (Spinhoven et al., 2013). As neuroticism is partly state dependent, our findings are limited by the fact that we could only use a single baseline assessment of neuroticism, which is likely to have been affected by the burden of psychiatric disease. However, it has been suggested that disorder-related state effects may reflect the true nature of personality (Riso et al., 2002; Spinhoven et al., 2013). Personality characteristics may change when depressive episodes remit, for example, due to a shared underlying etiology (see Costa et al., 2005). Future research could focus on comparison of state and trait effects of neuroticism on the course of depression and its individual symptoms, with trait being inferred from the mean neuroticism across several preceding measurements.

In conclusion, this study shows that a history of chronic depression and neuroticism predicted a higher severity of mood and cognitive symptoms and, to a lesser extent, severity of somatic/vegetative symptoms over the entire 9-year follow-up. Chronicity and high neuroticism may signal a specific disease cluster, since both variables are related to similar depression symptoms. In this context, future research might explore whether psychotherapeutic treatments that focus on low self-esteem or interpersonal sensitivity yield better outcomes for individual patients with high neuroticism and/or chronicity. It would be useful to examine whether such personalized interventions lead to better outcomes compared to standardized treatment protocols that approach MDD as a homogenous syndrome for all patients.

Author Statement

Contributors

BP is principal investigator of NESDA. Author WE performed the statistical analyzes and wrote the manuscript. EG contributed by frequent supervision and revision of the statistical-analyzes and writing of this manuscript. AH, IC, BP, and PS contributed by several revisions of the early stages as well as final stages of the manuscript. All authors have approved the final manuscript.

Role of the funding source

The funding source had no role in the design of this study, it's execution, analyzes, interpretation of the data, or decision to submit results.

Acknowledgements

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (VU University Medical Center, GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum).

Conflicts of interest

None.

References

1. Fried, Nesse, and Randolph, *Depression sum-scores don't add up: why analyzing specific depression symptoms is essential.* BMC medicine, 2015. **13**(1): p. 72.
2. Conradi, H., J. Ormel, and P. de Jonge, *Symptom profiles of the DSM-IV-defined remission, recoverym relapse, and recurrence of depression: the role of the core symptoms.* Depression and Anxiety, 2012. **29**(7): p. 638-645.
3. Colman, I., et al., *Predictors of long-term prognosis of depression.* Canadian Medical Association Journal, 2011. **183**(17): p. 1969-1976.
4. Kelly, K.M. and B. Mezuk, *Predictors of remission from generalized anxiety disorder and major depressive disorder.* J Affect Disord, 2017. **208**: p. 467-474.
5. Kotov, R., et al., *Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis*. 2010, American Psychological Association.
6. Boschloo, L., et al., *The Four-Year Course of Major Depressive Disorder: The Role of Staging and Risk Factor Determination.* Psychotherapy and Psychosomatics, 2014. **83**(5): p. 279-288.
7. Mulder, R.T., *Personality pathology and treatment outcome in major depression: a review.* American Journal of Psychiatry, 2002. **159**(3): p. 359-371.
8. Spinhoven, P., et al., *Positive and negative life events and personality traits in predicting course of depression and anxiety.* Acta Psychiatrica Scandinavica, 2011. **124**(6): p. 462-473.
9. Coryell, W., J. Endicott, and M. Keller, *Outcome of patients with chronic affective disorder: a five-year follow-up.* The American journal of psychiatry, 1990. **147**(12): p. 1627.
10. Mynors-Wallis, L. and D. Gath, *Predictors of treatment outcome for major depression in primary care.* Psychological medicine, 1997. **27**(3): p. 731-736.
11. Stegenga, B.T., et al., *The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study.* Social psychiatry and psychiatric epidemiology, 2012. **47**(1): p. 87-95.
12. Trivedi, M.H., et al., *Evaluation of outcomes with citalopram for depression using measurement-based care in STAR* D: implications for clinical practice.* American journal of Psychiatry, 2006. **163**(1): p. 28-40.
13. Association, A.P., *Diagnostic and statistical manual of mental disorders (DSM-5®)*. 2013: American Psychiatric Pub.
14. Jorm, A.F., et al., *Predicting anxiety and depression from personality: Is there a synergistic effect of neuroticism and extraversion?* Journal of abnormal psychology, 2000. **109**(1): p. 145.
15. Saklofske, D., I. Kelly, and B. Janzen, *Neuroticism, depression, and depression proneness.* Personality and individual differences, 1995. **18**(1): p. 27-31.
16. Jylhä, P. and E. Isometsä, *The relationship of neuroticism and extraversion to symptoms of anxiety and depression in the general population.* Depression and anxiety, 2006. **23**(5): p. 281-289.
17. Schmitz, N., J. Kugler, and J. Rollnik, *On the relation between neuroticism, self-esteem, and depression: results from the National Comorbidity Survey.* Comprehensive psychiatry, 2003. **44**(3): p. 169-176.
18. Duggan, C., et al., *Neuroticism: a vulnerability marker for depression evidence from a family study.* J Affect Disord, 1995. **35**(3): p. 139-43.
19. Cloninger, C.R., D.M. Svrakic, and T.R. Przybeck, *Can personality assessment predict future depression? A twelve-month follow-up of 631 subjects.* Journal of affective disorders, 2006. **92**(1): p. 35-44.
20. Shea, M.T. and S. Yen, *Personality traits/disorders and depression: A summary of conceptual and empirical findings.* 2005.
21. Malouff, J.M., E.B. Thorsteinsson, and N.S. Schutte, *The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis.* Journal of Psychopathology and Behavioral Assessment, 2005. **27**(2): p. 101-114.
22. Riso, L.P., R.K. Miyatake, and M.E. Thase, *The search for determinants of chronic depression: a review of six factors.* Journal of affective disorders, 2002. **70**(2): p. 103-115.
23. Schrader, G., *Chronic Depression: State or Trait?* The Journal of nervous and mental disease, 1994. **182**(10): p. 552-555.
24. Hartlage, S., K. Arduino, and L.B. Alloy, *Depressive personality characteristics: State dependent concomitants of depressive disorder and traits independent of current depression.* Journal of Abnormal Psychology, 1998. **107**(2): p. 349.

25. Steyer, R., M. Schmitt, and M. Eid, *Latent state–trait theory and research in personality and individual differences.* European Journal of Personality, 1999. **13**(5): p. 389-408.
26. Quilty, L.C., et al., *Dimensional personality traits and treatment outcome in patients with major depressive disorder.* Journal of Affective Disorders, 2008. **108**(3): p. 241-250.
27. Klein, D.N., et al., *Primary early-onset dysthymia: Comparison with primary nonbipolar nonchronic major depression on demographic, clinical, familial, personality, and socioenvironmental characteristics and short-term outcome.* Journal of Abnormal Psychology, 1988. **97**(4): p. 387.
28. Klein, D.N., et al., *Family study of early-onset dysthymia: Mood and personality disorders in relatives of outpatients with dysthymia and episodic major depression and normal controls.* Archives of General Psychiatry, 1995. **52**(6): p. 487-496.
29. Wiersma, J.E., et al., *The importance of childhood trauma and childhood life events for chronicity of depression in adults.* Journal of Clinical Psychiatry, 2009. **70**(7): p. 983.
30. Russell, J.M., et al., *Chronic depression and comorbid personality disorders: response to sertraline versus imipramine.* The Journal of clinical psychiatry, 2003. **64**(5): p. 554-561.
31. Trivedi, M.H., et al., *The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation.* Psychological Medicine, 2004. **34**(1): p. 73-82.
32. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): Psychometric properties.* Psychological Medicine, 1996. **26**(3): p. 477-486.
33. Beck, A.T., R.A. Steer, and M.G. Carbin, *Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation.* Clinical psychology review, 1988. **8**(1): p. 77-100.
34. Keller, M.C., M.C. Neale, and K.S. Kendler, *Association of different adverse life events with distinct patterns of depressive symptoms.* Am J Psychiatry, 2007. **164**(10): p. 1521-1529.
35. Fried, E.I., et al., *The differential influence of life stress on individual symptoms of depression.* Acta Psychiatrica Scandinavica, 2015. **131**(6): p. 465-471.
36. Fried, et al., *Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors.* Psychol. Med., 2014. **44**(10): p. 2067-2076.
37. Kendler, K.S. and S.H. Aggen, *Symptoms of major depression: Their stability, familiality, and prediction by genetic, temperamental, and childhood environmental risk factors.* Depress Anxiety, 2017. **34**(2): p. 171-177.
38. Lux, V. and K.S. Kendler, *Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria.* Psychol Med, 2010. **40**(10): p. 1679-90.
39. Angst, J., et al., *Long-term depression versus episodic major depression: results from the prospective Zurich study of a community sample.* Journal of affective disorders, 2009. **115**(1): p. 112-121.
40. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* International Journal of Methods in Psychiatric Research, 2008. **17**(3): p. 121-40.
41. Lyketsos, C.G., et al., *The Life Chart Interview: a standardized method to describe the course of psychopathology.* International Journal of Methods in Psychiatric Research, 1994.
42. Murray, G., et al., *NEO Five-Factor Inventory scores: Psychometric properties in a community sample.* Measurement and Evaluation in Counseling and Development, 2003. **36**(3): p. 140-150.
43. Costa, P.T. and R.R. McCrae, *Normal personality assessment in clinical practice: The NEO Personality Inventory.* Psychological assessment, 1992. **4**(1): p. 5.
44. McCrae, R.R. and P.T. Costa, *A contemplated revision of the NEO Five-Factor Inventory.* Personality and individual differences, 2004. **36**(3): p. 587-596.
45. Schaakxs, R., et al., *Age-related variability in the presentation of symptoms of major depressive disorder.* Psychological Medicine, 2017. **47**(3): p. 543-552.
46. Davidson, J., et al., *Symptoms of interpersonal sensitivity in depression.* Comprehensive Psychiatry, 1989. **30**(5): p. 357-368.
47. Sowislo, J.F. and U. Orth, *Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies*. 2013, American Psychological Association.
48. Verhoeven, F.E., et al., *Seeing the signs: Using the course of residual depressive symptomatology to predict patterns of relapse and recurrence of major depressive disorder.* Depression and anxiety, 2018. **35**(2): p. 148-159.
49. De Rubeis, J., et al., *Rejection sensitivity as a vulnerability marker for depressive symptom deterioration in men.* PloS one, 2017. **12**(10): p. e0185802.

50. Fried, E.I., et al., *What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis.* Journal of Affective Disorders, 2016. **189**: p. 314-320.
51. Beard, C., et al., *Network analysis of depression and anxiety symptom relationships in a psychiatric sample.* Psychological Medicine, 2016. **46**(16): p. 3359-3369.
52. Borkulo, C.v., et al., *Association of symptom network structure with the course of longitudinal depression.(Report).* JAMA Psychiatry, 2015. **72**(12): p. 1219.
53. Korrelboom, K., M. Maarsingh, and I. Huijbrechts, *Competitive memory training (COMET) for treating low self-esteem in patients with depressive disorders: a randomized clinical trail.* Depression and anxiety, 2012. **29**(2): p. 102-110.
54. Korrelboom, K., M. Marissen, and T. van Assendelft, *Competitive memory training (COMET) for low self-esteem in patients with personality disorders: A randomized effectiveness study.* Behavioural and Cognitive Psychotherapy, 2011. **39**(1): p. 1-19.
55. Fjorback, L.O., et al., *Mindfulness-Based Stress Reduction and Mindfulness-Based Cognitive Therapy–a systematic review of randomized controlled trials.* Acta Psychiatrica Scandinavica, 2011. **124**(2): p. 102-119.
56. Thompson, B.L. and J.A. Waltz, *Mindfulness, self-esteem, and unconditional self-acceptance.* Journal of Rational-Emotive & Cognitive-Behavior Therapy, 2008. **26**(2): p. 119-126.
57. Klerman, G.L. and M.M. Weissman, *Interpersonal psychotherapy of depression: A brief, focused, specific strategy*. 1994: Jason Aronson, Incorporated.

Supplementary material

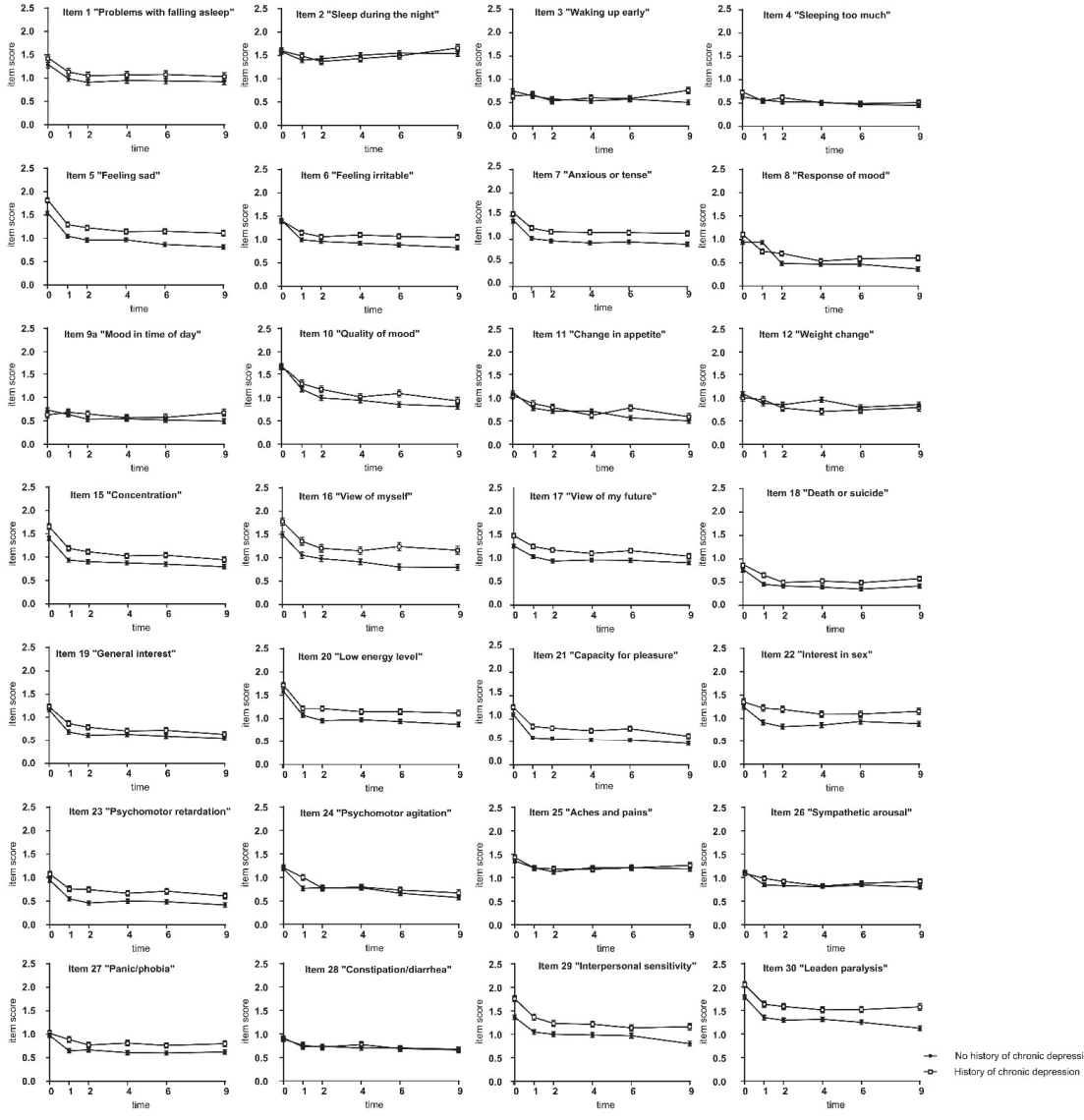


Figure 1 supplementary material. Estimated mean values of individual symptom scores over the 9-year follow-up in 560 MDD patients according to a history of chronicity (depressed 24 of 48 months depressed before baseline).

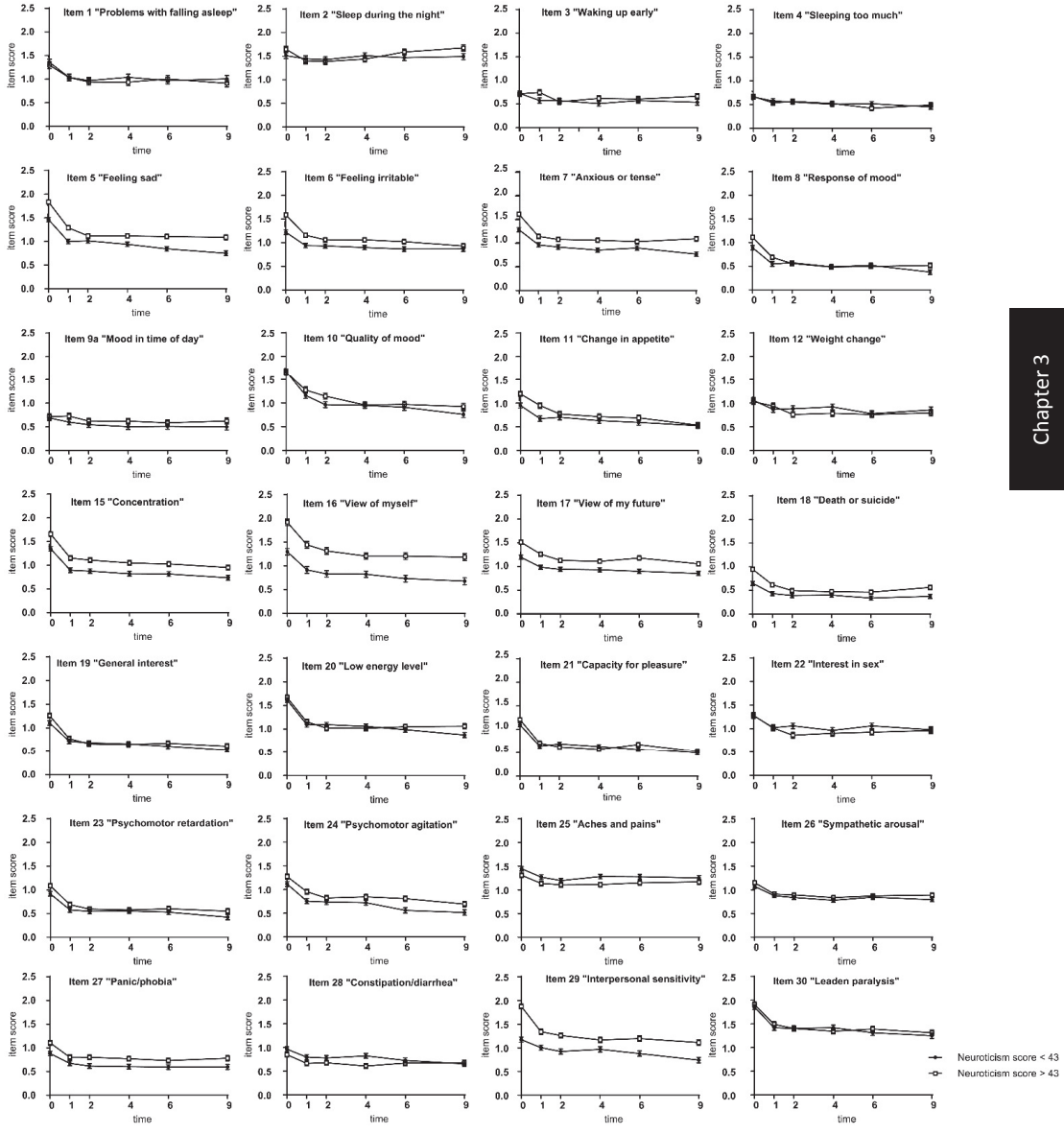


Figure 2 supplementary material. Estimated mean values of individual symptom scores over the 9-year follow-up in 560 MDD patients according to median split neuroticism score at baseline

[1] Department of Psychiatry, Leiden University Medical Centre (LUMC), Albinusdreef 2, Leiden, Zuid Holland, The Netherlands
[2] Clinical Psychology Unit, Institute of Psychology, Leiden University, Wassenaarseweg 52, Leiden, Zuid Holland, The Netherlands

Chapter 4

Prognostic Value of Pathological Personality Traits for Treatment Outcome in Anxiety and Depressive Disorders: The Leiden Routine Outcome Monitoring Study

van Eeden, W.A.[1], van Hemert, A.M.[1], Giltay, E.J.[1], Spinhoven, P.[1,2], de Beurs[2], E., Carlier, I.V.E[1]

(2022). The journal of nervous and mental disease, 10-1097

Abstract

Previous studies have failed to take baseline severity into account when assessing the effects of pathological personality traits (PPT) on treatment outcome. This study assessed the prognostic value of PPT (Dimensional Assessment of Personality Pathology-Short Form, DAPP-SF) on treatment outcome (Brief Symptom Inventory, BSI-posttreatment) among patients with depressive and/or anxiety disorders (*N* = 5,689). Baseline symptom level (BSI-pretreatment) was taken into account as a mediator- or moderator variable. Results showed significant effects of PPT on outcome, of which Emotional Dysregulation demonstrated the largest association, β=0.43, *p*<.001. When including baseline BSI score as a mediator variable, a direct effect (β=0.11; *p*<.001) remained of approximately one-third of the total effect. The effects of Emotional Dysregulation (interaction-effect β=0.061, *p*<.001) and Inhibition (interaction-effect β=0.062, *p*<.001), but not Compulsivity or Dissocial Behavior, were moderated by the baseline symptom level. PPT predicts higher symptom levels, both before and after treatment, but yields relatively small direct effects on symptom decline when the effect of pretreatment severity is taken into account.

Keywords: pathological personality traits, depression, anxiety disorders, treatment outcome, Dimensional Assessment of Personality Pathology Short Form (DAPP-SF)

4.1 Introduction

Pathological personality has often been linked to other psychiatric disorders, such as depressive and anxiety disorders [1-3]. Pathological personality can be considered from a categorical as well as a dimensional perspective. From a categorical perspective, personality pathology is assumed to be present when a patient meets the criteria for a personality disorder according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) or according to the Classification of Mental and Behavioral Disorders, the tenth revision [4-6]. Meta-analysis demonstrated that the risk of comorbid personality disorders for major depressive disorder has been estimated at 45% [7]; the risk ranged from 35% to 52% for anxiety disorders [3]. Moreover, in multiple reviews and meta-analyses researchers assessed the associations between personality disorders and treatment outcome of depressive and anxiety disorders [8-14]. It was found that the odds for poor outcome more than doubled when a comorbid personality disorder was present [13]. Evidence regarding anxiety disorders was less conclusive; some researchers found significant negative effects of personality disorders comorbidity [9, 11], but others did not [10, 11]. In one meta-analysis, Olatunji et al. (2010) found no significant effect of comorbid personality disorders on treatment outcome among patients with anxiety disorder.

There is clear empirical evidence that personality disorders are in fact better represented by a dimensional model than by the categorical model [15], in which personality pathology exists on a continuum, ranging from healthy/normal to maladaptive/abnormal psychopathology [16]. Several alternative dimensional approaches for personality disorders are proposed [see for an overview: 17]. A major effort has been made in this regard by Livesley and colleagues, who reorganized lower-order traits described among 100 self-report scales into 18 factors [18, 19]. These 18 factors formed the basis for the development of a self-report scale – The Dimensional Assessment of Personality Pathology [DAPP; 20]. Beside differences in methodology, subsequent studies found a considerable overlap with other models, such as with the five factor model [21]. The DAPP also demonstrated a considerable overlap in pathological personality traits (PPT) with other relevant scales such as the NEO Personality Inventory [NEO-PI; 21], Personality Inventory for DSM-5 [PID-5; 22, 23]), Schedule for Nonadaptive and Adaptive Personality [SNAP; 24], and Severity Indices of Personality

Functioning [SIPP; 25, 26]. Moreover, the identified pathological personality traits (PPT) are often used as a proxy measure of the Alternative DSM-5 model of personality disorders B-criterium personality traits [27].

Within the Leiden Routine Outcome Monitoring Study, it was demonstrated that patients with combined depressive and anxiety disorders displayed the highest mean values of PPT measured with the Dimensional Assessment of Personality Pathology - Short Form (DAPP-SF), followed by patients with singular depressive disorders. Mean values of PPT were lowest for patients with singular anxiety disorders [28]. Van Noorden et al. (2012) and Schat, van Noorden [29] found that PPT predicted an unfavorable treatment outcome (50% reduction of measured psychological distress) in patients with mood-, anxiety-, and somatoform disorders, with a hazard ratio ranging from 0.92 (95% confidence interval; CI [0.81, 1.05]) to 1.30 [95% CI 1.12–1.51; 30]. The present study builds upon this existing work with an extension of the sample, by using continuous outcome measures, and by explicitly taking the effects of baseline symptom level into account.

The effects of PPT on treatment outcome may be substantially lower when taking baseline symptom level into account, usually interpreted as severity. Baseline symptom level of depression and anxiety consistently influences posttest outcomes for depressive and anxiety disorders [8, 10]. The effect of PPT on treatment outcome or disorder persistence is attenuated when baseline symptom level is taken into account [8, 31, 32]. For instance, the effects of neuroticism on the persistence of a depressive disorder over the course of two years decreased from 1.57 RR, 95% CI [1.35, 1.83] to 1.20 RR, [0.92, 1.57], and on the persistence of an anxiety disorder from 1.67, [1.42, 1.95] to 1.09, [0.87, 1.36], after adjusting for baseline symptom level [32]. Adjusting the relationship between PPT and treatment outcome for baseline severity may be too simplistic. After all, patients with high levels of PPT may report higher levels of depression and anxiety. Baseline severity may serve as a mediating factor between PPT and treatment outcome [33]. Candrian et al. (2007) investigated this and found that the effect of personality disorder on an 8 week open-label treatment of fluoxetine was fully mediated by baseline symptom level. Moreover, previous studies found differential clinical characteristics of high and low severe depression and anxiety [34-37]. Baseline symptom severity could be an important moderator of treatment outcome as is demonstrated for patients suffering substance use disorders [38] and borderline personality

disorder [39, 40]. Possibly PPT may be especially predictive for treatment outcome in patients suffering from higher baseline symptom levels. PPT may hamper coping with high disease severity of depression and anxiety [41], in which case baseline severity could be a moderator variable of the effect of PPT on treatment outcome. Surprisingly, the likely intermediary effects (either as a mediator variable, or a moderator variable) of baseline severity on the relationship between PPT and treatment outcome have received little attention in the current literature [38-40].

Our aim was to investigate the prognostic value of dimensional PPT on treatment outcome among patients with anxiety disorders and/or depression while taking the effects of baseline symptom level into account. We first assessed the association between PPT and treatment outcome. Thereafter, we assessed how this possible association was affected by baseline severity. We assessed both the potential of mediation and moderation of baseline symptom level in the relationship between PPT and treatment outcome. The mediation analysis gave us insight into the role of baseline severity within the relationship between PPT and treatment outcome. Moderation analysis gave us insight into whether the effects of PPT on treatment outcome were different for patients with high baseline severity compared to low baseline severity. We used the Dimensional Assessment of Personality Pathology - Short Form (DAPP-SF) to measure a wide variety of maladaptive personality traits [42]. Based on previous research [8, 30, 31, 43], we hypothesized that PPT would be associated with higher symptom levels, both at baseline and after treatment. To assess the potential differential effects of PPT for depression, anxiety, and combined depression/anxiety [28], we performed additional analyses for each diagnostic group separately.

4.2 Methods

4.2.1 Participants

In this study, we used data from a sample of 5,755 psychiatric outpatients who received treatment for anxiety- and/or mood disorders at the mental health care provider GGZ Rivierduinen or at the Department of Psychiatry of the Leiden University Medical Centre (LUMC), both located in the Netherlands. We included adult patients (18 years or older) with anxiety disorders and/or depressive disorders of whom data was collected as part of the Leiden Routine Outcome Monitoring Study (2004–2013), and who had completed both the DAPP-SF at baseline and the Brief Symptom Inventory (BSI) at baseline and at 6 to 8 months posttreatment (see Instruments). Patients were recruited in policlinic departments for mood- and/or anxiety disorder. When patients had other primary diagnoses they were referred to other departments and therefore not included in the present study. As data collection in the form of Routing Outcome Monitoring is part of the routine care, this resulted in a representable sample of outpatients with anxiety disorders and/or depressive disorders.

4.2.2 Design and Procedure

Routine outcome monitoring (ROM) data were derived from a prospective cohort study, which was carried out to assess treatment outcome for patients with mood-, anxiety-, and/or somatoform disorders in a naturalistic setting [44]. For our analyses, we used data from assessments collected at the start of treatment and after 6 to 8 months of treatment. The first assessment occurred during an intake procedure; in order to diagnose patients in a standardized and reliable method; research nurses interviewed patients using the Mini International Neuropsychiatric Interview-Plus [MINI-Plus; 45]. Additionally, patients completed a number of self-report questionnaires. For further details regarding our ROM procedure, see de Beurs, den Hollander-Gijsman [44] and Carlier, Andree Wiltens [46]. Patients were treated in accordance with (inter)national evidence-based guidelines, consisting of pharmacotherapy, psychotherapy (e.g., cognitive behavioral therapy or interpersonal therapy), or a combination [e.g., 47, 48].

4.2.3 Instruments

4.2.3.1 Pathological personality traits

The DAPP-SF is a 136-item self-report questionnaire used to assess maladaptive personality traits. Participants rated items on a 5-point scale, ranging from 1 (*very unlike me*) to 5 (*very like me*). The items are clustered into 18 subscales and four higher order constructs. The subscales Submissiveness, Cognitive Distortion, Identity Problems, Affective Lability, Oppositionality, Anxiousness, Suspiciousness, Social Avoidance, Narcissism, Insecure Attachment, and Self-Harm are clustered under *Emotional Dysregulation* as the first higher order construct with 78 items. The subscales Intimacy Problems and Restricted Expression are clustered under *Inhibition* as the second higher order construct with 16 items. The subscales Stimulus Seeking, Callousness, Rejection, and Conduct Problems are clustered under *Dissocial Behavior* as the third higher order construct with 34 items. Finally, the subscale Compulsivity equals the fourth higher order construct *Compulsivity* with 8 items [49].

In accordance with the DAPP-SF manual, subscale scores and higher order construct scores are calculated as the mean of the item scores (see Table 1). Although the DAPP-SF subscales are associated with Cluster A-, B-, and C- Personality Disorders, they can be considered as dimensional scales ranging from “normal” to maladaptive PPT. Psychometric evaluations, both in the community and in clinical samples (i.e., patients with both Axis-I and Axis-II DSM-IV disorders), demonstrated good internal consistency, with Cronbach’s α between 0.78 and 0.89 [42]. The DAPP-SF score ranges from 1–5 and was used in our study as the independent variable (IV), with the higher order constructs serving as primary predictor variables.

4.2.3.2 General Psychopathology

The BSI is a 53-item self-report questionnaire used to assess symptoms of depression, anxiety, somatization, obsessive–compulsivity, interpersonal sensitivity, hostility, phobic anxiety, paranoid ideation, and psychoticism [50]. Participants rate items on a 5-point scale, ranging from 0 (*not at all*) to 4 (*extremely*). A psychometric evaluation of the BSI was performed in a large population of psychiatric patients, and it demonstrated good test–retest reliability and good internal consistency, with Cronbach’s α between 0.71 and 0.84 [51]. The BSI score (total) ranges from 0–4 and was used in our study as a dependent variable (DV) for our statistical analyses.

4.2.4 Statistical Analyses

We took several steps in our analyses to investigate the prognostic value of dimensional levels of PPT and the intermediary effects of baseline symptom level on treatment outcome of patients with anxiety- and depressive disorders. First, we conducted a mediation analysis using Preacher and Hayes [52] mediation model. This procedure allowed us to test the effects of an independent variable (IV; higher order PPT constructs) on BSI posttest (dependent variable; DV), either with or without a mediator (BSI baseline; M). This is demonstrated in Figure 1 A, where the c path denotes the effect of PPT (IV) on treatment outcome (DV) without mediation by baseline symptom levels. Figure 1 B demonstrates the *a* path which denotes the effect of PPT (IV) on BSI (DV) at baseline (M), the *b* path denotes the effect of M on DV, and the *c'* path denotes the direct effect after controlling for the mediator (M) baseline symptom level. Mediation was determined by testing the indirect effect of the IV on the DV via M ($a \times b$). This is quantified as the product of the effect of the IV on M (*a* path) and the effect of M on the DV (*b* path). We used a bootstrapping approach with 5,000 estimates of the $a \times b$ path to estimate the indirect effect. We computed 95% CIs for the empirical distribution, using cutoffs for the 2.5% highest and lowest scores. Mediating effects were considered to be significant when the CI did not include zero. For detailed information about the statistical procedures of the mediation analyses, see Hayes [53] and Loose, Acier [54]. Second, we performed a moderation analysis, in which PPT served as the IV, treatment outcome as the DV, and baseline symptom level as the moderation variable. We assessed whether there was an interaction between PPT and baseline symptom level in relation to treatment outcome. Thereafter, we assessed the effects of PPT for patients with one *SD* lower baseline symptom level and for patients with one *SD* higher baseline symptom level. We repeated these analyses for the 18 underlying DAPP-SF subscales clustered under the four higher order constructs, and we performed additional analyses for each diagnostic group separately (depression, anxiety, or combined depression/anxiety groups) which is included in the Appendix. All outcomes and IVs were standardized (i.e., *Z* scores) to yield standardized beta coefficients that could be compared between measures. Analyses were performed using R, version 3.4.1.

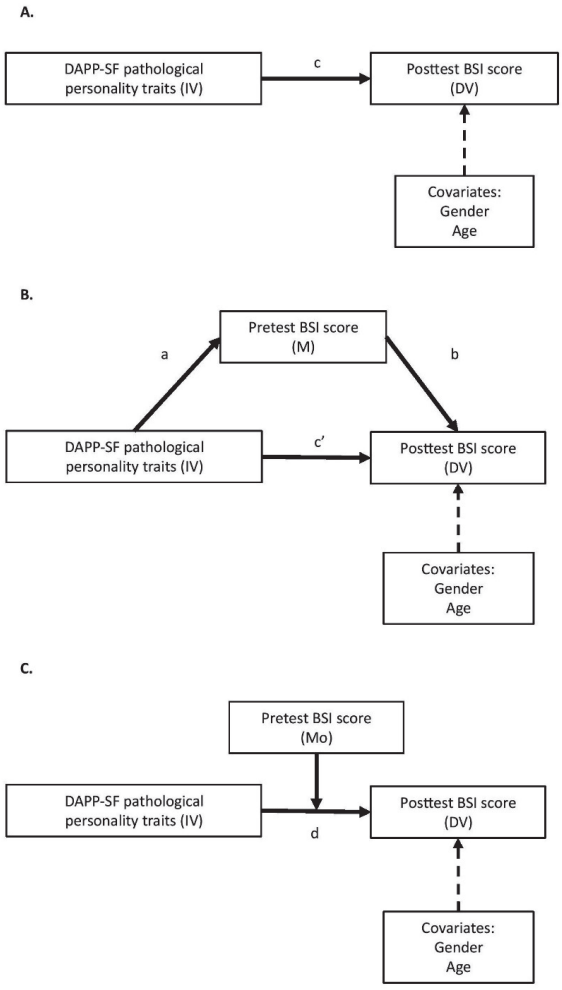


Figure 1. Model of psychopathology (DAPP-SF dimensions), baseline level of symptoms (baseline BSI score), and treatment outcome (posttest BSI score), suggesting that an increased baseline symptom level is an intermediate factor between psychopathology and treatment outcome. "IV" denotes independent variable (DAPP-SF). "DV" denotes dependent variable (posttest BSI score). "M" denotes mediating variable (baseline BSI). "Mo" denotes moderating variable (baseline BSI). "*c*" denotes total effect of IV on DV. "*a*" denotes effect of IV on M. "*b*" denotes effect of M on DV. "*c*" denotes direct effect of IV on DV. "d" denotes the moderated effect of IV on DV.

4.3 Results

4.3.1 Sample Characteristics

Table 1 presents the sample characteristics. On average, patients were 38 years old (*SD* = 12.5), and women (62.8%) were overrepresented compared to men (37.2%). The mean BSI score was 1.33 (*SD* = 0.70) at baseline, and 0.85 (*SD* = 0.72) after 6 to 8 months of treatment. The highest BSI scores were found among the combined depression and anxiety group, *p <0.001* (see Appendix Table 1). The DAPP-SF higher order PPT constructs ranged from 1.90 (Dissocial Behavior) to 2.93 (Compulsivity). The highest levels of PPT were found among the combined subgroup compared to the depression and anxiety subgroups (see Appendix Table 1).

Table 1

Demographic and Clinical Sample Characteristics at Baseline.

	Total sample (*n* = 5,689)
Variable	**Mean (*SD*) *n* (%)**
Age	38.8 (12.5)
Gender (female)	3572 (62.8)
BSI baseline score	1.33 (0.70)
BSI posttreatment score	0.85 (0.72)
MDD – single episode	1451 (25.5)
MDD – recurrent episode	2668 (46.9)
Dysthymia	682 (12.0)
Posttraumatic Stress Disorder	794 (13.6)
Social Phobia	776 (8.5)
Generalized Anxiety Disorder	481 (8.5)
Panic Disorder	1392 (24.5)
Obsessive-compulsive disorder	414 (7.3)
DAPP-SF (sub)scales	
Emotional Dysregulation	2.7 (0.66)
Submissiveness	2.9 (0.92)
Cognitive distortion	2.3 (0.95)
Identity problems	3.1 (0.99)
Affective lability	3.2 (0.85)
Oppositionality	2.8 (0.89)
Anxiousness	3.4 (0.92)
Suspiciousness	2.2 (0.98)
Social avoidance	3.0 (1.06)
Narcissism	2.4 (0.82)
Insecure attachment	2.9 (1.11)
Self-harm	1.8 (0.95)
Inhibition	2.8 (0.65)
Intimacy problems	2.4 (0.84)
Restricted expression	3.2 (0.85)
Compulsivity	2.9 (0.95)
Dissocial Behavior	1.9 (0.54)
Stimulus seeking	2.1 (0.81)
Callousness	1.8 (0.60)
Rejection	2.3 (0.82)
Conduct problems	1.4 (0.57)

Note. "BSI" denotes the Brief Symptom Inventory, DAPP-SF denotes Dimensional Assessment of Personality Pathology Short Form, "MDD" denotes major depressive disorder. DAPP-SF scales are demonstrated as mean item score (1–5).

4.3.2 Total Effect of PPT on Treatment Outcome (Figure 1A)

The total effect of PPT on treatment outcome is presented in Table 2 under "Total effect of PPT (IV) on treatment outcome (DV)'" and Table 3. Table 2 shows the total effect of PPT on treatment outcome, which is defined as the posttreatment BSI score. All higher order constructs of PPT were significantly associated with treatment outcome (i.e. less improvement), ranging from β = 0.10 (*SE* = 0.02, *p* < .001) for Compulsivity to β = 0.43 (*SE* = 0.02, *p* < .001) for Emotional Dysregulation. We found similar results for the subgroups anxiety, depression, or combined group (see Appendix Table 2).

Table 2

Predicting treatment outcome with DAPP-SF higher order constructs of pathological personality traits (PPT) mediated by baseline level of symptoms within patients with depression and/or a anxiety disorder (see also Figure 1 A and B)

Independent variable (IV)	Total effect of PPT (IV) on treatment outcome (DV)	Direct effect of PPT (IV) on treatment outcome (DV)	Effect of PPT (IV) on baseline symptom level (M)	Effect of baseline symptom level (M) on treatment outcome (DV)	Mediating effect
In Figure 1 denoted as:	**c**	**c'**	**a**	**b**	**a × b; 95% CI**
Total (*n* = 5,689)					
Emotional Dysregulation	0.43***	0.11 ***	0.67***	0.45***	0.31 [0.28, 0.33]
Inhibition	0.24***	0.08***	0.32***	0.51***	0.17 [0.15, 0.18]
Compulsivity	0.10***	−0.02	0.22***	0.54***	0.12 [0.11, 0.14]
Dissocial Behavior	0.15***	0.04 **	0.22***	0.53***	0.12 [0.10, 0.13]

Note. All variables are standardized. DAPP-SF subscale represents the independent variable (IV), baseline (BSI sum score at baseline) represents the mediating variable (M), and posttest (BSI sum score at follow up) represents the dependent variable (DV). "*c*" denotes direct effect, "*c*" denotes total effect, "*a*" denotes effect of IV on M, "*b*" denotes effect of M on Y, "*a* × *b*" denotes indirect mediating effect. Analyses are adjusted for age and gender.

****p value* <.001;

***p* value <.01

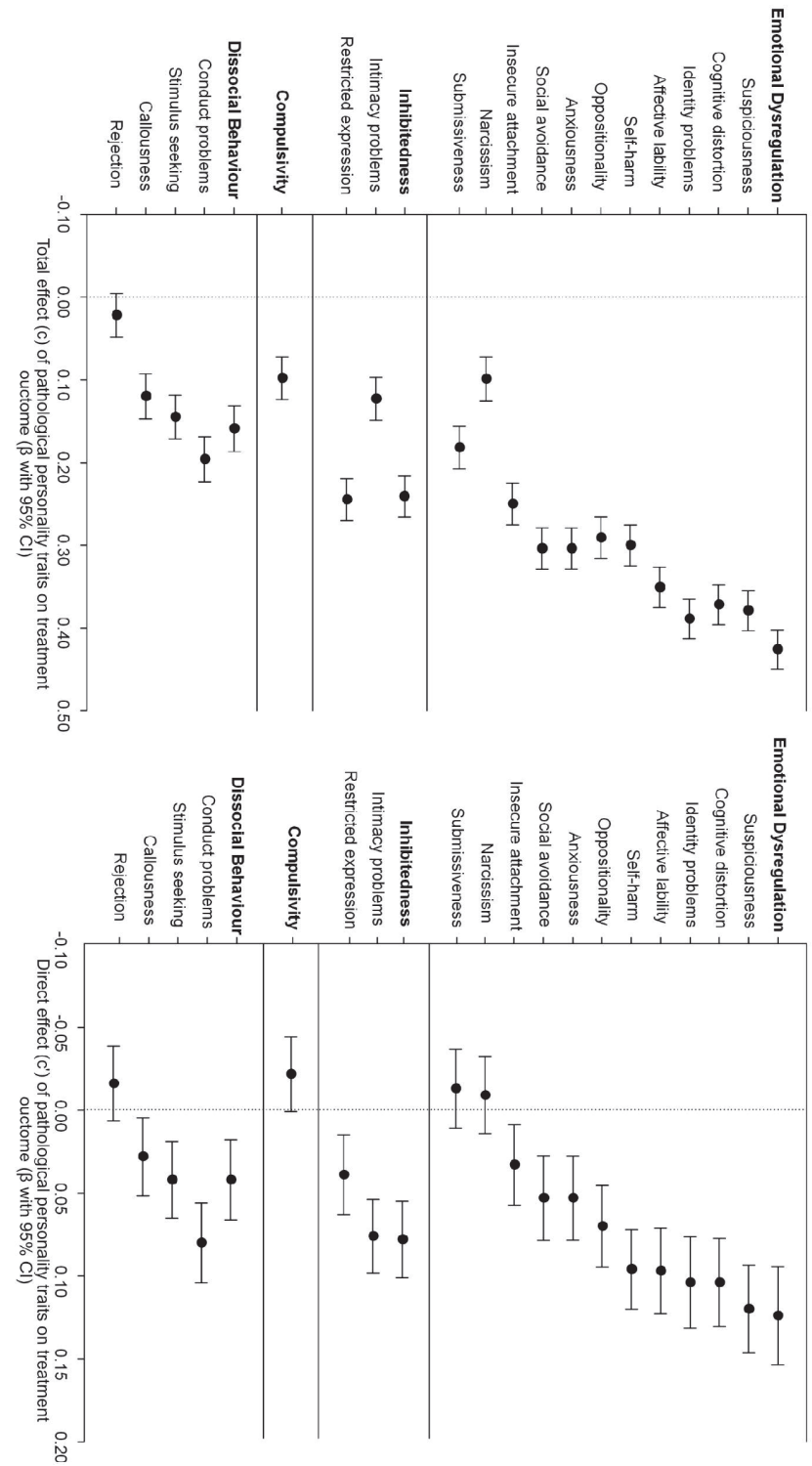

A.
Emotional Dysregulation
Suspiciousness
Cognitive distortion
Identity problems
Affective lability
Self-harm
Oppositionality
Anxiousness
Social avoidance
Insecure attachment
Narcissism
Submissiveness
Inhibitedness
Intimacy problems
Restricted expression
Compulsivity
Dissocial Behaviour
Conduct problems
Stimulus seeking
Callousness
Rejection
-0.10 0.00 0.10 0.20 0.30 0.40 0.50
Total effect (c) of pathological personality traits on treatment outcome (β with 95% CI)
B.
Emotional Dysregulation
Suspiciousness
Cognitive distortion
Identity problems
Affective lability
Self-harm
Oppositionality
Anxiousness
Social avoidance
Insecure attachment
Narcissism
Submissiveness
Inhibitedness
Intimacy problems
Restricted expression
Compulsivity
Dissocial Behaviour
Conduct problems
Stimulus seeking
Callousness
Rejection
-0.10 -0.05 0.00 0.05 0.10 0.15 0.20
Direct effect (c') of pathological personality traits on treatment outcome (β with 95% CI)

Figure 2. A demonstrates the total effect (c) of individual DAPP-SF pathological personality traitson treatment outcome (posttest BSI score). B demonstrates the direct effect (c') of individual DAPP-SF pathological personality traits on treatment outcome (posttest BSI score).

Regarding the individual subscales underlying the higher order constructs (Figure 2A and Appendix Table 3), we found beta-coefficients ranging from β = 0.02 (*SE* = 0.01, *p* = .09) for Rejection to β = 0.39 (*SE* = 0.01, *p* < .001) for Identity Problem. The subscales Identity Problems (β = 0.39, *SE* = 0.01, *p* < .001), Suspiciousness (β = 0.38, *SE* = 0.01, *p* < .001), Cognitive Distortion (β = 0.37, *SE* = 0.01, *p* < .001), and Affective Lability (β = 0.36, *SE* = 0.02, *p* < .001) demonstrated the strongest effects and were all part of the Emotional Dysregulation higher order construct. The subscale Rejection (part of the Dissocial Behavior construct) demonstrated a remarkably lower effect on treatment outcome compared to the other subscales.

4.3.3 Association between PPT and Baseline Symptom Level (Figure 1B)

The relationships between the DAPP-SF higher order constructs and BSI baseline symptom level for the total group are presented in Table 2 under "Effect of PPT (IV) on baseline symptom level (M)". We found that all constructs were significantly, *p* < .001, related to baseline BSI symptom level, ranging from 0.22 (*SE* = 0.02, *p* < .001) for Dissocial Behavior to β = 0.67 (*SE* = 0.02, *p* < .001) for Emotional Dysregulation within the total sample. We found no consistent differences in the magnitude of this association between the subgroups (see Appendix Table 2).

When assessing the underlying DAPP subscales of the higher order constructs, we found large differences in association with baseline symptom level. The subscales Identity Problems (β = 0.61, *SE* = 0.01, *p* < .001), Cognitive Distortion (β = 0.57, *SE* = 0.01, *p* < .001), Suspiciousness (β = 0.56, *SE* = 0.01, *p* < .001), and Affective Lability (β = 0.53, *SE* = 0.011, *p* < .001) demonstrated the strongest associations with baseline symptom level and were all part of the Emotional Dysregulation construct. Subscales Rejection (β = 0.08, *SE* = 0.01, *p* < .001; "Rejecting others") and Intimacy Problems (β = 0.09, *SE* = 0.01, *p* < .001) demonstrated the lowest associations regarding baseline symptom level and were part of Dissocial Behavior and Inhibition, respectively (see Appendix Table 3).

4.3.4 Mediation of Baseline Symptom Level (Figure 1B)

The relationship between PPT and treatment outcome was mediated by baseline symptom level. Table 2 under "Mediating effect" shows the results of the mediation analysis of PPT in relation to treatment outcome, with baseline symptom level as the M (mediator). We found

a strong mediating effect ($a \times b$) of baseline symptom level, with coefficients ranging from β = 0.12, 95% CI [0.10, 0.13], for Dissocial Behavior to β = 0.31, [0.28, 0.33], for Emotional Dysregulation.

The direct effect of PPT (c'), which takes into account the mediating effect of pretreatment level of symptoms, was approximately one third of the total effect and remained significant for Emotional Dysregulation, Inhibition, and Dissocial Behavior but was no longer significant for Compulsivity. This suggests that the effect is largely, but not entirely, mediated through the effects of baseline symptom level. The direct effect ranged from β = −0.02 (*SE* = 0.02, *p* = .071) for Compulsivity to β = 0.11 (*SE* = 0.02, *p* < .001) for Emotional Dysregulation. Individual DAPP-SF subscales demonstrated similar proportions of the total effect being mediated through baseline symptom level (see Figure 2B). The direct effect was no longer significant for the subscales Narcissism, Submissiveness, and Rejection. On average, Emotional Dysregulation demonstrated the strongest effect on treatment outcome. There were no consistent differences in the diagnostic subgroups (see Appendix Table 3).

Table 3

Moderating effects of baseline level of symptoms when predicting treatment outcome with DAPP-SF higher order constructs of pathological personality traits (PPT), within patients with depression and/or a anxiety disorder (see also Figure 1 C)

Treatment Outcome: posttreatment BSI score	**Interaction PPT (IV) with Baseline symptom level (Mo)**		**Effect PPT (IV) for 1 *SD* below Mean baseline level of symptoms (Mo)**		**Effect PPT (IV) for mean baseline level of symptoms (Mo)**		**Effect PPT (IV) for 1 *SD* above Mean baseline level of symptoms (Mo)**	
In Figure 1 denoted as:			**d – low baseline symptoms**		**d**		**d – high baseline symptoms**	
Independent variable (IV)	**Beta (SE)**	***p* value**	**Beta (SE)**	***p* value**	**Beta (SE)**	***p* value**	**Beta (SE)**	***p* value**
Total (*n* = 5,689)								
Emotional Dysregulation	0.061 (0.010)	<.001	0.070 (0.017)	<.001	0.130 (0.015)	<.001	0.191 (0.019)	<.001
Inhibition	0.062 (0.010)	<.001	0.012 (0.016)	.464	0.043 (0.012)	<.001	0.135 (0.015)	<.001
Compulsivity	−0.009 (0.011)	.378	−0.010 (0.016)	.546	−0.019 (0.011)	.096	−0.028 (0.015)	.061
Dissocial Behavior	−0.012 (0.011)	<.265	0.052 (0.018)	.003	0.039 (0.012)	.001	0.028 (0.015)	.066

Note. DAPP-SF subscale represents the independent variable (IV). Baseline BSI score represents the moderator variable (Mo). Beta denotes standardized regression coefficients. SE denotes standard error. Analyses are adjusted for age and gender.

4.3.5 Moderation of Baseline Symptom Level (Figure 1C)

Baseline symptom level was examined as a moderator of the relationship between PPT and treatment outcome and is demonstrated in Table 3. Baseline symptom level was a significant moderator of the relationship between Emotional Dysregulation and Inhibition and treatment outcome. Interaction effects between PPT and baseline symptom level were statistically significant for Emotional Dysregulation (β = 0.061, *SE* = 0.010, $p < .001$) and Inhibition (β = 0.062, *SE* = 0.062, $p < .001$). No significant interaction effect was found for Compulsivity and Dissocial Behavior. The standardized simple slope of Emotional Dysregulation for participants with one *SD* below the mean of baseline was 0.070, the standardized simple slope for participants with a mean level of baseline severity was 0.130, and the standardized simple slope for participants with one *SD* above mean baseline severity was 0.191. The standardized simple slope of Inhibition for participants with one *SD* below the mean of baseline was 0.012, the standardized simple slope for participants with a mean level of baseline severity was 0.043, and the standardized simple slope for participants with one *SD* above mean baseline severity was 0.135. Thus, Emotional Dysregulation and Inhibition were most predictive of high BSI score after treatment among participants with high baseline symptom level. These results were similar across separate diagnostic groups, though for the anxiety subgroup the interaction between Inhibition and baseline symptom level was no longer statistical significant. The results for each diagnostic group separately is demonstrated in Appendix Table 5.

All subscales that were part of Emotional Dysregulation and Inhibition and with the addition of Rejection demonstrated significant interaction effects (see Appendix Table 4). Interestingly, among patients with high baseline symptom level Narcissism had a beneficial effect on treatment outcome, though with a small effect size (β = −0.34, *SE* = 0.016, $p = .032$).

4.4 Discussion

We examined the effects of dimensional levels of PPT on treatment outcome after 6 to 8 months of treatment in a large sample of outpatients with depressive disorders, anxiety disorders, and combined depressive/anxiety disorders. The findings support our hypothesis that PPT is strongly related to higher symptom levels both before and after treatment, even when patients do not meet criteria for a personality disorder. Patients with one *SD* higher dimensional level of PPT had on average 0.20 to 0.43 *SD* higher levels of general psychopathology (BSI) after receiving treatment. At first glance, this suggests that dimensional levels of PPT had a significant and seemingly clinically relevant predictive effect on treatment outcome. However, when taking baseline symptom level into account, we found that patients with high symptom levels at baseline had substantially higher symptom levels after treatment regardless of PPT level. Baseline symptom level could be considered an important mediator of the relationship between PPT and treatment outcome. PPT was related to higher baseline symptom levels. The direct adverse effect (c') of PPT on outcome when baseline symptom level was taken into account was approximately one third of the total. This direct effect was no longer significant for Compulsivity. Furthermore, we found that the baseline symptom level moderated the predictive effects of Emotional Dysregulation and Inhibition, which were slightly more predictive of treatment outcome among participants with high baseline symptom level. However, the effect size of this interaction was small. We found a similar effect of PPT on treatment outcome among the three patient groups (see Appendix).

Our results replicate findings of previous studies, in which PPT was found to have a negative impact on treatment outcome in patients with anxiety- and depressive disorders [29, 30, 55-58]. Many studies, however, did not factor in the importance of baseline symptom levels. Because baseline symptom levels proved to have a strong and consistent relation to treatment outcome in the present and in previous studies, it is plausible that PPT has less prognostic value when researchers adjust for baseline symptom levels [8, 37]. Previous studies have also found higher levels of symptomatology (both pre- and posttreatment) when PPT was present, but with a similar symptom decline during treatment [43]. Studies that adjusted for baseline symptom levels found (at most) a small effect of PPT on treatment

outcome for both depressive- and anxiety disorders, or no effect [10, 59]. In this regard, the findings of the current study are in line with prior literature. We approached baseline symptom level as a mediating variable in which PPT is related to higher symptom severity and perceived stress at baseline/ which in turn leads to higher levels of symptoms after treatment [60]. Moreover, for PPT constructs Emotional Dysregulation and Inhibition, baseline symptom level served as a moderator variable, in which PPT was more predictive for adverse treatment outcome when patients experienced high symptom severity. This is in line with previous literature which found that baseline symptom severity was a moderator for treatment outcome for substance use disorders [38] and borderline personality disorder [39, 40]. The present study is the first to assess the moderating effects of baseline severity on treatment outcome among depression and anxiety patients.

Conventionally, the relationship between PPT and depression/anxiety may be considered as an etiological one, in which PPT causes higher symptom levels of psychopathology. Researchers have demonstrated that PPT can be a predictor for future psychopathology in response to life stress [61]. Furthermore, PPT can cause increased levels of distress because it contributes to problems in physical health, increased financial difficulties, dissolution of relationships, and other negative life outcomes [62]. PPT likely hampers patients to cope with the burden of depression or anxiety [63]. In line with this, we found that PPT was associated with higher symptom levels of depression and anxiety at both pre- and posttreatment. In particular, we found that Suspiciousness, Cognitive Distortion, Identity Problems, and Affective Lability related strongly to symptom level before and after treatment; these constructs may be especially linked to maladaptive reactions to life events.

PP is generally thought to be present before depression and anxiety; however, Widiger (2011) posited the presence of a *pathoplastic* as well as a *spectrum* relationship in addition to an *etiological* one. A pathoplastic relationship would suggest that the presentation and expression of PPT and psychopathology (in this case depression and/or anxiety) would bidirectionally influence each other. Both PPT and depression/anxiety are considered impairments to how an individual thinks, feels, and behaves in relation to others. A priori PPT results in higher levels of impairment in these areas, resulting in higher levels of reported depression/anxiety, but high levels of psychopathology may also influence the reported level of PP. Patients who are very anxious or depressed may fail to provide accurate self-

descriptions [64-66]. Although some may consider the above as self-report bias, others argue that PPT causes patients to respond to stress with (or relapse in) depression. Thus, self-reported levels of depression are considered accurate expressions of underlying PP. Subsequently, patients who report lower (depression) symptom levels after treatment may also display a decrease in levels of PPT [67]. In further support of a pathoplastic relationship, levels of reported PPT were substantially higher when patients were diagnosed with both a depression and an anxiety disorder and had a higher BSI baseline symptom level. Unfortunately, we only measured PPT at baseline and therefore cannot make statements about the posttreatment decrease of PPT alongside the decrease of depression and anxiety. Alternatively, our findings can be interpreted in terms of a spectrum relationship. PPT and depression/anxiety can be (partly) considered as manifestations of one and the same underlying common spectrum [65]. In support of a spectrum relationship, we found the strongest associations with the higher order construct of Emotional Dysregulation, which has demonstrated overlap with depression and anxiety. Symptoms of anxiety and depression may lie in the same spectrum as Emotional Dysregulation. In our study, PPT was measured at the same time point as baseline symptom level. According to earlier findings [59] and the theory of the pathoplastic and spectrum relationships, PPT was likely influenced by an individual's current depressive or anxious state, which could have affected our mediation analyses.

Our findings could be valuable for clinical practice with regard of making prognosis. We found that baseline symptom level had far greater prognostic value compared to PPT measured with the DAPP-SF. The DAPP-SF, however, was still of added predictive value. Moreover, the DAPP-SF may provide relevant patient-specific information, which may be a focus for psychological therapy [27, 68]. With regard of treatment, we found that patients with high levels of PPT experience higher symptom levels after 6 to 8 months of treatment for depression and anxiety. The implicationsregarding to treatment can be interpreted in several ways. One can argue that patients with concurrent high levels of PPT do benefit from a treatment that does not necessarily focusses on Personality Pathology. An additional treatment aimed at PPT may be appropriate only for patients who remain symptomatic in spite of treatment. Moreover, it is likely that patients with higher levels of PPT simply need to be treated longer in order to achieve full remission in symptoms [69]. However, one could also argue that patients with high PPT should be treated differently or more intensely, in order to achieve the same

symptom level after 6 to 8 months of treatment as their lower PPT counterparts [70]. Both of these treatment options need further research and policymaking, in which clinical aspects and efficiency play a role [71, 72].

4.4.1 Strengths and Limitations

The strengths of our study include its large sample size and the distinction of diagnostic groups of depression and anxiety. By collecting data in a naturalistic setting, we were able to analyze data from a clinical sample, which was representative of day-to-day patient care. We also measured PPT dimensionally, which is considered a strength in light of how PPT is currently conceptualized. Previous studies have consistently criticized categorical definitions of PPT (i.e., personality disorders), and there is still no consensus on how to best classify patients with personality problems [13, 17]. Dimensional levels of PPT do not equate to personality disorders, but there is evidence that PPT could be a reasonable proxy for the personality disorder diagnosis itself [73-76]. Contrary to most studies, we assessed the intermediary effects of baseline symptom severity as both a mediator and a moderator in the prospective relation of PPT to treatment outcome.

Our findings should also be considered in light of their limitations. First, personality pathology is a broad concept, which could also include other definitions such as psychodynamic functioning, personality organization, coping styles, attachment constructs, etc. Though the DAPP-SF is based on 18 empirically sound factors [18, 19] and increasingly used as a proxy measure for the Alternative DSM-5 model of personality disorders B-criterium personality traits [27], caution is warranted when generalizations are made to other realms of personality. Second, with the current study design, causality between PPT and baseline symptom level was assumed but could not be formally analyzed because both were measured at the same time point. Mediation analysis is fitting when the results are interpreted as a etiological relationship between PPT and depression/anxiety. As discussed, the reality may be more complex. Third, we limited the assessment of outcome to 6 to 8 months of treatment. Some patients did not complete their follow-up and were left out of the analysis, potentially introducing selection bias [77]. Fourth, we lacked information regarding the type of treatment patients received (psychotherapy, medication, or both). This may be relevant because certain treatments may be better suited to patients with PPT than others [78]. Fourth, patients with personality disorders as primary diagnoses were referred to other departments and therefore

not included in the present study. Therefore. our sample might not have been representative for patients with the highest levels of PP. Lastly, PPT was only measured once, and not repeatedly. Earlier studies demonstrated that a decrease of (self-reported) PPT can occur after psychopathology is treated and has declined [64-66].

4.4.2 Conclusions

We expanded the way in which researchers can examine the prognostic value of PPT for treatment outcome in depressive- and/or anxiety disorders. Our results showed that PPT had a negative effect on treatment for patients with anxiety- and depressive disorders, of which the PPT constructs Emotional Dysregulation and Inhibition among participants with high baseline symptom level demonstrated the strongest effect. This effect was to a large extend mediated by baseline symptom levels. High PPT was related to both higher symptom levels before and after treatment. and the added (direct) effect of PPT on symptom decline after treatment was relatively small. Moreover, the effects of Emotional Dysregulation and Inhibition were also moderated, and demonstrated to have a stronger effect on treatment outcome when patients experienced high baseline severity, although with a small effect size.

References

1. Bienvenu, O.J., et al., *Normal personality traits and comorbidity among phobic, panic and major depressive disorders.* Psychiatry research, 2001. **102**(1): p. 73-85.
2. Shea, M.T. and S. Yen, *Personality traits/disorders and depression: A summary of conceptual and empirical findings.* 2005.
3. Friborg, O., et al., *Comorbidity of personality disorders in anxiety disorders: A meta-analysis of 30 years of research.* Journal of affective disorders, 2013. **145**(2): p. 143-155.
4. American-Psychiatric-Association, *Diagnostic and statistical manual of mental disorders (DSM-5®)*. 2013: American Psychiatric Pub.
5. Organization, W.H., *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. 1992: Geneva: World Health Organization.
6. First, M.B., et al., *The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders.* World Psychiatry, 2015. **14**(1): p. 82-90.
7. Friborg, O., et al., *Comorbidity of personality disorders in mood disorders: a meta-analytic review of 122 studies from 1988 to 2010.* Journal of affective disorders, 2014. **152**: p. 1-11.
8. Mulder, R.T., *Personality pathology and treatment outcome in major depression: a review.* American Journal of Psychiatry, 2002. **159**(3): p. 359-371.
9. Feske, U., et al., *Avoidant personality disorder as a predictor for treatment outcome among generalized social phobics.* Journal of personality disorders, 1996. **10**(2): p. 174-184.
10. Kampman, M., et al., *Outcome prediction of cognitive behaviour therapy for panic disorder: Initial symptom severity is predictive for treatment outcome, comorbid anxiety or depressive disorder, cluster C personality disorders and initial motivation are not.* Behavioural and Cognitive Psychotherapy, 2008. **36**(1): p. 99-112.
11. Reich, J., *The effect of Axis II disorders on the outcome of treatment of anxiety and unipolar depressive disorders: a review.* Journal of personality disorders, 2003. **17**(5): p. 387-405.
12. Newton-Howes, G., P. Tyrer, and T. Johnson, *Personality disorder and the outcome of depression: meta-analysis of published studies.* The British Journal of Psychiatry, 2006. **188**(1): p. 13-20.
13. Newton-Howes, G., et al., *Influence of personality on the outcome of treatment in depression: systematic review and meta-analysis.* Journal of Personality Disorders, 2014. **28**(4): p. 577-593.
14. Olatunji, B.O., J.M. Cisler, and D.F. Tolin, *A meta-analysis of the influence of comorbidity on treatment outcome in the anxiety disorders.* Clinical psychology review, 2010. **30**(6): p. 642-654.
15. Widiger, T.A. and D.B. Samuel, *Diagnostic categories or dimensions? A question for the Diagnostic and statistical manual of mental disorders.* Journal of abnormal psychology, 2005. **114**(4): p. 494.
16. Haslam, N., E. Holland, and P. Kuppens, *Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research.* Psychological medicine, 2012. **42**(5): p. 903-920.
17. Widiger and Simonsen, *Alternative dimensional models of personality disorder: Finding a common ground.* Journal of personality disorders, 2005. **19**(2): p. 110-130.
18. Livesley, W.J., D.N. Jackson, and M.L. Schroeder, *A study of the factorial structure of personality pathology.* Journal of Personality Disorders, 1989. **3**(4): p. 292-306.
19. Livesley, W.J., *The phenotypic and genotypic structure of psychopathic traits*, in *Psychopathy: Theory, research and implications for society*. 1998, Springer. p. 69-79.
20. Livesley, W.J. and R.M. Larstone, *The Dimensional Assessment of Personality Pathology (DAPP).* 2008.
21. Clark, L.A. and W.J. Livesley, *Two approaches to identifying the dimensions of personality disorder: Convergence on the five-factor model.* 2002.
22. Gutiérrez, F., et al., *Toward an integrated model of pathological personality traits: Common hierarchical structure of the PID-5 and the DAPP-BQ.* Journal of personality disorders, 2020. **34**(Supplement C): p. 25-39.
23. Van den Broeck, J., et al., *Hierarchical structure of maladaptive personality traits in older adults: Joint factor analysis of the PID-5 and the DAPP-BQ.* Journal of personality disorders, 2014. **28**(2): p. 198-211.
24. Pryor, L.R., J.D. Miller, and E.T. Gaughan, *Testing two alternative pathological personality measures in the assessment of psychopathy: An examination of the SNAP and DAPP-BQ.* Journal of Personality Disorders, 2009. **23**(1): p. 85-100.

25. Rossi, G., I. Debast, and S. Van Alphen, *Measuring personality functioning in older adults: construct validity of the Severity Indices of Personality Functioning–Short Form (SIPP-SF).* Aging & mental health, 2017. **21**(7): p. 703-711.
26. Berghuis, H., J.H. Kamphuis, and R. Verheul, *Specific personality traits and general personality dysfunction as predictors of the presence and severity of personality disorders in a clinical sample.* Journal of personality assessment, 2014. **96**(4): p. 410-416.
27. Berghuis, H., et al., *Assessment of pathological traits in DSM-5 personality disorders by the DAPP-BQ: How do these traits relate to the six personality disorder types of the alternative model?* Journal of Personality Disorders, 2019. **33**(1): p. 49-70.
28. Carlier, et al., *A comparative analysis of personality pathology profiles among patients with pure depressive-, pure anxiety-, and pure somatoform disorders.* Journal of affective disorders, 2014. **168**: p. 322-330.
29. Schat, A., et al., *Predictors of outcome in outpatients with anxiety disorders: the Leiden routine outcome monitoring study.* Journal of psychiatric research, 2013. **47**(12): p. 1876-1885.
30. Van Noorden, M.S., et al., *Predicting outcomes of mood, anxiety and somatoform disorders: The Leiden routine outcome monitoring study.* Journal of Affective Disorders, 2012. **142**(1-3): p. 122-131.
31. Boschloo, L., et al., *The Four-Year Course of Major Depressive Disorder: The Role of Staging and Risk Factor Determination.* Psychotherapy and Psychosomatics, 2014. **83**(5): p. 279-288.
32. Spinhoven, P., et al., *Positive and negative life events and personality traits in predicting course of depression and anxiety.* Acta Psychiatrica Scandinavica, 2011. **124**(6): p. 462-473.
33. Ananth, C.V. and E.F. Schisterman, *Confounding, causality, and confusion: the role of intermediate variables in interpreting observational studies in obstetrics.* American journal of obstetrics and gynecology, 2017. **217**(2): p. 167-175.
34. Rhebergen, D., et al., *Course trajectories of unipolar depressive disorders identified by latent class growth analysis.* Psychological medicine, 2012. **42**(7): p. 1383-1396.
35. Wardenaar, K.J., H.J. Conradi, and P. de Jonge, *Data-driven course trajectories in primary care patients with major depressive disorder.* Depression and anxiety, 2014. **31**(9): p. 778-786.
36. Batelaan, N.M., et al., *Two-year course trajectories of anxiety disorders: do DSM classifications matter?* The Journal of clinical psychiatry, 2014. **75**(9): p. 985-993.
37. Kvarstein, E.H. and S. Karterud, *Large variation of severity and longitudinal change of symptom distress among patients with personality disorders.* Personality and Mental Health, 2013. **7**(4): p. 265-276.
38. Ball, S.A., et al., *Reliability of personality disorder symptoms and personality traits in substance-dependent inpatients.* Journal of abnormal psychology, 2001. **110**(2): p. 341.
39. Bos, E.H., et al., *Effectiveness of systems training for emotional predictability and problem solving (STEPPS) for borderline personality problems in a 'real-world'sample: Moderation by diagnosis or severity?* Psychotherapy and psychosomatics, 2011. **80**(3): p. 173-181.
40. Sahin, Z., et al., *Clinical severity as a moderator of outcome in psychodynamic and dialectical behavior therapies for borderline personality disorder.* Personality Disorders: Theory, Research, and Treatment, 2018. **9**(5): p. 437.
41. Pereira-Morales, A.J., et al., *Personality traits and health-related quality of life: the mediator role of coping strategies and psychological distress.* Annals of general psychiatry, 2018. **17**(1): p. 25.
42. van Kampen, D., E. de Beurs, and H. Andrea, *A short form of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ): the DAPP-SF.* Psychiatry Research, 2008. **160**(1): p. 115-128.
43. van Eeden, W.A., et al., *Neuroticism and chronicity as predictors of 9-year course of individual depressive symptoms.* Journal of affective disorders, 2019. **252**: p. 484-492.
44. de Beurs, et al., *Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice.* Clinical psychology & psychotherapy, 2011. **18**(1): p. 1-12.
45. Van Vliet, I. and E. De Beurs, *The MINI-International Neuropsychiatric Interview. A brief structured diagnostic psychiatric interview for DSM-IV en ICD-10 psychiatric disorders.* Tijdschrift voor psychiatrie, 2007. **49**(6): p. 393.
46. Carlier, I., et al., *Treatment course and its predictors in patients with somatoform disorders: A routine outcome monitoring study in secondary psychiatric care.* Clinical psychology & psychotherapy, 2018. **25**(4): p. 550-564.
47. Cuijpers, P., et al., *The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: A meta-analysis of direct comparisons.* World Psychiatry, 2013. **12**(2): p. 137-148.

48. van Fenema, E.M., et al., *Vitality predicts level of guideline-concordant care in routine treatment of mood, anxiety and somatoform disorders.* Journal of evaluation in clinical practice, 2012. **18**(2): p. 441-448.
49. de Beurs, et al., *Reliability and validity of the Dutch Dimensional Assessment of Personality Pathology-Short Form (DAPP-SF), a shortened version of the DAPP-Basic Questionnaire.* Journal of Personality Disorders, 2009. **23**(3): p. 308-326.
50. Derogatis, L.R. and N. Melisaratos, *The brief symptom inventory: an introductory report.* Psychological medicine, 1983. **13**(03): p. 595-605.
51. De Beurs and Zitman, *The Brief Symptom Inventory (BSI): reliability and validity of a practical alternative to SCL-90.* MGV, 2006. **61**: p. 120-41.
52. Preacher, K.J. and A.F. Hayes, *Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models.* Behavior research methods, 2008. **40**(3): p. 879-891.
53. Hayes, A.F., *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. 2017: Guilford Publications.
54. Loose, T., D. Acier, and G. El-Baalbaki, *Drinking motives as mediators between personality traits and alcohol use among young French people.* Personality and Individual Differences, 2018. **134**: p. 268-274.
55. Telch, M.J., J.H. Kamphuis, and N.B. Schmidt, *The effects of comorbid personality disorders on cognitive behavioral treatment for panic disorder.* Journal of Psychiatric Research, 2011. **45**(4): p. 469-474.
56. Goddard, E., J. Wingrove, and P. Moran, *The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety.* 2015. **73**: p. 1-7.
57. Shea, M.T., et al., *Course of depressive symptoms over follow-up: findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program.* Archives of general psychiatry, 1992. **49**(10): p. 782-787.
58. van den Hout, M., C. Brouwers, and J. Oomen, *Clinically diagnosed axis II co-morbidity and the short term outcome of CBT for axis I disorders.* Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 2006. **13**(1): p. 56-63.
59. Blom, M.B., et al., *Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression.* Journal of affective disorders, 2007. **104**(1-3): p. 119-126.
60. Candrian, M., et al., *Perceived stress and cognitive vulnerability mediate the effects of personality disorder comorbidity on treatment outcome in major depressive disorder: a path analysis study.* The Journal of nervous and mental disease, 2007. **195**(9): p. 729-737.
61. Malouff, J.M., E.B. Thorsteinsson, and N.S. Schutte, *The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis.* Journal of Psychopathology and Behavioral Assessment, 2005. **27**(2): p. 101-114.
62. Lahey, B.B., *Public health significance of neuroticism.* American Psychologist, 2009. **64**(4): p. 241.
63. Dunkley, D.M., D.C. Zuroff, and K.R. Blankstein, *Self-critical perfectionism and daily affect: dispositional and situational influences on stress and coping.* Journal of personality and social psychology, 2003. **84**(1): p. 234.
64. Fava, M., et al., *Personality disorder comorbidity with major depression and response to fluoxetine treatment.* Psychotherapy and Psychosomatics, 1994. **62**(3-4): p. 160-167.
65. Widiger, A.T., *Personality and psychopathology.* World Psychiatry, 2011. **10**(2): p. 103-106.
66. Gunderson, J.G., et al., *Plausibility and possible determinants of sudden "remissions" in borderline patients.* Psychiatry: Interpersonal and Biological Processes, 2003. **66**(2): p. 111-119.
67. Costa Jr, P.T., et al., *Personality self-reports are concurrently reliable and valid during acute depressive episodes.* Journal of affective disorders, 2005. **89**(1-3): p. 45-55.
68. Antony, M.M. and D.H. Barlow, *Handbook of assessment and treatment planning for psychological disorders*. 2020: Guilford Publications.
69. Boer, S., et al., *Prediction of prolonged treatment course for depressive and anxiety disorders in an outpatient setting: The Leiden routine outcome monitoring study.* Journal of affective disorders, 2019. **247**: p. 81-87.
70. Angstman, K. and N.H. Rasmussen, *Personality disorders: review and clinical application in daily practice.* American family physician, 2011. **84**(11): p. 1253-1260.
71. Donohue, J.M. and H.A. Pincus, *Reducing the societal burden of depression.* Pharmacoeconomics, 2007. **25**(1): p. 7-24.
72. Helmchen, H. and M. Linden, *Subthreshold disorders in psychiatry: clinical reality, methodological artifact, and the double-threshold problem.* Comprehensive Psychiatry, 2000. **41**(2): p. 1-7.

73. Katz, A.C., et al., *A family study of the DSM-5 section III personality pathology model using the personality inventory for the DSM-5 (PID-5).* Journal of personality disorders, 2018. **32**(6): p. 753-765.
74. Widiger and Costa, *Integrating normal and abnormal personality structure: the five-factor model.* Journal of Personality, 2012. **80**(6): p. 1471-1506.
75. Few, L.R., et al., *Trait-based assessment of borderline personality disorder using the NEO Five-Factor Inventory: Phenotypic and genetic support.* Psychological assessment, 2016. **28**(1): p. 39.
76. Bernstein, D.P., et al., *Opinions of personality disorder experts regarding the DSM-IV personality disorders classification system.* Journal of Personality Disorders, 2007. **21**(5): p. 536-551.
77. Hoenders, R.H., et al., *Pitfalls in the assessment, analysis, and interpretation of routine outcome monitoring (ROM) data: Results from an outpatient clinic for integrative mental health.* Administration and Policy in Mental Health and Mental Health Services Research, 2014. **41**(5): p. 647-659.
78. van Bronswijk, S.C., et al., *The impact of personality disorder pathology on the effectiveness of cognitive therapy and interpersonal psychotherapy for major depressive disorder.* Journal of affective disorders, 2018. **225**: p. 530-538.

Acknowledgements

We gratefully acknowledge the essential contributions made by the participants of this study and the mental health care provider GGZ Rivierduinen. We thank M. Shahabi for her preparatory work in the context of her master thesis (Leiden University). We also thank Prof. dr. O.M. Dekkers (LUMC) for providing consultation regarding the statistical analyses.

Conflicts of Interest statement and sources of funding

The research project was funded by the Dutch Mental Health Care Provider GGZ Rivierduinen. The authors declare no conflict of interest.

Ethical Considerations

The Medical Ethical Committee of the LUMC approved the general study protocol—with Routine Outcome Monitoring (ROM) being integral to the treatment process (no written informed consent was required). All participants gave permission for the anonymized use of their data for research purposes.

Author Statement

Wessel A. van Eeden: conceptualisation, methodology, statistical analysis, writing – original draft. Albert M. van Hemert: writing – review and editing. Erik J. Giltay: methodology, statistical analysis, writing – figures, review and editing. Philip Spinhoven: writing – review and editing. Edwin de Beurs: writing – review and editing. Ingrid V. E. Carlier: conceptualization, writing – original draft, supervision. All authors have read and approved the submitted manuscript.

Disclosures

The authors declare no conflicts of interest.

Appendix

Appendix Table 1

Variable	Anxiety group (*n* = 1,993) Mean (*SD*) *n* (%)	Depression group (*n* = 1,664) Mean (*SD*) *n* (%)	Combined depression and anxiety group (*n* = 2,032) Mean (*SD*) *n* (%)
Age	36.7 (12.6)	41.6 (12.7)	38.6 (11.8)
Gender (female)	1293 (64.9)	955 (57.4)	1324 (65.2)
BSI baseline item score	1.00 (0.59)	1.32 (0.62)	1.659 (0.70)
BSI posttreatment item score	0.63 (0.57)	0.79 (0.65)	1.12 (0.82)
MDD – single episode	-	591 (35.5)	584 (28.7)
MDD – recurrent episode	-	958 (57.6)	1237 (60.9)
Dysthymia	-	249 (15.0)	366 (18.0)
Posttraumatic Stress Disorder	248 (12.4)	-	535 (26.3)
Social Phobia	351 (17.6)	-	415 (20.4)
Generalized Anxiety Disorder	220 (11.0)	-	254 (12.5)
Panic Disorder	718 (36.0)	-	671 (33.0)
Obsessive-compulsive disorder	219 (11.0)	-	192 (9.4)
DAPP-SF (sub)scales			
Emotional Dysregulation	2.49 (0.64)	2.72 (0.62)	2.96 (0.64)
Submissiveness	2.83 (0.93)	2.92 (0.89)	3.15 (0.92)
Cognitive distortion	2.06 (0.87)	2.33 (0.91)	2.60 (0.97)
Identity problems	2.65 (0.96)	3.24 (0.92)	3.45 (0.90)
Affective lability	3.01 (0.87)	3.22 (0.82)	3.49 (0.80)
Oppositionality	2.50 (0.84)	2.92 (0.87)	3.03 (0.87)
Anxiousness	3.17 (0.92)	3.33 (0.90)	3.64 (0.87)
Suspiciousness	1.92 (0.88)	2.10 (0.93)	2.41 (1.04)
Social avoidance	2.75 (1.07)	2.89 (1.00)	3.27 (1.02)
Narcissism	2.34 (0.82)	2.35 (0.81)	2.40 (0.82)
Insecure attachment	2.74 (1.08)	2.73 (1.09)	3.15 (1.09)
Self-harm	1.40 (0.70)	1.91 (0.98)	1.98 (1.03)
Inhibition	2.64 (0.62)	2.88 (0.63)	2.95 (0.65)
Intimacy problems	2.29 (0.78)	2.48 (0.83)	2.47 (0.88)
Restricted expression	2.99 (0.85)	3.27 (0.83)	3.42 (0.82)
Compulsivity	2.89 (0.94)	2.87 (0.92)	3.01 (0.97)
Dissocial Behavior	1.86 (0.50)	1.95 (0.55)	1.93 (0.55)
Stimulus seeking	1.99 (0.75)	2.18 (0.83)	2.16 (0.83)
Callousness	1.76 (0.58)	1.78 (0.60)	1.79 (0.61)
Rejection	2.32 (0.81)	2.39 (0.83)	2.30 (0.83)
Conduct problems	1.36 (0.52)	1.45 (0.59)	1.48 (0.60)

Note. "BSI" denotes the Brief Symptom Inventory, DAPP-SF denotes Dimensional Assessment of Personality Pathology Short Form, "MDD" denotes major depressive disorder. DAPP-SF scales are demonstrated as mean item score (1–5).

Appendix Table 2

Predicting treatment outcome with DAPP-SF higher order constructs of Pathological Personality Traits (PPT) mediated by baseline level of symptoms within a depression group, anxiety group, or combined group (see also Figure 1)

Independent variable (IV)	Total effect of PPT (IV) on treatment outcome (DV)	Direct effect of PPT (IV) on treatment outcome (DV)	Effect of PPT (IV) on baseline symptom level (M)	Effect of baseline symptom level (M) on treatment outcome (DV)	Mediating effect
In Figure 1 denoted as:	**c**	**c'**	**a**	**b**	**a × b; 95% CI**
Anxiety group (*n* = 1,993)					
Emotional Dysregulation	0.38***	0.18***	0.56***	0.34***	0.20 [0.16, 0.23]
Inhibition	0.17***	0.06***	0.24**	0.45***	0.11 [0.09, 0.13]
Compulsivity	0.10***	0.01	0.18***	0.47***	0.09 [0.07, 0.11]
Dissocial Behavior	0.17***	0.07***	0.21***	0.45***	0.10 [0.07, 0.12]
Depression group (*n* = 1,664)					
Emotional Dysregulation	0.35***	0.11***	0.58***	0.42***	0.24 [0.20, 0.29]
Inhibition	0.15***	0.06**	0.21***	0.48***	0.10 [0.07, 0.12]
Compulsivity	0.07**	−0.02	0.19***	0.50***	0.10 [0.07, 0.12]
Dissocial Behavior	0.14***	0.05*	0.19***	0.48***	0.09 [0.06, 0.12]
Combined depression and anxiety group (*n* = 2,032)					
Emotional Dysregulation	0.39***	0.07**	0.65***	0.48***	0.31 [0.27, 0.36]
Inhibition	0.24***	0.09***	0.30***	0.50***	0.15 [0.12, 0.17]
Compulsivity	0.08***	−0.04	0.24***	0.53***	0.13 [0.10, 0.15]
Dissocial Behavior	0.10***	0.01	0.18***	0.52***	0.09 [0.07, 0.12]

Note. All variables are standardized. DAPP-SF subscale represents the independent variable (IV), baseline (BSI sum score at baseline) represents the mediating variable (M), and posttest (BSI sum score at follow up) represents the dependent variable (DV). "*c*" denotes direct effect, "*c*" denotes total effect, "*a*" denotes effect of IV on M, "*b*" denotes effect of M on Y, "*a* × *b*" denotes indirect mediating effect. Analyses are adjusted for age and gender.

***p value <.001;

**p value <.01;

*p value <.05

Appendix Table 3

Predicting treatment outcome with DAPP-SF subscales of pathological personality traits (PPT) mediated by baseline level of symptoms within a group of patients with an anxiety disorder, depression, or both (n = 5,689)

Independent variable (IV)		Total effect of PPT (IV) on treatment outcome (DV)	Direct effect of PPT (IV) on treatment outcome (DV)	Effect of PPT (IV) on baseline symptom level (M)	Effect of baseline symptom level (M) on treatment outcome (DV)	Mediating effect
In Figure 1 denoted as:		***c***	***c'***	***a***	***b***	***a × b; 95% CI***
Emotional Dysregulation	Submissiveness	0.20***	−0.01	0.38***	0.54***	0.20[0.19, 0.22]
	Cognitive distortion	0.37***	0.10***	0.57***	0.48***	0.27 [0.25, 0.29]
	Identity problems	0.39 ***	0.10 ***	0.61***	0.47***	0.29 [0.27, 0.31]
	Affective lability	0.36 ***	0.10 ***	0.53***	0.48***	0.26 [0.24, 0.28]
	Oppositionality	0.29***	0.07***	0.44***	0.51***	0.22 [0.21, 0.24]
	Anxiousness	0.31***	0.05***	0.50***	0.51***	0.25 [0.24, 0.27]
	Suspiciousness	0.38***	0.11***	0.56***	0.47***	0.26 [0.34, 0.29]
	Social avoidance	0.31***	0.05***	0.50***	0.51***	0.25 [0.23, 0.27]
	Narcissism	0.11***	−0.01	0.23***	0.54***	0.11 [0.10, 0.13]
	Insecure attachment	0.25***	0.03**	0.42***	0.52***	0.22 [0.20, 0.24]
	Self-harm	0.30***	0.09***	0.42***	0.50***	0.21 [0.19, 0.22]
Inhibition	Intimacy problems	0.12***	0.08***	0.09***	0.53***	0.05 [0.03, 0.06]
	Restricted expression	0.24***	0.04**	0.40***	0.52***	0.21 [0.19, 0.23]
Compulsivity	Compulsivity	0.10***	−0.02	0.22***	0.54***	0.12 [0.11, 0.14]
Dissocial Behavior	Stimulus seeking	0.15***	0.04***	0.20***	0.53***	0.11 [0.09, 0.12]
	Callousness	0.12***	0.03*	0.17***	0.53***	0.09 [0.08, 0.11]
	Rejection	0.02	−0.02	0.08***	0.54***	0.03 [0.03, 0.05]
	Conduct problems	0.20***	0.08***	0.22***	0.52***	0.12 [0.10, 0.13]

Note. All variables are standardized. DAPP-SF subscale represents the independent variable (IV), baseline (BSI sum score at baseline) represents the mediating variable (M), and posttest (BSI sum score at follow up) represents the dependent variable (DV). "*c*" denotes direct effect, "*c*" denotes total effect, "*a*" denotes effect of IV on M, "*b*" denotes effect of M on Y, "*a* × *b*" denotes indirect mediating effect. Analyses are adjusted for age and gender.

****p* value <.001;

***p* value <.01;

**p* value <.05

Appendix Table 4

Moderating effects of baseline level of symptoms (Mo) when predicting treatment outcome with DAPP-SF subscales of pathological personality traits (PPT), within patients with depression and/or a anxiety disorder (see also Figure 1 C)

Treatment Outcome: posttreatment BSI score	**Interaction PPT (IV) with Baseline symptom level (Mo)**		**Effect PPT (IV) for 1 *SD* below Mean baseline level of symptoms (Mo)**		**Effect PPT (IV) for mean baseline level of symptoms (Mo)**		**Effect PPT (IV) for 1 *SD* above Mean baseline level of symptoms (Mo)**	
In Figure 1 denoted as:			**d – low baseline symptoms**		**d**		**d – high baseline symptoms**	
Independent variable (IV)	**Beta (SE)**	***p* value**	**Beta (SE)**	***p* value**	**Beta (SE)**	***p* value**	**Beta (SE)**	***p* value**
Emotional Dysregulation								
Submissiveness	0.022 (0.011)	.043	−0.028 (0.022)	.068	−0.007 (0.012)	.588	0.016 (0.017)	.357
Cognitive distortion	0.028 (0.010)	.006	0.069 (0.018)	<.001	0.097 (0.014)	<.001	0.125 (0.016)	<.001
Identity problems	0.069 (0.011)	<.001	0.049 (0.016)	.003	0.118 (0.014)	<.001	0.187 (0.019)	<.001
Affective lability	0.066 (0.011)	<.001	0.048 (0.016)	.002	0.114 (0.013)	<.001	0.180 (0.018)	<.001
Oppositionality	0.048 (0.011)	<.001	0.021 (0.016)	.188	0.070 (0.012)	<.001	0.118 (0.016)	<.001
Anxiousness	0.048 (0.011)	<.001	0.015 (0.016)	.324	0.063 (0.013)	<.001	0.110 (0.018)	<.001
Suspiciousness	0.040 (0.010)	<.001	0.059 (0.019)	.002	0.099 (0.014)	<.001	0.139 (0.015)	<.001
Social avoidance	0.060 (0.011)	<.001	0.001 (0.016)	.944	0.061 (0.013)	<.001	0.121 (0.018)	<.001
Narcissism	−0.025 (0.011)	.023	0.016 (0.016)	.338	−0.009 (0.012)	.438	−0.034 (0.016)	.032
Insecure attachment	0.038 (0.011)	<.001	−0.005 (0.017)	.741	0.032 (0.012)	.009	0.070 (0.016)	<.001
Self-harm	0.027 (0.011)	.010	0.056 (0.020)	.004	0.083 (0.013)	<.001	0.111 (0.014)	<.001
Inhibition								
Intimacy problems	0.047 (0.010)	<.001	0.023 (0.016)	.145	0.070 (0.011)	<.001	0.117 (0.014)	<.001
Restricted expression	0.049 (0.011)	<.001	−0.001 (0.016)	.629	0.041 (0.012)	<.001	0.089 (0.017)	<.001
Compulsivity								
Compulsivity	−0.009 (0.011)	.378	−0.010 (0.016)	.546	−0.019 (0.011)	.096	−0.028 (0.015)	.061
Dissocial Behavior								
Stimulus seeking	−0.013 (0.011)	.233	0.055 (0.017)	.001	0.042 (0.012)	<.001	0.029 (0.015)	.051
Callousness	−0.003 (0.011)	.808	0.028 (0.017)	.098	0.026 (0.012)	.033	0.023 (0.015)	.120
Rejection	−0.022 (0.011)	.039	0.006 (0.016)	.691	−0.016 (0.012)	.177	−0.038 (0.015)	.012
Conduct problems	−0.002 (0.011)	.889	0.083 (0.019)	<.001	0.082 (0.013)	<.001	0.080 (0.014)	<.001

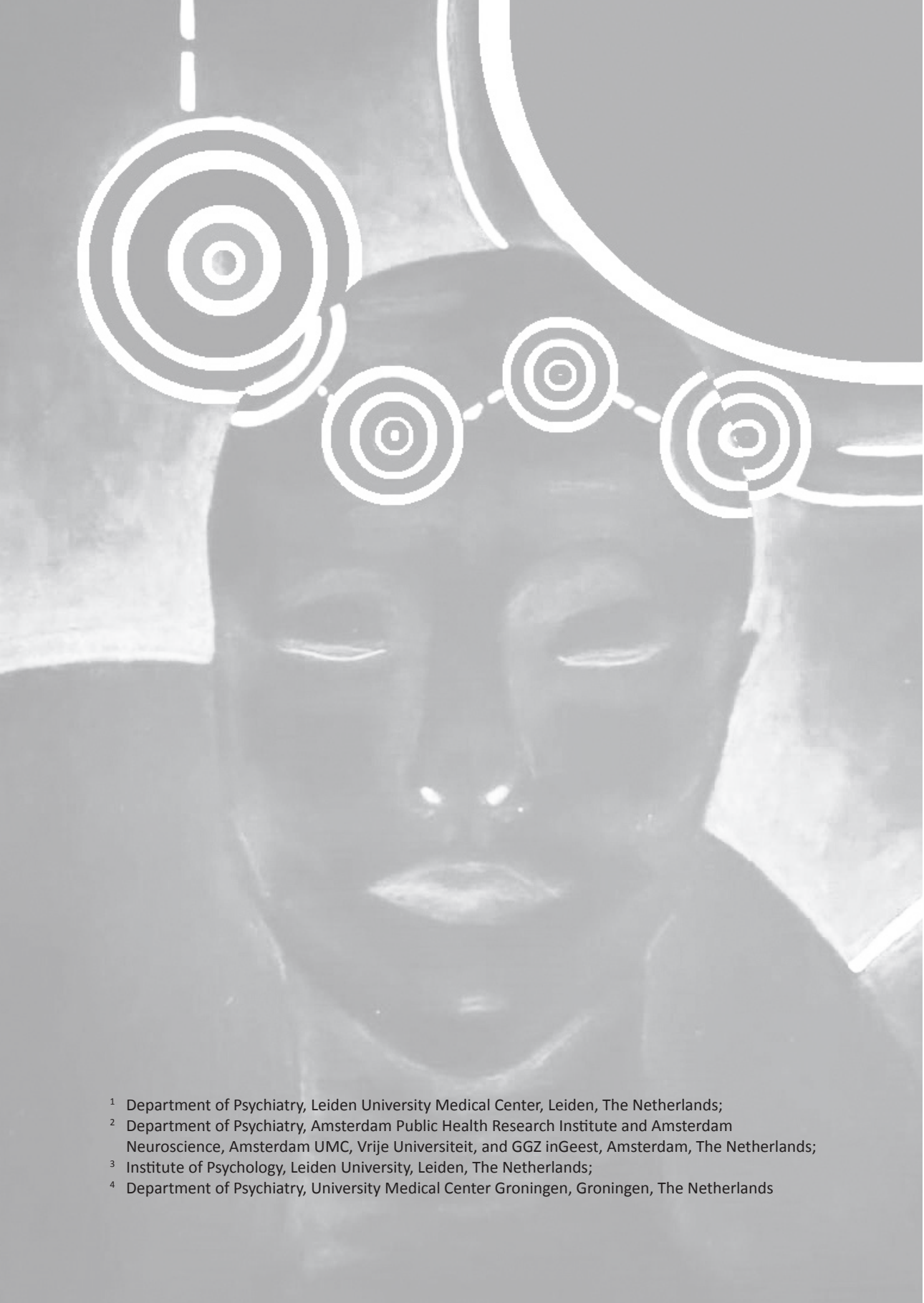
Note. DAPP-SF subscale represents the independent variable (IV). Baseline BSI score represents the moderator variable (Mo). Beta denotes standardized regression coefficients. SE denotes standard error. Analyses are adjusted for age and gender.

Appendix Table 5

Moderating effects of baseline level of symptoms (Mo) when predicting treatment outcome with DAPP-SF higher order constructs of pathological personality traits (PPT), within a depression group, anxiety group, or combined group (see also Figure 1 C)

Treatment Outcome: posttreatment BSI score	Interaction PPT (IV) with Baseline symptom level (Mo)		Effect PPT (IV) for 1 *SD* below Mean baseline level of symptoms (Mo)		Effect PPT (IV) for mean baseline level of symptoms (Mo)		Effect PPT (IV) for 1 *SD* above Mean baseline level of symptoms (Mo)	
In Figure 1 denoted as:			**d – low baseline symptoms**		**d**		**d – high baseline symptoms**	
Independent variable (IV)	**Beta (SE)**	**p value**	**Beta (SE)**	**p value**	**Beta (SE)**	**p value**	**Beta (SE)**	**p value**
Anxiety group (*n* = 1,993)								
Emotional Dysregulation	0.080 (0.016)	<.001	0.161 (0.021)	<.001	0.227 (0.023)	<.001	0.294 (0.031)	<.001
Inhibition	0.078 (0.018)	<.001	0.035 (0.018)	.046	0.100 (0.018)	<.001	0.166 (0.029)	<.001
Compulsivity	0.015 (0.017)	.390	0.004 (0.017)	.830	0.016 (0.017)	.345	0.029 (0.027)	.283
Dissocial Behavior	0.007 (0.018)	.701	0.070 (0.020)	<.001	0.076 (0.019)	<.001	0.082 (0.027)	.003
Depression group (*n* = 1,664)								
Emotional Dysregulation	0.109 (0.019)	<.001	0.022 (0.030)	.468	0.118 (0.058)	<.001	0.215 (0.032)	<.001
Inhibition	0.098 (0.021)	.063	−0.035 (0.028)	.206	0.051 (0.020)	.011	0.138 (0.027)	<.001
Compulsivity	−0.012 (0.021)	.554	−0.012 (0.027)	.654	−0.023 (0.020)	.253	−0.034 (0.027)	.212
Dissocial Behavior	<0.001 (0.020)	.991	0.051 (0.028)	.073	0.051 (0.021)	.016	0.051 (0.026)	.054
Combined depression and anxiety group (*n* = 2,023)								
Emotional Dysregulation	0.049 (0.021)	.021	0.006 (0.042)	.892	0.055 (0.031)	.073	0.104 (0.033)	.001
Inhibition	0.041 (0.020)	.044	0.031 (0.038)	.414	0.072 (0.025)	.004	0.113 (0.025)	<.001
Compulsivity	−0.013 (0.021)	.548	−0.021 (0.039)	.588	−0.034 (0.025)	.177	−0.046 (0.024)	.058
Dissocial Behavior	−0.014 (0.021)	.522	0.032 (0.042)	.449	0.018 (0.027)	.503	0.004 (0.025)	.858

Note. DAPP-SF subscale represents the independent variable (IV). Baseline BSI score represents the moderator variable (Mo). Beta denotes standardized regression coefficients. SE denotes standard error. Analyses are adjusted for age and gender.



[1] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands;
[2] Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit, and GGZ inGeest, Amsterdam, The Netherlands;
[3] Institute of Psychology, Leiden University, Leiden, The Netherlands;
[4] Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands

Chapter 5

Basal and LPS-stimulated inflammatory markers and the course of individual symptoms of depression

van Eeden, W. A.[1], van Hemert, A. M.[1], Carlier, I. V.[1], Penninx, B. W.[2], Lamers, F.[2], Fried, E. I.[3], Schoevers, R.[4], Giltay, E. J.[1]

(2020). Translational psychiatry, 10(1), 1-12

Abstract

Background: Multiple studies show an association between inflammatory markers and major depressive disorder (MDD). People with chronic low-grade inflammation may be at an increased risk of MDD, often in the form of sickness behaviors. We hypothesized that inflammation is predictive of the severity and the course of a subset of MDD symptoms, especially symptoms that overlap with sickness behavior, such as anhedonia, anorexia, low concentration, low energy, loss of libido, psychomotor slowness, irritability, and malaise.

Methods: We tested the association between basal and lipopolysaccharide (LPS)-induced inflammatory markers with individual MDD symptoms (measured using the Inventory of Depressive Symptomatology Self-Report) over a period of up to 9 years using multivariate-adjusted mixed models in 1147 to 2872 Netherlands Study of Depression and Anxiety (NESDA) participants.

Results: At baseline, participants were on average 42.2 years old, 66.5% were women, and 53.9% had a current mood or anxiety disorder. We found that basal and LPS-stimulated inflammatory markers were more strongly associated with sickness behavior symptoms at up to 9-year follow up compared to non-sickness behavior symptoms of depression. However, we also found significant associations with some symptoms that are not typical of sickness behavior (e.g., sympathetic arousal among others).

Conclusions: Inflammation was not related to depression as a unified syndrome but rather to the presence and the course of specific MDD symptoms, of which the majority were related to sickness behavior. Anti-inflammatory strategies should be tested in the subgroup of MDD patients who report depressive symptoms related to sickness behavior.

5.1 Introduction

Inflammatory markers and depression have an intricate and complex relationship [1, 2]. Evidence from meta-analyzes suggests that depressed subjects have higher circulating concentrations of acute-phase proteins and pro-inflammatory cytokines compared to healthy subjects [3-8]. During an inflammatory response, the innate and adaptive immune systems are activated. Pro-inflammatory cytokines are produced by macrophages, monocytes, and other cells that stimulate the liver to produce acute-phase proteins. Chronically increased levels of peripheral blood interleukin (IL-6), tumor necrosis factor-alpha (TNF-α), and C-reactive protein (CRP), all of which indicate low-grade inflammation, are often associated with depression [1]. Other studies, however, have not found significant associations [7, 8].

Another approach to assess inflammation is to stimulating the immune cells and study the clinically important immune disturbances [9, 10]. After ex vivo induction of lipopolysaccharide (LPS: the cell membrane of Gram-negative bacteria that strongly induces immunological responses) in whole blood samples, a wide array of pro-inflammatory cytokines are released, which can be measured in the supernatant [9, 10]. Fewer studies exist on LPS-induced inflammation's putative importance for depression [11-13]. Previous studies have found an association between LPS-stimulated inflammatory markers and depression. Sum scores of the Beck's depression inventory (BDI) were associated with higher levels of inflammatory markers interleukin-1beta (IL-1β), tumor necrosis factor-alpha (TNF-α), and interleukin-8 (IL-8), after LPS induction in whole blood. Additionally, depressed men had higher monocyte chemoattractant protein-1 (MCP-1) levels, and depressed women had higher IL-1α levels [11, 12]. In a previous cross-sectional analysis of the NESDA cohort, higher levels of LPS-induced inflammatory markers were found among patients with a remitted or current depression compared to healthy controls [13]. LPS-induced inflammatory markers were especially elevated among MDD patients with the DSM-5 anxious distress specifier [14]. Results remained statistically significant for LPS induced but not for basal levels of inflammatory markers, after adjusting for lifestyle and somatic health-related covariates ([13].

Researchers have speculated on the existence of crosstalk between several inflammatory pathways and neurocircuits that may lead to sickness behavior [1, 15, 16]. Sickness behavior as a syndrome is still rather ill-defined and has varied across time, disciplines, and studies but

is generally regarded as an organized group of reward oriented behavioral and motivational changes that accompany inflammation and infections [1, 15, 17, 18]. Researchers have theorized that sickness behavior holds some evolutionary advantages and has protective mechanisms for the individual (e.g., recovery), because it preserves energy resources needed for healing infection or other diseases and may help prevent the transmission of its potential infectious agent [1, 18]. The causal chain may involve somatic triggers inducing an inflammatory response followed by sickness behavior. Sickness behavior in turn overlaps with and induces depression, with additional positive feedback loops between (neuro) inflammation and (neuro) degenerative processes [1, 16, 18]. Sickness behavior symptoms show a considerable overlap with depressive symptoms like anhedonia, anorexia, low concentration, low energy, low libido, psychomotor slowness, irritability; and researchers have hypothesized that depression is a maladaptive or exacerbated form of sickness behavior in some patients with chronic low-grade inflammation [15-20]. Besides their reward-sensitivity related symptoms, recent studies suggest that also trauma- and anxiety-related symptoms are related to inflammatory markers, resulting in a mix of overlapping symptoms of mood, anxiety, and post-traumatic stress disorder [13, 21-23]. A causal pathway in which inflammation causes symptoms of anxiety is less established as studies show that inflammatory levels increase when study participants became anxious [24, 25], and a large longitudinal study found that anxiety predicted inflammation in the future but not vice versa [26].

Inflammatory markers and depression have been linked, but effect sizes were generally small [27] with limited clinical relevance for the individual patient (1). Because depression is a heterogeneous disorder with large between-person variation [28] and symptomatology [29, 30], low-grade inflammation may only be strongly linked to a subset of depressive symptoms [31, 32]. Thus, inflammation may be involved in the pathogenesis of a subset of MDD patients. Identifying associations between pro-inflammatory markers and specific depressive symptoms could advance personalized medicine [27]. Nevertheless, few clinical studies have analyzed whether inflammatory markers are associated with specific MDD symptoms [27, 31, 33, 34].

Inflammation has been repeatedly linked to sickness-behavior symptoms such as certain sleeping problems, low energy, changes in appetite, low mood, and cognitive symptoms [27,

31, 33, 34]. Two recent cross-sectional analyzes in the current NESDA cohort found that inflammatory markers demonstrated the strongest associations with sleep and energy level, appetite/weight, and aches and pains, but associations were reduced or disappeared completely when adjusted for demographic-, lifestyle-, and disease-related factors such as BMI, activity, chronic somatic diseases and gender [31, 32]. Adjusting for certain variables is necessary in order to avoid confounding. However, overadjustment must also be avoided as variables such as activity, BMI, and somatic diseases may be part of the causal pathway between low-grade inflammation (which could be induced by somatic disease) on the one hand, and sickness behavior (which includes reduced activity and anorexia) and depression on the other hand [15, 17-20]. There is still no consensus in the field about how to approach these demographic, somatic and lifestyle variables, and studies show that taking these variables into account as either confounders, or as part of the causal pathway, greatly influences the effect size of the relation between inflammation and depression [35]. We are not aware of previous studies that examined the symptom-specific associations with LPS-induced inflammatory agents. Moreover, examining individual symptoms longitudinally is important as inflammation may be related differently to depression symptoms longitudinally [36-40]. A recent longitudinal study for example found that inflammation was especially related to atypical symptoms [40]. Moreover, one meta-analysis demonstrated that increased inflammation can be associated with the development of late-life and the persistence of depression [39]. The present study extends on the current literature as we examined associations between basal levels and LPS-induced inflammatory markers and individual MDD symptoms in a large cohort over the course of 9 years. We hypothesized that persistent low-grade inflammation will show the strongest associations with symptoms characteristic of sickness behavior.

5.2 Methods and Materials

5.2.1 Study sample and procedures

We evaluated baseline and follow-up data from 2872 out of 2981 participants from the NESDA cohort. NESDA investigated the course and consequences of depressive and anxiety disorders. NESDA included patients and healthy controls from a diverse array of (health-care) settings and applied a limited number of exclusion criteria, namely not being fluent in Dutch and the presence of other clinically overt psychiatric disorders (e.g. addiction, psychotic, bipolar). With this method, NESDA aimed for a cohort that is representative for diverse populations of healthy controls and patients with depression and anxiety [41]. The first measurement wave (baseline) ran from 2004 to September 2007; the sixth wave at the 9-year follow up finished in October 2016. All procedures involving human subjects/patients were approved by Ethical Review Board of the VU University Medical Centre and subsequently by local review boards of each participating center. Written informed consent was obtained from all participants. Where verbal consent was obtained this must be followed by a statement such as: Verbal consent was witnessed and formally recorded. More detailed design and sampling procedures are published elsewhere [41].Basal serum levels of inflammatory markers were collected from 2867 participants. For logistical reasons, LPS induction in blood was only assessed during the last year of baseline sample collection. Consequently, data of LPS-stimulated inflammatory markers were available from 1229 out of 2867 participants. Of all the demographics and clinical characteristics mentioned in Table 1, this sub-selection did not differ from participants with missing data ($p > 0.05$), with the exception of age because the LPS subgroup was on average 1 year older. About 40% of the sample had a chronic somatic disease. A wide variety of diseases were assessed through a self-report questionnaire, asking for the presence of 20 common chronic diseases including asthma, chronic bronchitis or pulmonary emphysema, heart diseases or infarct, diabetes, stroke or CVA, arthritis or arthrosis, rheumatic complaints, tumor and/or metastasis, stomach or intestinal disorders, liver disease or liver cirrhosis, epilepsy, thyroid gland disease, or another chronic disease for which the patient receives treatment. A count was made of the number of chronic diseases for which a person reported receiving treatment. More details regarding this variable can be found elsewhere [42].

5.2.2 Measures

5.2.2.1 Demographics and clinical features

The Composite International Diagnostic Interview (CIDI WHO, version 2.1) was used to assess the presence of depressive and anxiety disorders according to the DSM-IV. The CIDI is a fully standardized diagnostic interview with validated psychometric characteristics [41, 43].

Demographic variables were described and included gender, age, ethnicity (yes/no regarding Northern European heritage), and level of education (elementary or less; general intermediate/secondary education; college/university). Patients also indicated whether they had a fever or cold in the week prior to blood draw (sickness prior to interview).

Medication use was determined by inspecting participants' medication containers. Antidepressant use included selective serotonin reuptake inhibitors (SSRIs; ATC code: N06AB), tricyclic antidepressants (TCAs; ATC code: N06AA), and other antidepressants (ATC codes: N06AF, N06AG, N06AX). The use of statins (ATC code: C10AA) and anti-inflammatory, anti-rheumatic, and anti-allergic medications (ATC codes: M01A, M01B, A07EB, A07EC) was also assessed (further referred to as anti-inflammatory medication).

5.2.2.2 Independent variables: inflammatory markers

Baseline inflammatory markers CRP, IL-6, and TNF-α were assessed using fasting blood plasma levels (see the supplementary information (SI)). Intra- and inter-assay coefficients of variation for CRP levels were 5% and 10%, respectively. Intra- and inter-assay coefficients of variation for IL-6 levels were 8% and 12%, respectively. Intra- and inter-assay coefficients of variation for TNF-α levels were 10% and 15%, respectively.

Inflammation is likely to occur when multiple cytokines are elevated. We did not form specific hypotheses about individual inflammatory markers, so we created a basal inflammation index, representing the mean value of $\log_e$-transformed (due to non-normality) and standardized levels of CRP, IL-6, and TNF-α [13].

5.2.2.3 Independent variables: inflammatory markers after LPS induction

The innate immune response of 12 cytokines was assessed in ex vivo stimulated blood using LPS (see the SI). For all available samples, we simultaneously assessed levels of interferon-γ (IFN-γ), macrophage inflammatory protein-α (MIP-1α), IL-2, IL-6, IL-8, IL-10, IL-18, MCP-1,

macrophage inflammatory protein-α (MIP-1α), MIP-1β, matrix metallopeptidase-2 (MMP-2), TNF-α, and TNF-β using a multi-analytic profile (Human CytokineMAP A v.1.0; Myriad RBM, Austin, TX, USA). Cytokine distributions were skewed to the right and therefore $\log_e$-transformed to normalize their distributions.

We created an LPS-induced inflammation index composed from the mean standardized value of all available LPS-induced markers, further referred to as LPS-induced inflammation index. To avoid loss of information, we conducted an exploratory factor analysis [EFA; 44], which resulted into two LPS-induced inflammation indexes, further referred to as LPS-induced inflammation index-1 and LPS-induced inflammation index-2. Markers IFN-γ, IL-10, IL-2, IL-6, MMP-2, TNF-α, and TNF-β loaded on LPS-induced inflammation index-1 with factor loadings between 0.41 and 0.88 and a raw alpha of 0.86. IL-8, IL-18, MCP-1, MIP-1α, and MIP-1β loaded on LPS-induced inflammation index-2. See SI for the correlations between individual markers within each index (SI Figure 1) and a more detailed description of the EFA procedures. Subsequently, two LPS-induced inflammation indexes were calculated as the mean of $\log_e$-transformed and standardized markers.

5.2.2.4 Dependent variables: IDS items

The sum score of the Inventory of Depressive Symptomatology Self-Report (IDS-SR) was used as the outcome measure for severity and course of depression on syndrome level, and the separate items were used for the symptom analyzes [45, 46]. The IDS-SR consists of 30 equally weighted items, rated on a 4-point Likert scale (0–3), and includes all symptoms of depression: melancholic, atypical, and anxious symptoms. Several additional symptoms were included: sympathetic arousal, pessimism, and interest in sex. We hypothesized that the following 16 IDS-SR items would be associated with inflammation at baseline because they can identify sickness-behavior symptoms [15, 17-20]: sleeping too much (Item 4), feeling irritable (Item 6), responsiveness of mood (Item 8), decrease in appetite (Item 11), decrease in weight (Item 12), concentration (Item 15), pessimism (Item 17), general interest (Item 19), low energy level (Item 20), capacity for pleasure (Item 21), interest in sex (Item 22), psychomotor retardation (Item 23), aches and pains (Item 25), sympathetic arousal (Item 26), constipation or diarrhea (Item 28), and leaden paralysis (Item 30).

5.2.3 Statistical analysis

We used a multivariate linear mixed model with IDS-SR item scores as outcome variables and inflammatory markers as the main independent variables. Because of the heterogeneity of our sample (healthy and depressed participants at baseline), the intercepts and slopes were considered as random variables, which resulted in a significantly better fit compared to a non-random model. (For the model with the basal inflammation index, the log likelihood (LL)-ratio increased by 80932.5, $p < 0.001$; for LPS-induced inflammation index-1, LL-ratio increased by 36887.2, $p < 0.001$; and for LPS-induced inflammation index-2, LL-ratio increased by 38640.1, $p < 0.001$.) Adding an interaction between time and inflammatory markers resulted a minimal increase of model fit. (For the model with the basal inflammation index, the LL-ratio increased by 12.2, $p < 0.001$; for LPS-induced inflammation index-1, the LL-ratio increased by 1.9, $p = 0.167$; and for LPS-induced inflammation index-2, the LL-ratio increased by 12.5, $p < 0.001$.) This small effect could be attributed to regression to the mean, so we decided not to include the interaction terms in our final models. Doing so resulted in mixed models for each individual IDS item with random intercepts and slopes over time, that analyzed whether participants with elevated levels of inflammation were more likely to have higher symptom levels at baseline and during the 9-year follow-up period. Models were adjusted for certain baseline variables: gender, age, sickness prior to interview, and the use of anti-inflammatory medication. In sensitivity analyzes, we repeated the analysis for MDD patients (~30% of the total sample; SI table 1 and SI figure 2) and for the LPS-inflammatory composite index score (SI table 2 and SI figure 3). Moreover, sensitivity analyzes were executed which additionally adjusted for chronic somatic diseases and antidepressants (SI figure 4). Subsequently, we adjusted the outcomes of the inflammation indexes for multiple testing using the Benjamin–Hochberg procedure [47]. Means of subscale scores (i.e., sickness behavior vs non-sickness behavior) were computed and presented in line graphs for the effects over time. In order to yield beta coefficients that can be compared among symptoms, all outcome and independent variables were standardized (i.e., *z* scores) with two-sided p-values. All models were run in R, version 3.4.3.

5.3 Results

5.3.1 Sociodemographic and clinical characteristics at baseline

Our study sample was 66.7% female (*n* = 1975), and the ages ranged from 18 to 64 years at baseline (mean 42.9 years, *SD* 13.1; see Table 1 for demographics). The sample consisted of 35.4% one-month recency MDD patients (*n* = 796), 2.8% with minor depression (*n* = 84), 9.3% with dysthymia (*n* = 277), 43.6% with a (comorbid) anxiety disorder (*n* = 1299), and 46.1% without a mood or anxiety diagnosis at baseline (*n* = 1368), of whom 54.2% never had a psychiatric diagnosis (*n* = 742).

Table 1. Sociodemographic and clinical characteristics

	Whole sample n = 2872	**LPS-induced subsample n = 1229**
Age in years (mean, SD)	41.9 (13.0)	42.8 (12.7)
Female (%)	66.5	65.6
North-european etnicity (%)	94.9	94.8
BMI (mean, SD)	25.6 (5.0)	25.7 (5.0)
smoking status (%)		
Never smoker	28.0	29.0
Former smoker	33.6	34.2
Current smoker	38.4	36.8
Education level (%)		
Elementary or lower	6.49	6.4
Secondary education	58.2	56.7
College or university	35.4	36.9
Sickness prior to interview (%)	27.9	30.1
Chronic somatic disease, yes (%)	40.4	44.3
Anti-inflam. med., yes (%)	4.9	3.1
MDD, yes (%)	35.4	28.8
Minor depression, yes (%)	2.8	2.1
Dysthymia, yes (%)	9.3	10.4
Anxiety disorder, yes (%)	43.6	44.4
No Disorder (%)	46.1	46.3
No lifetime disorder (%)	34.1	36.3
Total score IDS at baseline (SD)	21.184 (14.6)	20.86 (14.6)
Antidepressants		
TCA (%)	3.7	2.9
SSRI (%)	16.8	16.5
Other (%)	5.5	5.6
no AD (%)	75.5	75.9
Inflammattory markers (mean, sd)		
TNF-α (pg/ml)	1.09 (1.41)	
IL-6 (pg/ml)	1.55 (13.5)	
CRP (mg/L)	2.82 (5.12)	
Inflammattory markers after LPS induction (mean, sd)		
IFN-γ (pg/ml)		12.80 (10.8)
IL-10 (pg/ml)		300.28 (294.9)
IL-18 (pg/ml)		262.39 (91.9)
IL-2 (pg/ml)		10.06 (5.0)
IL-6 (ng/ml)		27.36 (15.6)
IL-8 (ng/ml)		12.02 (7.7)
MCP-1 (ng/ml)		1.72 (1.1)
MIP-1α (ng/ml)		19.38 (12.0)
MIP-1β (ng/ml)		245.52 (123.3)
MMP-2 (pg/ml)		72.13 (19.3)
TNF-α (ng/ml)		3.19 (2.0)
TNF-β (pg/ml)		324.21 (126.6)

Table 1. Demographic and clinical sample characteristics. BMI = body mass index. MDD = major depressive disorder. TCA = tricyclic antidepressants. SSRI = selective serotonin reuptake inhibitors. AD = antidepressants. Tumor necrosis factor = TNF (median). Interleukin = IL. C-reactive protein = CRP. Interferon-γ = IFN-γ. Higher monocyte chemoattractant protein-1 = MCP-1. Macrophage inflammatory protein = MIP. Matrix metallopeptidase-2 = MMP-2.

5.3.2 Basal inflammation

We found a small but significant association between the basal inflammatory index and IDS-scores adjusted for age, gender, and anti-inflammatory medication (β = 0.039; $p < 0.001$). Thus, participants with a higher inflammatory index tended to have a 0.039 SD higher IDS-30 score over the course of 9 years, compared to participants with a 1 *SD* lower inflammatory index. This comes down to a absolute value of 1.12 IDS-SR sum score.

Next, we analyzed the associations between the basal inflammation index for each of the 30 IDS items. Table 2 and Figure 1 present the standardized beta coefficients of the basal inflammation index adjusted for age, gender, sickness prior to interview, and anti-inflammatory medication. All individual symptoms were positively related to high levels of basal inflammation. The beta sizes ranged from 0.005 (Item 2: Sleep during the night) to 0.085 (Item 25: Aches and pains). The course of tertiles of mean scores of sickness behavior symptoms versus non-sickness behavior symptoms is presented in Figure 2. As expected, both sub-scores declined steeply after baseline due to regression to the mean effects of anxiety and MDD patients who were initially selected for the NESDA cohort. Symptoms related to sickness behavior more strongly associated with basal inflammatory markers than other symptoms, the mean scores of which remained relatively elevated during the 9 years. Beta coefficients were statistically significant for quality of mood (Item 10; β = 0.028, $p = 0.049$) and all other items with beta coefficients above 0.028 (see Figure 1). After adjusting for multiple testing for all tests summarized in Table 2, *p* values remained statistically significant for 17 items. Of the symptoms related to sickness behavior, 14 out of 16 symptoms were significantly associated with inflammation, compared to six out of 14 non-sickness-behavior symptoms. We found similar results with MDD patients only ($n = 908$), albeit with overall weaker effects due to lower variance and a smaller sample size (see SI Table 1 and SI Figure 2). Among patients with MDD at baseline, eight out of 16 sickness-behavior-related symptoms were significantly associated with inflammation compared to three out of 14 non-sickness-behavior-related symptoms.

Table 2A. Basal serum inflammatory markers in relation to IDS symptoms over the course of nine years

	Basal Serum inflammation index		
	CRP, TNF-α, IL-6		
Item	Beta (SE)		p-value
1. Falling asleep	0.025	(0.015)	0.096
2. Sleep during the night	0.005	(0.014)	0.723
3. Waking up too early	0.015	(0.014)	0.270
4. Sleeping too much	0.053	(0.014)	<0.001*
5. Feeling Sad	0.033	(0.014)	0.022*
6. Feeling irritable	0.016	(0.014)	0.252
7. Anxious or tense	0.018	(0.014)	0.213
8. Response of mood	0.038	(0.013)	0.004*
9a. Mood in time of day	0.012	(0.013)	0.361
10. Quality of mood	0.028	(0.014)	0.049*
11. Decreased appetite	0.039	(0.012)	0.001*
12. Increased appetite	0.050	(0.013)	<0.001*
13. Decreased weight	0.041	(0.010)	<0.001*
14. Increased weight	0.031	(0.011)	0.006*
15. Concentration	0.025	(0.014)	0.071
16. View of myself	0.034	(0.014)	0.018*
17. View of my future	0.051	(0.014)	<0.001*
18. Death or suicide	0.034	(0.014)	0.017*
19. General interest	0.057	(0.014)	<0.001*
20. Energy level	0.076	(0.014)	<0.001*
21. Capacity for pleasure	0.057	(0.014)	<0.001*
22. Interest in sex	0.053	(0.014)	<0.001*
23. Psychomotor retardation	0.061	(0.014)	<0.001*
24. Psychomotor agitation	0.018	(0.014)	0.220
25. Aches and pains	0.085	(0.014)	<0.001*
26. Sympathetic arousal	0.055	(0.014)	<0.001*
27. Panic/Phobic	0.016	(0.015)	0.288
28. Constipation/diarrhea	0.041	(0.014)	0.003*
29. Interpersonal sensitivity	0.006	(0.014)	0.683
30. Leaden paralysis	0.072	(0.014)	<0.001*

Table 2A. Standardized beta coefficients of the association between basal serum inflammatory markers and individual depressive symptoms. Linear mixed models fitted with repeated measures, using standardized IDS-SR item-scores as outcome variables, which were assessed up to six times over 9 years of follow up. Standardized beta coefficients were adjusted for gender, age, sickness prior to interview, and the use of anti-inflammatory medication.

**P* values that remained significant (< 0.05) after correcting for multiple testing using the Benjamin–Hochberg procedure.

5.3.3 LPS-induced inflammation

The overall LPS-induced inflammation index (β = 0.036; p = 0.014) and the LPS-induced inflammation index-2 (β = 0.056, p < 0.001) were significantly related to the IDS score averaged over 30 items, and the LPS-induced inflammation index-1 indicated a relationship that approached significance (β = 0.026; p = 0.072). In absolute values this would translates in IDS-SR sum-scores difference of 1.12, 0.82, 1.71 for each SD increase of the LPS-induced inflammation index, LPS-induced inflammation index-1, and LPS-induced inflammation index-2 respectively.

The LPS-induced inflammation index-2 more strongly related to sickness-behavior symptoms, compared to non-sickness-behavior symptoms, than LPS-induced inflammation index-1, the beta coefficients of which ranged from −0.005 (mood related to time of the day) to 0.049 (feeling irritable) and were statistically significant for feeling irritable (Item 6; β = 0.049, *p* = 0.035) and panic/phobia (Item 27; β = 0.056, *p* = 0.018). After adjusting for multiple testing, only panic/phobia remained statistically significant.

Regarding LPS-induced inflammation index-2, beta coefficients ranged from −0.004 (waking up too early) to 0.105 (aches and pains). Betas were statistically significant for 13 out of 16 sickness-behavior symptoms and for six out of 14 non-sickness-behavior symptoms, with significant betas for decreased weight (Item 13; β = 0.050, *p* = 0.002) and all other items with betas greater than 0.050 (see Figure 1). Sickness-behavior symptoms remained elevated over the 9 years (Figure 2). After adjusting for multiple testing, *p* values remained significant for 19 items.

In a sensitivity analysis, we analyzed the association of the composite LPS-induced inflammation index for all LPS-induced markers. Only seven out of 30 symptoms indicated significant associations (see SI Table 2 and SI Figure 3). However, findings were no longer statistically significant after we adjusted for multiple testing. The LPS-induced inflammation index was equally related to sickness- and non-sickness-behavior symptoms.

Table 2B. LPS-induced inflammatory markers in realtion to IDS symptoms over the course of nine years

	LPS-induced index inflammation factor 1			LPS-induced index inflammation factor 2		
	IL-2, IL-6, IL-10, MMP-2, TNF-α, TNF-β, IFN-y			IL-8, IL-18, MCP-1, MIP-1α, MIP-1β		
Item	Beta	(SE)	p-value	Beta	(SE)	p-value
1. Falling asleep	0.018	(0.024)	0.445	0.013	(0.024)	0.570
2. Sleep during the night	0.032	(0.022)	0.132	0.017	(0.022)	0.429
3. Waking up too early	-0.003	(0.023)	0.879	-0.004	(0.023)	0.853
4. Sleeping too much	0.010	(0.023)	0.653	0.008	(0.023)	0.720
5. Feeling Sad	0.018	(0.024)	0.466	0.028	(0.024)	0.247
6. Feeling irritable	0.049	(0.023)	0.035*	0.075	(0.023)	0.001*
7. Anxious or tense	0.045	(0.023)	0.055	0.064	(0.023)	0.007*
8. Response of mood	0.030	(0.021)	0.159	0.071	(0.021)	0.001*
9a. Mood in time of day	-0.005	(0.021)	0.820	0.032	(0.021)	0.132
10. Quality of mood	0.039	(0.023)	0.093	0.071	(0.023)	0.002*
11. Decreased appetite	0.017	(0.019)	0.359	0.071	(0.018)	<0.001*
12. Increased appetite	-0.004	(0.021)	0.852	0.012	(0.021)	0.567
13. Decreased weight	0.003	(0.016)	0.854	0.050	(0.016)	0.002*
14. Increased weight	0.013	(0.018)	0.469	0.030	(0.018)	0.088
15. Concentration	0.024	(0.023)	0.290	0.071	(0.023)	0.002*
16. View of myself	0.006	(0.024)	0.799	0.080	(0.023)	0.001*
17. View of my future	0.042	(0.024)	0.081	0.072	(0.024)	0.003*
18. Death or suicide	0.041	(0.023)	0.082	0.040	(0.023)	0.082
19. General interest	0.026	(0.022)	0.242	0.068	(0.022)	0.002*
20. Energy level	0.027	(0.023)	0.233	0.085	(0.022)	<0.001*
21. Capacity for pleasure	0.014	(0.023)	0.523	0.070	(0.022)	0.002
22. Interest in sex	0.000	(0.022)	0.987	0.040	(0.022)	0.070
23. Psychomotor retardation	0.013	(0.023)	0.561	0.068	(0.022)	0.003*
24. Psychomotor agitation	0.011	(0.023)	0.650	0.065	(0.023)	0.005*
25. Aches and pains	0.045	(0.023)	0.052	0.105	(0.023)	<0.001*
26. Sympathetic arousal	0.025	(0.023)	0.260	0.073	(0.022)	0.001*
27. Panic/Phobic	0.056	(0.024)	0.018*	0.068	(0.024)	0.004*
28. Constipation/diarrhea	0.039	(0.022)	0.085	0.024	(0.022)	0.282
29. Interpersonal sensitivity	0.023	(0.024)	0.332	0.061	(0.023)	0.009*
30. Leaden paralysis	0.021	(0.024)	0.369	0.077	(0.023)	0.001*

Table 2B. Standardized beta coefficients of the association between LPS-induced inflammatory markers and individual depressive symptoms. Linear mixed models fitted with repeated measures, using standardized IDS-SR item-scores as outcome variables, which were assessed up to six times over 9 years of follow up. Standardized beta coefficients were adjusted for gender, age, sickness prior to interview, and the use of anti-inflammatory medication.

*P values that remained significant (< 0.05) after correcting for multiple testing using the Benjamin–Hochberg procedure.

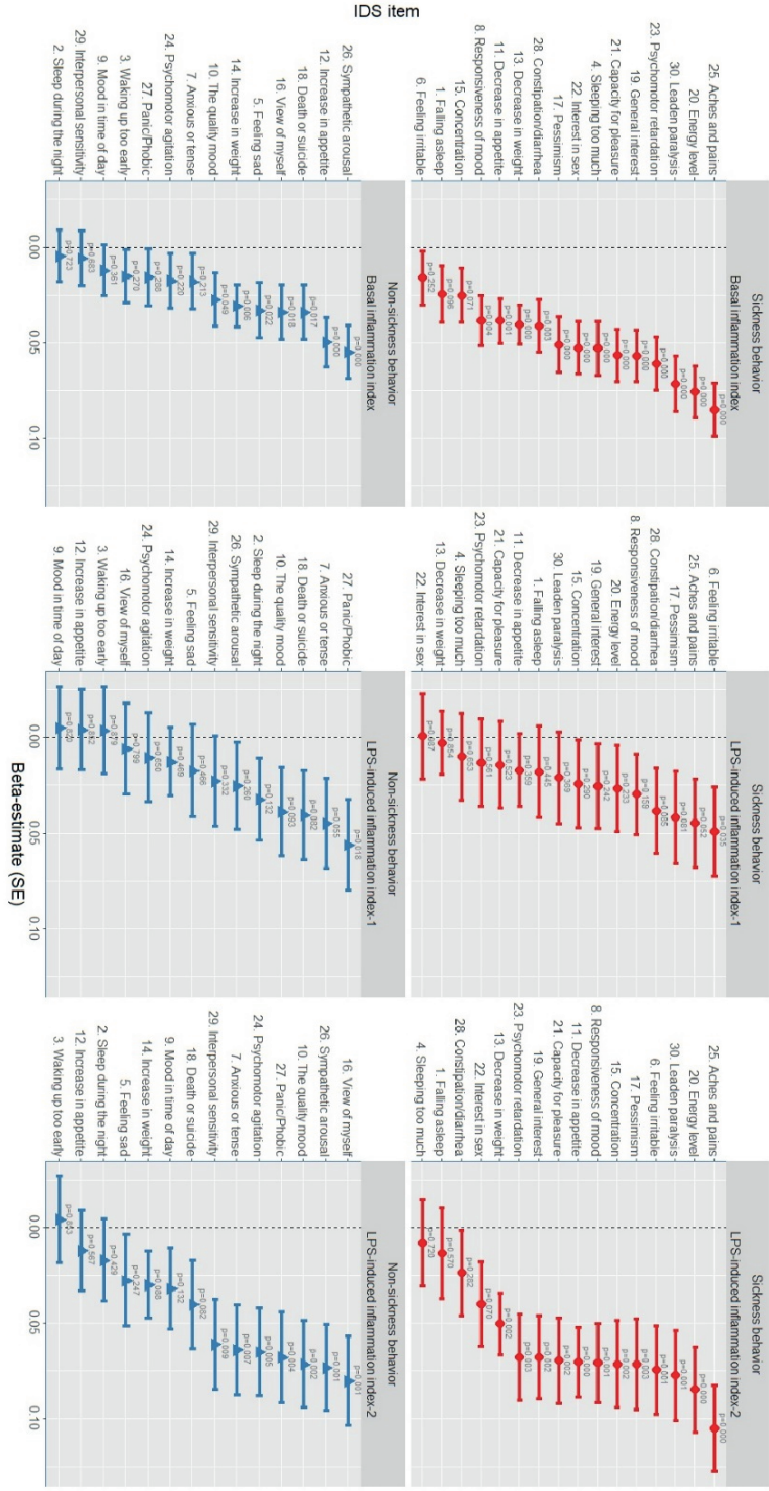

Sickness behavior
Basal inflammation index
25. Aches and pains
20. Energy level
30. Leaden paralysis
23. Psychomotor retardation
19. General interest
21. Capacity for pleasure
4. Sleeping too much
22. Interest in sex
17. Pessimism
28. Constipation/diarrhea
13. Decrease in weight
11. Decrease in appetite
8. Responsiveness of mood
15. Concentration
1. Falling asleep
6. Feeling irritable
Non-sickness behavior
Basal inflammation index
26. Sympathetic arousal
12. Increase in appetite
18. Death or suicide
16. View of myself
5. Feeling sad
14. Increase in weight
10. The quality mood
7. Anxious or tense
24. Psychomotor agitation
27. Panic/Phobic
3. Waking up too early
9. Mood in time of day
29. Interpersonal sensitivity
2. Sleep during the night
Sickness behavior
LPS-induced inflammation index-1
6. Feeling irritable
25. Aches and pains
17. Pessimism
28. Constipation/diarrhea
8. Responsiveness of mood
20. Energy level
19. General interest
15. Concentration
30. Leaden paralysis
1. Falling asleep
11. Decrease in appetite
21. Capacity for pleasure
23. Psychomotor retardation
4. Sleeping too much
13. Decrease in weight
22. Interest in sex
Non-sickness behavior
LPS-induced inflammation index-1
27. Panic/Phobic
7. Anxious or tense
18. Death or suicide
10. The quality mood
2. Sleep during the night
26. Sympathetic arousal
29. Interpersonal sensitivity
5. Feeling sad
14. Increase in weight
24. Psychomotor agitation
16. View of myself
3. Waking up too early
12. Increase in appetite
9. Mood in time of day
Sickness behavior
LPS-induced inflammation index-2
25. Aches and pains
20. Energy level
30. Leaden paralysis
6. Feeling irritable
17. Pessimism
15. Concentration
8. Responsiveness of mood
11. Decrease in appetite
21. Capacity for pleasure
19. General interest
23. Psychomotor retardation
13. Decrease in weight
22. Interest in sex
28. Constipation/diarrhea
1. Falling asleep
4. Sleeping too much
Non-sickness behavior
LPS-induced inflammation index-2
16. View of myself
26. Sympathetic arousal
10. The quality mood
27. Panic/Phobic
24. Psychomotor agitation
7. Anxious or tense
29. Interpersonal sensitivity
18. Death or suicide
9. Mood in time of day
14. Increase in weight
5. Feeling sad
2. Sleep during the night
12. Increase in appetite
3. Waking up too early
IDS item
0.00
0.05
0.10
Beta-estimate (SE)

Figure 1. Associations of the basal inflammation index (n = 2872), LPS-induced inflammation index-1 (n = 1147), and LPS-induced inflammation index-2 (n = 1229) with individual depressive symptoms during 9 years. Standardized beta coefficients with error bars representing standard errors of the predictive values of inflammatory indexes in relation to individual depressive symptoms over 9 years of follow up. The red dots represent depressive symptoms that are assumed to be related to sickness behavior. The blue dots represent depressive symptoms that are not related to sickness behavior. Beta coefficients translates a "the amount of SD that that particular symptom is elevated averaged over nine years, for each increased SD of inflammatory marker". Assessments conducted using linear mixed models with repeated measures, adjusting for gender, age, use of anti-inflammatory drugs, and sickness prior to interview.

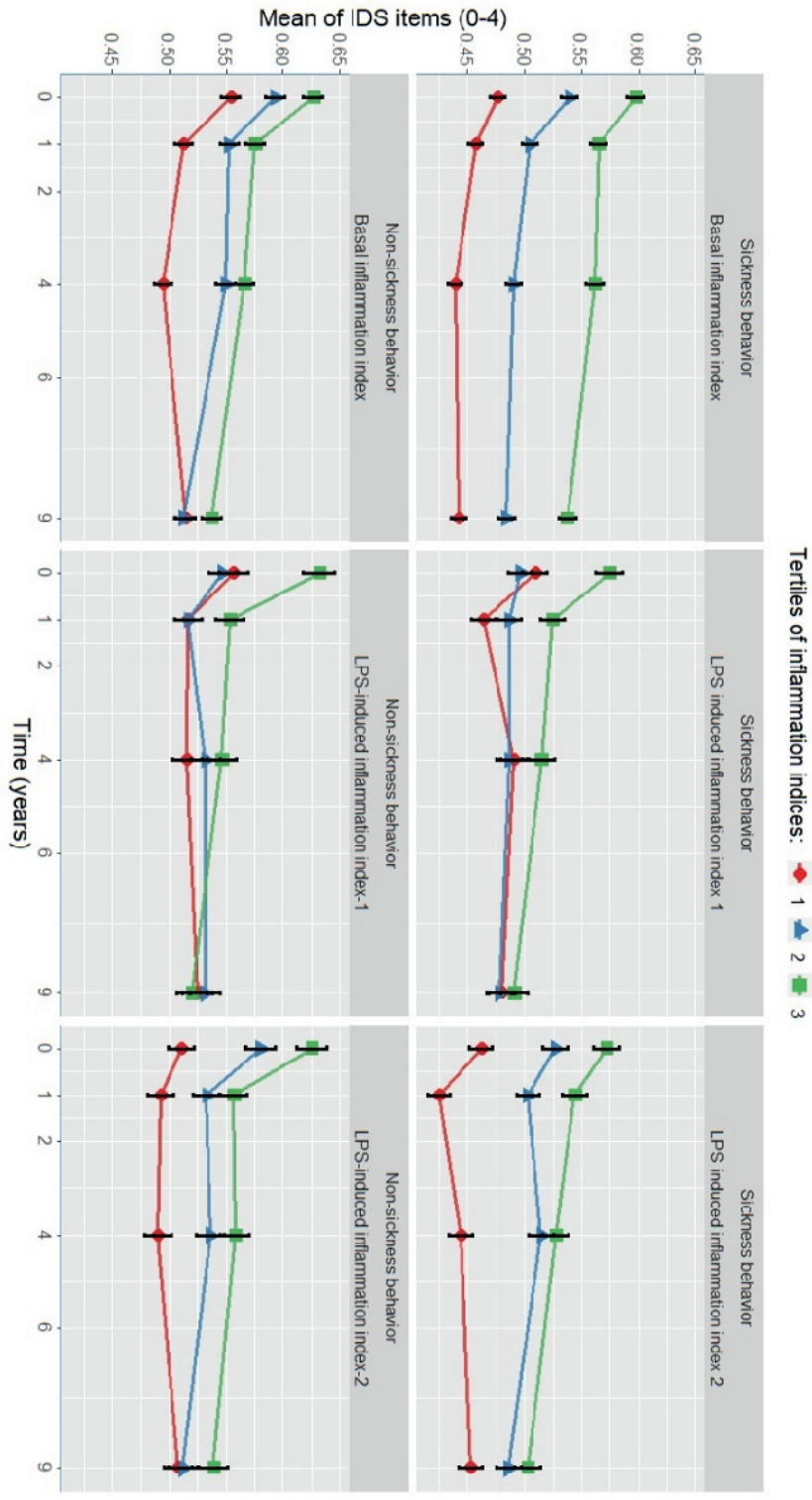

Tertiles of inflammation indices: 1 2 3
Mean of IDS items (0-4)
0.45
0.50
0.55
0.60
0.65
Time (years)
0
1
2
4
6
9
Sickness behavior
Basal inflammation index
Non-sickness behavior
Basal inflammation index
Sickness behavior
LPS induced inflammation index 1
Non-sickness behavior
LPS-induced inflammation index-1
Sickness behavior
LPS induced inflammation index 2
Non-sickness behavior
LPS-induced inflammation index-2

Figure 2. Tertiles of the basal inflammation index, LPS-induced inflammation index-1, and LPS-induced inflammation index-2 related to IDS-SR item scores of sickness-behavior symptoms and non-sickness-behavior symptoms over the course of 9 years. Inflammation indexes are divided into tertiles of equal proportions of the sample distribution (1. lowest inflammatory markers: 0.0 – 0.33; 2. middle: 0.33 – 0.66; 3. highest: 0.66 – 1.0). Y-axis represent absolute mean values of IDS-SR item-scores (0 - 3). Error bars representing standard errors. IDS items related to sickness behavior: sleeping too much (Item 4), feeling irritable (Item 6), responsiveness of mood (Item 8), decrease in appetite (Item 11), decrease in weight (Item 12), concentration (Item 15), pessimism (Item 17), general interest (Item 19), low energy level (Item 20), capacity for pleasure (Item 21), interest in sex (Item 22), psychomotor retardation (Item 23), aches and pains (Item 25), constipation or diarrhea (Item 28) and leaden paralysis (Item 30). Non-sickness behavior IDS items: falling asleep (Item 1), sleep during the night (Item 2), waking up too early (Item 3), feeling sad (Item 5), anxious or tense (Item 7), mood in time of day (Item 9a), quality of mood (Item 10), increased appetite (Item 12), increased weight (Item 14), view of myself (Item 16), death or suicide (Item 18), psychomotor agitation (Item 24), sympathetic arousal (Item 26), panic/phobic (Item 27), and interpersonal sensitivity (Item 29).

5.4 Discussion

We aimed to examine whether diverse inflammatory markers could predict the trajectories of individual symptoms of depression over the course of 9 years, specifically looking at symptoms indicative of sickness behavior. We found that the basal inflammation index and the LPS-induced inflammation index-2 predicted many depressive symptoms over the course of 9 years. By conducting regression analysis for each individual symptom separately, we demonstrated that significant associations between inflammatory markers and the course of a particular individual symptom was more than twice as likely to be significant when that symptom was related to sickness behavior compared to non-sickness-related behavior. The sickness-behavior theory may explain the rather weak (or sometimes conflicting) relationships found between low-grade inflammation and MDD [18].

Four previous studies, three with cross-sectional [31-33] and one with a prospective design [34], have examined symptom-specific associations between basal serum inflammatory markers and depression. One study found that inflammation was specifically related to a change in appetite, poor sleep, and low energy [33]. Two of the cross-sectional studies were conducted within the current NESDA cohort and demonstrated that symptoms of sleeping problems, energy levels, appetite/weight changes, aches and pains and irritability were most likely to be positively associated with basal inflammatory markers [31, 32]. By using network analyzes, it was further demonstrated that the relation between basal inflammatory markers mostly runs through, and was affected by, lifestyle and disease-related covariates, such as BMI, activity level, and chronic somatic diseases [32]. Our study differed from these analyzes because we used index scores instead of individual inflammatory markers. Moreover, as recommended for future research directions [32, 48], the individual symptoms were measured longitudinally at six time points over the course of 9 years. We adjusted for two disease related variables (sickness prior to intake, and anti-inflammatory markers). Moreover, in a sensitivity analysis, we additionally adjusted for the count of self-report chronic somatic diseases and the use of antidepressants, which yielded a small attenuation of our results, but did not lead to different conclusions (SI figure 4). Our findings are largely consistent with previous findings; signs of low-grade inflammation at baseline were associated with the long-term symptomatology of sickness behavior [18], and elevated levels of inflammation could

lead to sickness behavior, which may explain some of the symptoms in certain cases of MDD [49-51]. However, we also found significant associations with symptoms that are not typical of sickness behavior (e.g., anxiety and low self-esteem). It is likely that much of the associations we found runs through lifestyle and disease related variables, as these factors are thought to be part of the causal pathway [16, 32, 52]. It is hypothesized that (chronic) somatic factors results in higher levels of inflammatory markers, which in its turn results in sickness behavior (including lifestyle factors such as lower activity) which is related to, and is part of the depressive symptomatology [16, 32, 52]. Another line of thought is that these somatic and lifestyle factors act as confounding variables as they are both related to inflammation and depression [52]. The fact that we found the strongest association to symptoms that are specifically related to sickness behavior over the course of nine years, suggests however that the sickness behavior theory is probable [16, 53].

To our knowledge, this is the first study that examined LPS-induced inflammatory markers in relation to the course of individual depressive symptoms. These markers reflect the cytokine production capacity when triggered by endogenous or exogenous triggers [9, 54], and are thought to be less affected by health and lifestyle factors such as BMI and chronic somatic diseases [13]. We found strong associations between LPS induced inflammation index-2 markers and depressive symptoms. However, LPS induced inflammation index-1 did not demonstrate such results. When looking at individual symptoms, LPS-induced, but not basal levels seem to be more specifically associated to symptoms of anxiety. Although this was not the focus of the current study, these findings are in line with the idea that anxiety-related symptoms may induce an inflammatory response [13, 14, 21]. Future research may focus on the potential role of LPS-induced markers in relation to the longitudinal course of anxiety related symptoms.

Cytokines contribute to many aspects of human biology and have evolved to enable the sensing and interpretation of environmental cues relevant to maintaining a healthy physiology [55]. Although these secretory (glycol)proteins are best known for their role as custodians of immune homeostasis and the inflammatory response to infection, trauma, or injury, this study confirms their additional effects on mood and behavior [56]. Cytokines often display heterogenetic, pleiotropic, and overlapping functional properties [57]. Although cytokines are considered to be a “family,” this is a functional (rather than structural) concept.

A common factor of the markers clustered in the LPS-induced inflammation index-2 is the link with T lymphocyte cells (T cells) and natural killer cells (NK cells). MCP-1, MIP-1α, and MIP-1β have a signaling function for monocytes and regulate T-cell activity. MIP-1β has an additional specificity for NK cells. IL-8 and IL-18 induce certain T-cell and NK-cell functions such as chemotaxis [58, 59] and locomotion [60, 61]. There are indications that some MDD patients have impaired neuroprotective and anti-inflammatory T-cell responses [1]. Also, researchers have found a reduced number of circulating NK cells for MDD patients compared to healthy controls [1, 62].

Depression is a heterogeneous syndrome with a substantial variety of symptoms among patients with symptom-specific risk factors [63]. Not all patients exhibit symptoms related to sickness behavior, and only one third of MDD patients exhibit elevated inflammatory markers [64]. Our findings could have implications for anti-inflammatory treatment [6, 65] and preventative care [66-69] in a subgroup of depressed patients with sickness-behavior-related symptoms [70]. Research is underway to investigate the effects of anti-TNF-alpha biologic infliximab on measures of anhedonia, motivational behavior and glutamatergic changes in the basal ganglia [71] and to investigate the effects of simvastatin for treatment-resistant MDD [72] and patients with comorbid obesity and MMD [73]. We recommend that future studies approach depression as a group of separate symptoms rather than as a unified construct. The construct of sickness behavior could be particularly promising in this regard.

Our study features several strengths, namely the substantial sample size and the 9-year follow-up period wherein we analyzed individual symptoms of depression. Multiple reviews have published about the sickness behavior theory and how this could relate to symptoms of depression. However not many papers exist that tested how this theory translates to data of self-report symptoms of depression (1-5). This study is novel in the sense that we explicitly categorized symptoms into sickness behavior symptoms and non-sickness behavior symptoms and found a convincing stronger association with the first. Moreover, a wide array of inflammatory markers were assessed at baseline, including LPS-induced markers. We did not have preliminary hypotheses regarding which markers would indicate certain depressive symptoms, so we constructed three inflammatory indexes based on inflammatory markers to enhance the interpretability of our results. We demonstrated the utility of these index scores for research purposes and it's potential for clinical practice. By averaging multiple markers

the effect of individual measurement errors is reduced which is an important methodological advantage [74]. Some limitations must also be discussed. First, some of the component markers of the index scores were only weakly intercorrelated. Moreover, we composed two indexes based on data driven methods (Factor analysis [44]), more research is needed regarding grouping of individual markers based on underlying properties. Second, we repeatedly use the ill-defined term “sickness behavior”; different fields of medicine should solidify the definition so as to develop this construct in more depth [17, 18]. Third, due to logistical reasons, LPS-stimulated markers were only assessed in a consecutive subsample of 1229 participants. Fourth, previous studies found that antidepressants might have anti-inflammatory effects. Rats treated with fluoxetine demonstrated lower IL-1β in plasma and brain after 90 and 120-day treatment [75]. Furthermore, two meta-analyzes demonstrated that among MDD patients antidepressant treatment decreases TNF-α, IL-4, IL-6, IL-10 and IL-1ß [76, 77]. In the present study, the use of antidepressants was not adjusted for in our first models, as their use may indicate more severe depressive symptoms (confounding-by-indication) and therefore may lead to overadjustment. However, sensitivity analyzes demonstrate adding this variable as a confounder had only a limited effect on our outcomes and conclusions (SI figure 4) . Finally, beta coefficients were statistically significant but still of small effect sizes, with questionable clinical relevance. However, self-reported IDS items were scored on crude four-point scales, potentially contributing to measurement error and reduced statistical power. Moreover, the NESDA cohort only used a single measurement of inflammatory markers; trajectory analyzes with sequential day-to-day measures of inflammatory markers would have increased the precision of the independent variable.

In conclusion, we found that basal levels of inflammation and LPS-induced inflammatory markers predicted the course of individual depressive symptoms, especially those related to the construct of sickness behavior. This association persisted over the course of 9 years. Our findings suggest that inflammation might not relate to depression as one unified syndrome but rather to the presence and course of a subset of symptoms. Future studies should develop inflammation-targeted treatment strategies for individuals with symptom profiles associated with low-grade inflammation.

Acknowledgements and financial disclosures

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (Amsterdam University Medical Centers (location VUmc), GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum). BP has received (non-related) research funding from Boehringer Ingelheim and Jansen Research. All remaining authors report no biomedical financial interests or potential

conflicts of interest.

References

1. Miller, A.H. and C.L. Raison, *The role of inflammation in depression: from evolutionary imperative to modern treatment target.* Nat Rev Immunol, 2016. **16**(1): p. 22-34.
2. Krogh, J., et al., *The association between depressive symptoms, cognitive function, and inflammation in major depression.* Brain, behavior, and immunity, 2014. **35**: p. 70-76.
3. Dowlati, Y., et al., *A meta-analysis of cytokines in major depression.* Biological psychiatry, 2010. **67**(5): p. 446-457.
4. Howren, M.B., D.M. Lamkin, and J. Suls, *Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis.* Psychosomatic medicine, 2009. **71**(2): p. 171-186.
5. Liu, Y., R.C.-M. Ho, and A. Mak, *Interleukin (IL)-6, tumour necrosis factor alpha (TNF-α) and soluble interleukin-2 receptors (sIL-2R) are elevated in patients with major depressive disorder: a meta-analysis and meta-regression.* Journal of affective disorders, 2012. **139**(3): p. 230-239.
6. 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, 2014. **71**(12): p. 1381-1391.
7. Valkanova, V., K.P. Ebmeier, and C.L. Allan, *CRP, IL-6 and depression: a systematic review and meta-analysis of longitudinal studies.* Journal of affective disorders, 2013. **150**(3): p. 736-744.
8. Haapakoski, R., et al., *Cumulative meta-analysis of interleukins 6 and 1β, tumour necrosis factor α and C-reactive protein in patients with major depressive disorder.* Brain, behavior, and immunity, 2015. **49**: p. 206-215.
9. van der Linden, M.W., et al., *Determination of tumour necrosis factor-alpha and interleukin-10 production in a whole blood stimulation system: assessment of laboratory error and individual variation.* J Immunol Methods, 1998. **218**(1-2): p. 63-71.
10. Kovach, N.L., et al., *Lipid IVA inhibits synthesis and release of tumor necrosis factor induced by lipopolysaccharide in human whole blood ex vivo.* J Exp Med, 1990. **172**(1): p. 77-84.
11. Suarez, E.C., R.R. Krishnan, and J.G. Lewis, *The relation of severity of depressive symptoms to monocyte-associated proinflammatory cytokines and chemokines in apparently healthy men.* Psychosomatic medicine, 2003. **65**(3): p. 362-368.
12. Suarez, E.C., et al., *Enhanced expression of cytokines and chemokines by blood monocytes to in vitro lipopolysaccharide stimulation are associated with hostility and severity of depressive symptoms in healthy women.* Psychoneuroendocrinology, 2004. **29**(9): p. 1119-1128.
13. Vogelzangs, N., et al., *Cytokine production capacity in depression and anxiety.* Translational psychiatry, 2016. **6**(5): p. e825.
14. Gaspersz, R., et al., *The role of anxious distress in immune dysregulation in patients with major depressive disorder.* Translational psychiatry, 2017. **7**(12): p. 1268.
15. Shattuck, E.C. and M.P. Muehlenbein, *Human sickness behavior: Ultimate and proximate explanations.* American Journal of Physical Anthropology, 2015. **157**(1): p. 1-18.
16. Maes, M., et al., *Depression and sickness behavior are Janus-faced responses to shared inflammatory pathways.* BMC Med, 2012. **10**(1): p. 66.
17. Miller, A.H., L. Capuron, and C.L. Raison, *Immunologic influences on emotion regulation.* Clinical Neuroscience Research, 2005. **4**(5-6): p. 325-333.
18. Dantzer, R. and K.W. Kelley, *Twenty years of research on cytokine-induced sickness behavior.* Brain, behavior, and immunity, 2007. **21**(2): p. 153-160.
19. Vollmer-Conna, U., et al., *Production of pro-inflammatory cytokines correlates with the symptoms of acute sickness behaviour in humans.* Psychological medicine, 2004. **34**(7): p. 1289-1297.
20. Smith, A.P., *Effects of the common cold on mood, psychomotor performance, the encoding of new information, speed of working memory and semantic processing.* Brain, behavior, and immunity, 2012. **26**(7): p. 1072-1076.
21. Felger, J.C., *Imaging the Role of Inflammation in Mood and Anxiety-related Disorders.* Curr Neuropharmacol, 2018. **16**(5): p. 533-558.
22. Naude, P.J.W., et al., *Anxiety disorders and CRP in a population cohort study with 54,326 participants: The LifeLines study.* World J Biol Psychiatry, 2018. **19**(6): p. 461-470.

23. Renna, M.E., et al., *The association between anxiety, traumatic stress, and obsessive–compulsive disorders and chronic inflammation: A systematic review and meta-analysis.* Depression and anxiety, 2018. **35**(11): p. 1081-1094.
24. Bierhaus, A., et al., *A mechanism converting psychosocial stress into mononuclear cell activation.* Proceedings of the National Academy of Sciences, 2003. **100**(4): p. 1920-1925.
25. Pace, T.W., et al., *Increased stress-induced inflammatory responses in male patients with major depression and increased early life stress.* Am J Psychiatry, 2006. **163**(9): p. 1630-3.
26. Glaus, et al., *The bidirectional relationship between anxiety disorders and circulating levels of inflammatory markers: Results from a large longitudinal population-based study.* Depression and anxiety, 2018. **35**(4): p. 360-371.
27. Köhler-Forsberg, O., et al., *Association between C-reactive protein (CRP) with depression symptom severity and specific depressive symptoms in major depression.* Brain, Behavior, and Immunity, 2017. **62**: p. 344-350.
28. Sjoholm, L., C. Lavebratt, and Y. Forsell, *A multifactorial developmental model for the etiology of major depression in a population-based sample.* J Affect Disord, 2009. **113**(1-2): p. 66-76.
29. Fried and Nesse, *Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR* D study.* Journal of affective disorders, 2015. **172**: p. 96-102.
30. van Eeden, W.A., et al., *Severity, course trajectory, and within-person variability of individual symptoms in patients with major depressive disorder.* Acta Psychiatr Scand, 2019. **139**(2): p. 194-205.
31. Lamers, et al., *Metabolic and inflammatory markers: associations with individual depressive symptoms.* Psychological Medicine, 2017: p. 1-11.
32. Fried, E.I. and J.M. Haslbeck, *Using network analysis to examine links between individual depression symptoms, inflammatory markers, and covariates.* Psychological Medicine, 2019: p. 1-9.
33. Jokela, M., et al., *Inflammation and specific symptoms of depression.* JAMA psychiatry, 2016. **73**(1): p. 87-88.
34. White, J., et al., *Association of inflammation with specific symptoms of depression in a general population of older people: The English Longitudinal Study of Ageing.* Brain, behavior, and immunity, 2017. **61**: p. 27-30.
35. Horn, S.R., et al., *Replication and reproducibility issues in the relationship between C-reactive protein and depression: A systematic review and focused meta-analysis.* Brain, behavior, and immunity, 2018. **73**: p. 85-114.
36. Uher, R., et al., *An inflammatory biomarker as a differential predictor of outcome of depression treatment with escitalopram and nortriptyline.* American Journal of Psychiatry, 2014. **171**(12): p. 1278-1286.
37. Huang, M., et al., *Longitudinal association of inflammation with depressive symptoms: A 7-year cross-lagged twin difference study.* Brain, behavior, and immunity, 2019. **75**: p. 200-207.
38. Lamers, F., et al., *Longitudinal Association Between Depression and Inflammatory Markers: Results From the Netherlands Study of Depression and Anxiety.* Biological psychiatry, 2019.
39. Ng, A., et al., *IL-1beta, IL-6, TNF- alpha and CRP in Elderly Patients with Depression or Alzheimer's disease: Systematic Review and Meta-Analysis.* Sci Rep, 2018. **8**(1): p. 12050.
40. Lamers, F., et al., *Depression profilers and immuno-metabolic dysregulation: Longitudinal results from the NESDA study.* Brain Behav Immun, 2020.
41. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* Int J Methods Psychiatr Res, 2008. **17**(3): p. 121-40.
42. Gerrits, M.M., et al., *The impact of chronic somatic diseases on the course of depressive and anxiety disorders.* Psychother Psychosom, 2013. **82**(1): p. 64-6.
43. Wittchen, H.-U., *Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review.* Journal of Psychiatric Research, 1994. **28**(1): p. 57-84.
44. Bandalos, D.L. and S.J. Finney, *Factor analysis: Exploratory and confirmatory*, in *The reviewer's guide to quantitative methods in the social sciences*. 2018, Routledge. p. 98-122.
45. Trivedi, M.H., et al., *The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation.* Psychol Med, 2004. **34**(1): p. 73-82.
46. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): psychometric properties.* Psychol Med, 1996. **26**(3): p. 477-86.

47. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the royal statistical society. Series B (Methodological), 1995: p. 289-300.
48. Smith, K.J., et al., *The association between C-reactive protein, Interleukin-6 and depression among older adults in the community: a systematic review and meta-analysis.* Experimental gerontology, 2018. **102**: p. 109-132.
49. Haroon, E., et al., *Antidepressant Treatment Resistance is Associated with Increased Inflammatory Markers in Patients with Major Depressive Disorder.* Psychoneuroendocrinology, 2018.
50. Duivis, H.E., et al., *Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: findings from the Netherlands Study of Depression and Anxiety (NESDA).* Psychoneuroendocrinology, 2013. **38**(9): p. 1573-1585.
51. Smith, R.S., *The macrophage theory of depression.* Medical hypotheses, 1991. **35**(4): p. 298-306.
52. Berk, M., et al., *So depression is an inflammatory disease, but where does the inflammation come from?* BMC medicine, 2013. **11**(1): p. 200.
53. Rosenblat, J.D., et al., *Inflamed moods: A review of the interactions between inflammation and mood disorders.* Progress in Neuro-Psychopharmacology and Biological Psychiatry, 2014. **53**: p. 23-34.
54. van den Biggelaar, et al., *Inflammation and interleukin-1 signaling network contribute to depressive symptoms but not cognitive decline in old age.* Experimental gerontology, 2007. **42**(7): p. 693-701.
55. Walker, K.J., *Chemokines : Types, Functions, and Structural Characteristics*. Cell Biology Research Progress. 2011, New York: Nova Science Publishers, Inc.
56. Dinarello, C.A., *Historical insights into cytokines.* European journal of immunology, 2007. **37**(S1): p. S34-S45.
57. Jones, S.A. and B.J. Jenkins, *Recent insights into targeting the IL-6 cytokine family in inflammatory diseases and cancer.* Nature Reviews Immunology, 2018: p. 1.
58. Srivastava, S., et al., *Effects of interleukin-18 on natural killer cells: costimulation of activation through Fc receptors for immunoglobulin.* Cancer Immunology, Immunotherapy, 2013. **62**(6): p. 1073-1082.
59. Wechsler, A.S., et al., *Induction of IL-8 expression in T cells uses the CD28 costimulatory pathway.* The Journal of Immunology, 1994. **153**(6): p. 2515-2523.
60. Sebok, K., et al., *IL-8 induces the locomotion of human IL-2-activated natural killer cells. Involvement of a guanine nucleotide binding (Go) protein.* The Journal of Immunology, 1993. **150**(4): p. 1524-1534.
61. Komai-Koma, M., et al., *Chemoattraction of human T cells by IL-18.* The Journal of Immunology, 2003. **170**(2): p. 1084-1090.
62. Majidi-Zolbanin, S., et al., *Correlation between major depressive disorder and circulating natural killer cells.* Archives of Medical Laboratory Sciences, 2015. **1**(2).
63. Fried, et al., *Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors.* Psychol. Med., 2014. **44**(10): p. 2067-2076.
64. Raison, C.L. and A.H. Miller, *Is depression an inflammatory disorder?* Current psychiatry reports, 2011. **13**(6): p. 467-475.
65. Jha, M. and M. Trivedi, *Personalized antidepressant selection and pathway to novel treatments: clinical utility of targeting inflammation.* International journal of molecular sciences, 2018. **19**(1): p. 233.
66. Gimeno, D., et al., *Associations of C-reactive protein and interleukin-6 with cognitive symptoms of depression: 12-year follow-up of the Whitehall II study.* Psychological medicine, 2009. **39**(3): p. 413-423.
67. Au, B., et al., *The longitudinal associations between C-reactive protein and depressive symptoms: evidence from the English Longitudinal Study of Ageing (ELSA).* International journal of geriatric psychiatry, 2015. **30**(9): p. 976-984.
68. Ridker, P.M., et al., *Prospective study of C-reactive protein and the risk of future cardiovascular events among apparently healthy women.* Circulation, 1998. **98**(8): p. 731-733.
69. Ridker, P.M., et al., *C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women.* New England journal of medicine, 2000. **342**(12): p. 836-843.
70. Kopschina Feltes, P., et al., *Anti-inflammatory treatment for major depressive disorder: implications for patients with an elevated immune profile and non-responders to standard antidepressant therapy.* Journal of Psychopharmacology, 2017. **31**(9): p. 1149-1165.
71. Lee, Y., et al., *Anti-cytokine agents for anhedonia: targeting inflammation and the immune system to treat dimensional disturbances in depression.* Ther Adv Psychopharmacol, 2018. **8**(12): p. 337-348.
72. Husain, M.I., et al., *Adjunctive simvastatin for treatment-resistant depression: study protocol of a 12-week randomised controlled trial.* BJPsych open, 2019. **5**(1).

73. *Simvastatin add-on to Escitalopram in patients with comorbid obesity and major depression: A multicenter, randomized, double-blind, placebo-controlled trial; EudraCT number 2018-002947-27*. 2019, Charité – Universitätsmedizin Berlin: Germany.
74. Hardikar, S., et al., *Intraindividual variability over time in plasma biomarkers of inflammation and effects of long-term storage.* Cancer Causes Control, 2014. **25**(8): p. 969-76.
75. Lu, Y., et al., *Chronic administration of fluoxetine and pro-inflammatory cytokine change in a rat model of depression.* PLoS One, 2017. **12**(10): p. e0186700.
76. Wiedlocha, M., et al., *Effect of antidepressant treatment on peripheral inflammation markers - A meta-analysis.* Prog Neuropsychopharmacol Biol Psychiatry, 2018. **80**(Pt C): p. 217-226.
77. Hannestad, J., N. DellaGioia, and M. Bloch, *The effect of antidepressant medication treatment on serum levels of inflammatory cytokines: a meta-analysis.* Neuropsychopharmacology, 2011. **36**(12): p. 2452-9.

Supplementary material

Basal inflammation data collection procedure

Baseline inflammatory markers CRP, IL-6, and TNF-α were determined from fasting blood plasma. After an overnight fast, 50 ml of blood was drawn, immediately transferred to a local laboratory, and kept frozen at −80°C. High sensitivity plasma levels of CRP were measured in duplicate by an in-house, high sensitivity enzyme-linked immunosorbent assay (ELISA), which is based on a purified protein and polyclonal anti-CRP antibodies (Dako, Glostrup, Denmark). The lower detection limit of CRP is 0.1 mg/l, and the sensitivity is 0.05 mg/l. Intra- and inter-assay coefficients of variation were 5% and 10%, respectively. Plasma IL-6 levels were measured in duplicate by a high sensitivity ELISA (PeliKine Compact™ ELISA, Sanquin, Amsterdam, the Netherlands). The lower detection limit of IL-6 is 0.35 pg/ml, and the sensitivity is 0.10 pg/ml. Intra- and inter-assay coefficients of variation were 8% and 12%, respectively. Plasma TNF-α levels were assayed in duplicate using a high sensitivity solid phase ELISA (Quantikine HS Human TNF-α Immunoassay, R&D systems, Minneapolis, MN, USA). The lower detection limit of TNF-α is 0.10 pg/ml, and the sensitivity is 0.11 pg/ml. Intra- and inter-assay coefficients of variation were 10% and 15%, respectively.

Inflammatory markers after the LPS-induction data-collection procedure

The innate immune response of 12 cytokines was assessed in blood that was ex vivo stimulated with LPS. Venous whole blood samples were obtained at baseline in single 7-ml heparin-coated tubes (Greiner Bio-One, Monroe, NC, USA). Between 10 and 60 min after blood draw, 2.5 ml of blood was transferred into a PAXgene tube (Qiagen, Valencia, CA, USA). Remaining blood (4.5 ml) was stimulated by addition of LPS (10 ng/ml − 1 blood; Escherichia coli, Sigma, St. Louis, MO, USA). LPS-stimulated samples were laid flat and incubated at a slow rotation for 5–6 hr at 37°C. A 2.5 ml sample of this LPS-stimulated blood was transferred into a PAXgene tube. This LPS procedure was carried out at four laboratories (Amsterdam, Leiden, Groningen, and Heerenveen, The Netherlands).

Levels of interferon-γ (IFN-γ), macrophage inflammatory protein-α (MIP-1α), IL-2, IL-6, IL-8, IL-10, IL-18, MCP-1, macrophage inflammatory protein-α (MIP-1α), MIP-1β, matrix metallopeptidase-2 (MMP-2), TNF-α, and TNF-β were assessed simultaneously for all available samples, using a multi-analytic profile (Human CytokineMAP A v.1.0; Myriad RBM,

Austin, TX, USA). This commercial platform adheres to stringent guidelines of quality control and has Clinical Laboratory Improvement Amendments (CLIA) approval. Cytokine distributions were skewed to the right and therefore $\log_e$-transformed to normalize their distributions.

We created an LPS-induced inflammation index from the mean standardized value of all available LPS-induced markers, further referred to as the LPS-induced inflammation index. To avoid loss of information, we conducted an exploratory factor analysis (EFA) with promax rotation on all LPS-induced inflammatory markers in order to reduce these into additional data-driven index scores. This resulted into two LPS-induced inflammation indexes with an eigenvalue of 0.55, further referred to as LPS-induced inflammation index-1 and LPS-induced inflammation index-2. Markers IFN-γ, IL-10, IL-2, IL-6, MMP-2, TNF-α, and TNF-β loaded on LPS-induced inflammation index-1 with factor loadings between 0.41 and 0.88 and a raw alpha of 0.86. IL-8, IL-18, MCP-1, MIP-1α, and MIP-1β loaded on LPS-induced inflammation index-2 with loadings between 0.34 and 0.94 and a raw alpha of 0.89. This two-factor solution fitted the data better but still poorly—Comparative Fit Index (CFI) = 0.867, Tucker-Lewis index (TLI) = 0.796, Root Mean Square Error of Approximation (RMSEA) = 0.187—compared to a one-factor solution: CFI = 0.794, TLI = 0.748, and RMSEA = 0.208. See SI Figure 1 for the correlations between individual markers within each index. Subsequently, two LPS-induced inflammation indexes were calculated as the mean of $\log_e$-transformed and standardized markers.

SI Table 1. IDS symptoms over the course of nine years in relation to inflammatory markers for MDD patients only

	Basal Serum inflammation index			LPS-induced index inflammation factor 1			LPS-induced index inflammation factor 2		
	CRP, TNF-α, IL-6			IL-2, IL-6, IL-10, MMP-2, TNF-α, TNF-β, IFN-γ			IL-8, IL-18, MCP-1, MIP-1α, MIP-1β		
Item	Beta	(SE)	p-value	Beta	(SE)	p-value	Beta	(SE)	p-value
1. Falling asleep	-0.001	(0.026)	0.955	-0.055	(0.046)	0.229	-0.026	(0.049)	0.598
2. Sleep during the night	0.023	(0.022)	0.286	-0.013	(0.037)	0.723	0.015	(0.041)	0.709
3. Waking up too early	0.043	(0.025)	0.087	-0.054	(0.045)	0.231	-0.026	(0.048)	0.584
4. Sleeping too much	0.016	(0.026)	0.544	-0.022	(0.044)	0.623	0.003	(0.048)	0.946
5. Feeling Sad	0.045	(0.021)	0.034*	0.002	(0.038)	0.951	0.067	(0.041)	0.104
6. Feeling irritable	0.008	(0.021)	0.712	0.021	(0.038)	0.572	0.101	(0.040)	0.013*
7. Anxious or tense	0.004	(0.022)	0.843	0.002	(0.038)	0.965	0.086	(0.041)	0.040
8. Response of mood	0.027	(0.025)	0.273	-0.027	(0.043)	0.532	0.017	(0.045)	0.703
9a. Mood in time of day	-0.036	(0.025)	0.153	0.020	(0.044)	0.649	0.029	(0.048)	0.536
10. Quality of mood	0.047	(0.023)	0.038	0.008	(0.040)	0.840	0.081	(0.043)	0.059
11. Decreased appetite	0.035	(0.023)	0.140	0.008	(0.040)	0.844	0.091	(0.042)	0.029*
12. Increased appetite	0.077	(0.026)	0.003*	-0.022	(0.046)	0.641	-0.024	(0.050)	0.633
13. Decreased weight	0.018	(0.019)	0.492	-0.028	(0.033)	0.409	0.051	(0.035)	0.139
14. Increased weight	0.031	(0.020)	0.120	0.011	(0.035)	0.752	0.039	(0.038)	0.298
15. Concentration	0.018	(0.021)	0.378	0.030	(0.036)	0.409	0.092	(0.039)	0.019*
16. View of myself	0.043	(0.024)	0.078	0.023	(0.043)	0.585	0.114	(0.046)	0.014*
17. View of my future	0.045	(0.020)	0.024*	0.041	(0.034)	0.228	0.093	(0.037)	0.014*
18. Death or suicide	0.041	(0.027)	0.130	0.007	(0.047)	0.890	0.032	(0.050)	0.532
19. General interest	0.055	(0.024)	0.020*	-0.032	(0.042)	0.453	0.051	(0.045)	0.255
20. Energy level	0.059	(0.021)	0.004*	-0.030	(0.036)	0.410	0.015	(0.039)	0.699
21. Capacity for pleasure	0.056	(0.024)	0.020*	-0.030	(0.043)	0.474	0.042	(0.046)	0.357
22. Interest in sex	0.037	(0.023)	0.112	-0.003	(0.043)	0.948	-0.008	(0.046)	0.860
23. Psychomotor retardation	0.067	(0.027)	0.013*	-0.042	(0.048)	0.380	0.058	(0.051)	0.258
24. Psychomotor agitation	0.016	(0.025)	0.518	0.020	(0.044)	0.651	0.070	(0.047)	0.136
25. Aches and pains	0.096	(0.021)	<0.001*	0.014	(0.038)	0.703	0.122	(0.041)	0.003*
26. Sympathetic arousal	0.048	(0.022)	0.030*	0.001	(0.040)	0.973	0.095	(0.043)	0.029*
27. Panic/Phobic	-0.008	(0.025)	0.758	-0.002	(0.045)	0.969	0.017	(0.048)	0.721
28. Constipation/diarrhea	0.031	(0.024)	0.201	-0.009	(0.043)	0.832	0.007	(0.046)	0.879
29. Interpersonal sensitivity	-0.022	(0.022)	0.325	0.003	(0.038)	0.940	0.106	(0.042)	0.011*
30. Leaden paralysis	0.043	(0.020)	0.028*	-0.006	(0.034)	0.861	0.015	(0.037)	0.687

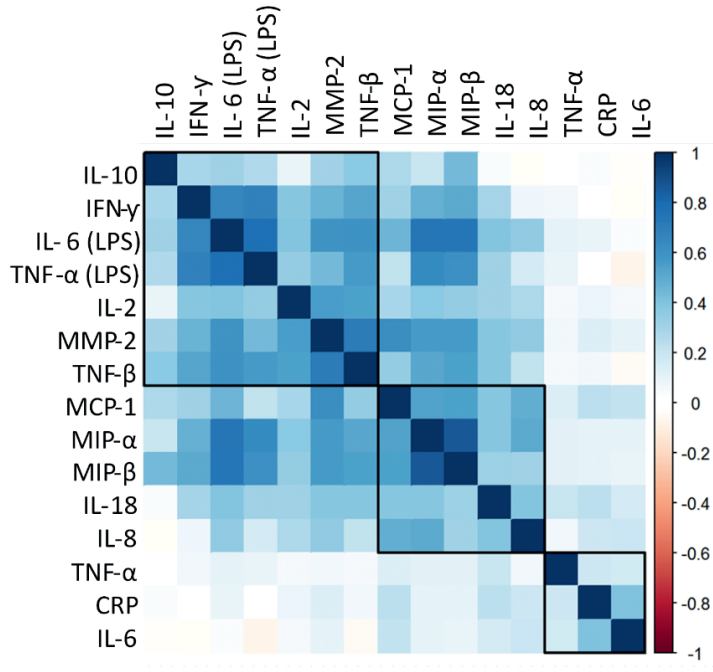
SI Table 1. Standardized beta coefficients of the association between inflammatory markers and individual depressive symptoms in a sample of MDD patients only. Standardized beta coefficients of linear mixed models with basal inflammation index and LPS-induced inflammation index-1 and -2 assessed with repeated measures, used to predict standardized IDS-SR item-scores measured over 9 years of follow up. Assessed at up to six time-points, adjusted for baseline variables of gender, age, sickness prior to interview, and the use of anti-inflammatory medication in a sample of MDD patients only.

**P* values that remained significant (< 0.05) after correcting for multiple testing using the Benjamin–Hochberg procedure.

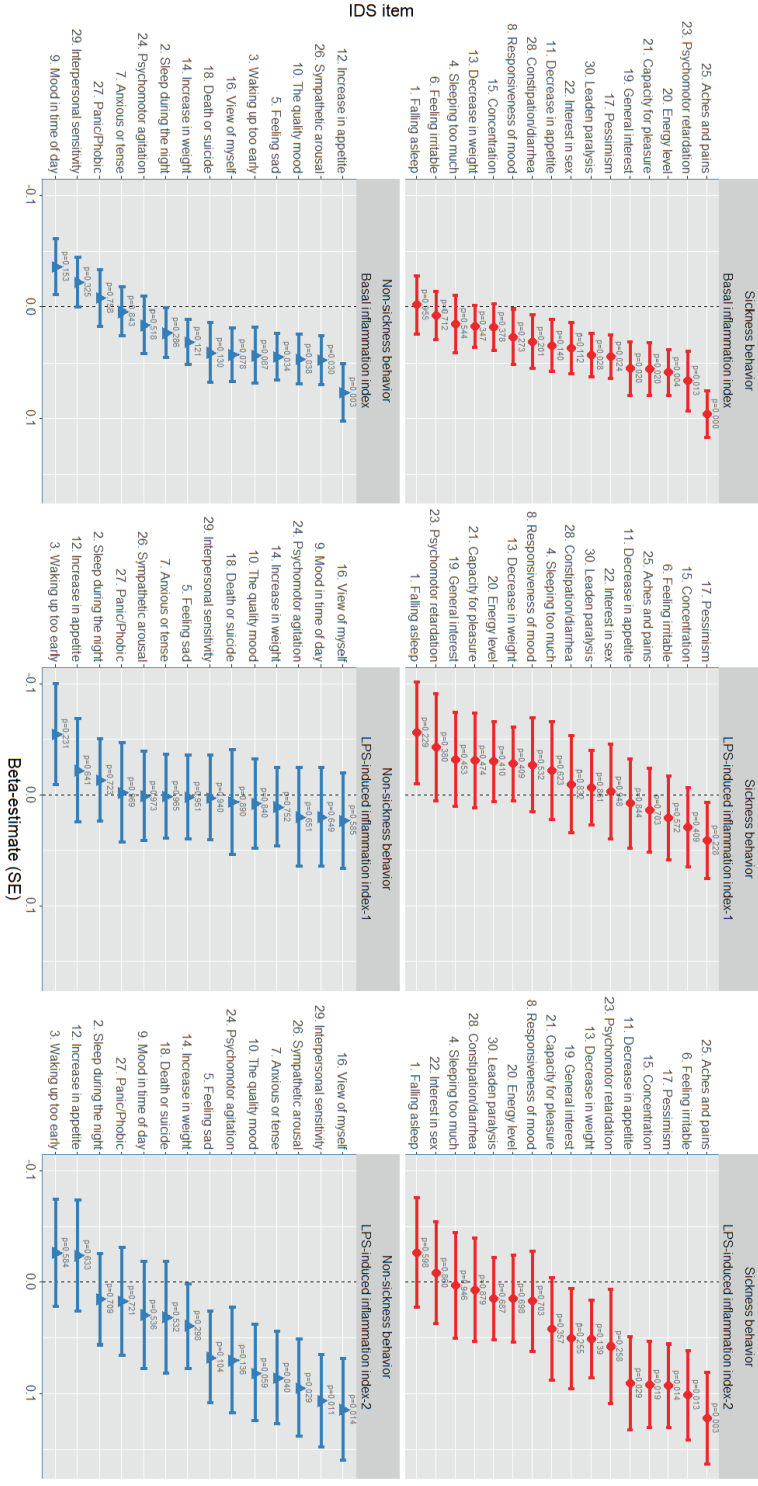
SI Table 2. IDS symptoms over the course of nine years in relation to LPS-induced inflammatory markers

	LPS-induced Inflammationindex	
	IL-2, IL-6, IL-10, MMP-2, TNF-α, TNF-β, IFN-γ, IL-8, IL-18, MCP-1, MIP-1α, MIP-1β	
Item	Beta (SE)	p-value
1. Falling asleep	0.014 (0.024)	0.552
2. Sleep during the night	0.029 (0.022)	0.186
3. Waking up too early	0.000 (0.023)	0.989
4. Sleeping too much	0.013 (0.023)	0.569
5. Feeling Sad	0.025 (0.024)	0.315
6. Feeling irritable	0.060 (0.024)	0.011*
7. Anxious or tense	0.051 (0.024)	0.033*
8. Response of mood	0.041 (0.021)	0.058
9a. Mood in time of day	0.003 (0.022)	0.894
10. Quality of mood	0.047 (0.023)	0.045
11. Decreased appetite	0.035 (0.019)	0.064
12. Increased appetite	0.000 (0.021)	0.993
13. Decreased weight	0.021 (0.017)	0.198
14. Increased weight	0.017 (0.018)	0.342
15. Concentration	0.041 (0.023)	0.082
16. View of myself	0.030 (0.024)	0.208
17. View of my future	0.055 (0.024)	0.023*
18. Death or suicide	0.039 (0.024)	0.098
19. General interest	0.037 (0.022)	0.094
20. Energy level	0.044 (0.023)	0.052
21. Capacity for pleasure	0.030 (0.023)	0.195
22. Interest in sex	0.012 (0.023)	0.586
23. Psychomotor retardation	0.032 (0.023)	0.163
24. Psychomotor agitation	0.021 (0.024)	0.381
25. Aches and pains	0.066 (0.023)	0.005*
26. Sympathetic arousal	0.051 (0.023)	0.026*
27. Panic/Phobic	0.062 (0.024)	0.010*
28. Constipation/diarrhea	0.040 (0.023)	0.080
29. Interpersonal sensitivity	0.035 (0.024)	0.141
30. Leaden paralysis	0.045 (0.024)	0.064

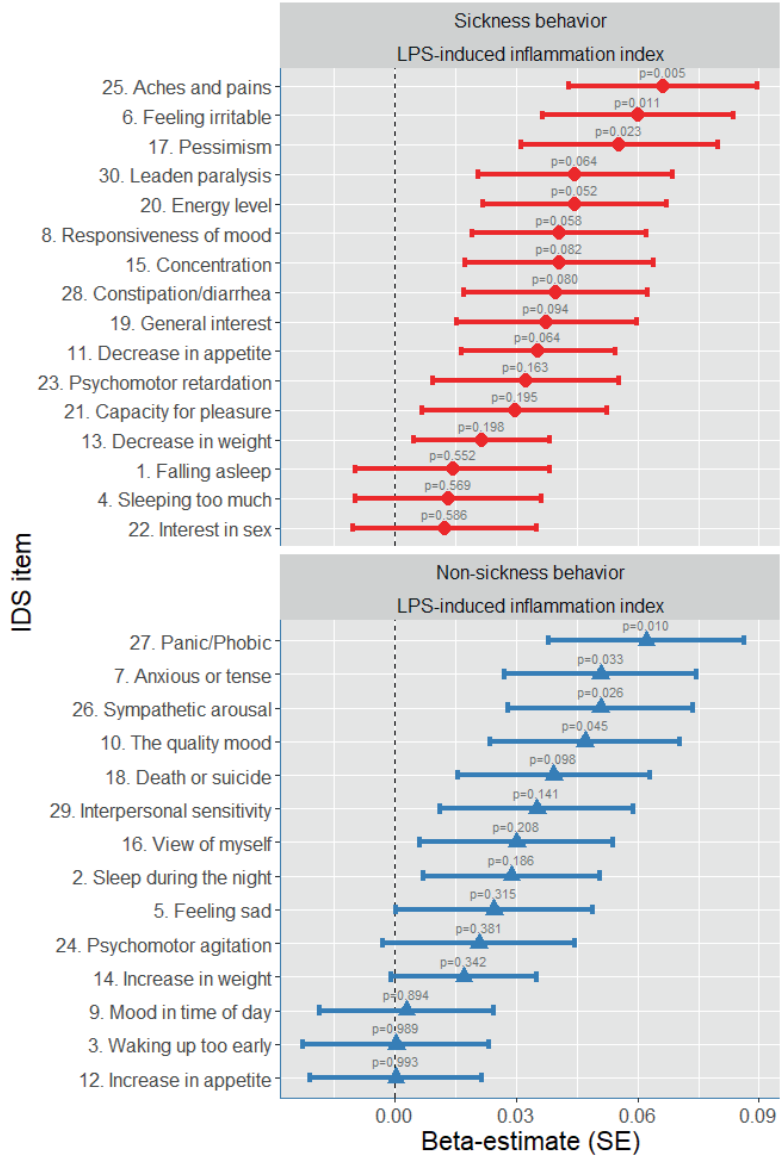
SI Table 2: Standardized beta coefficients of the association between LPS-induced inflammatory markers and individual depressive symptoms. Standardized beta coefficients of the LPS-induced inflammation index, assessed using a mixed model with repeated measures with standardized IDS-SR item score as the outcome variable. Assessed at six time points over the 9 years of follow up and adjusted for baseline variables of gender, age, sickness prior to interview, and the use of anti-inflammatory medication. **P* values that remained significant (< 0.05) after correcting for multiple testing using the Benjamin–Hochberg procedure.


IL-10
IFN-γ
IL-6 (LPS)
TNF-α (LPS)
IL-2
MMP-2
TNF-β
MCP-1
MIP-α
MIP-β
IL-18
IL-8
TNF-α
CRP
IL-6
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

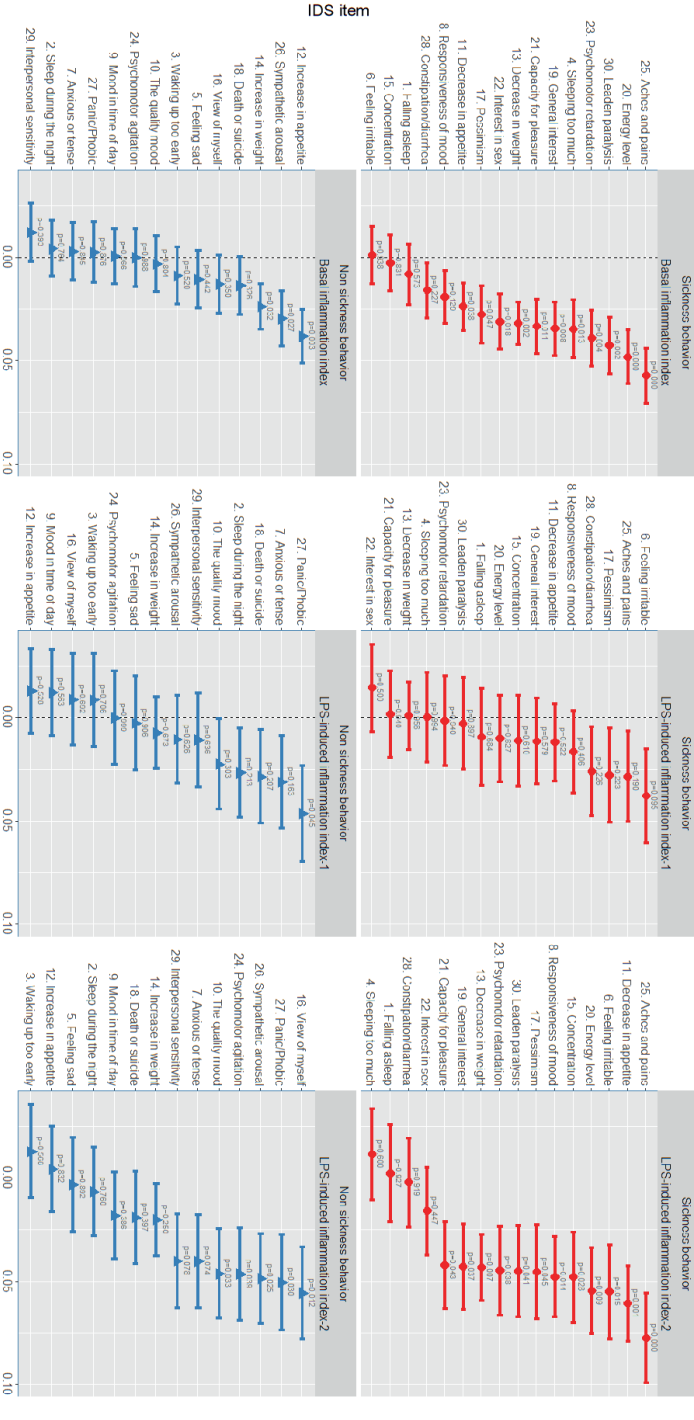
SI Figure 1. Correlations between inflammatory markers. The black lines demonstrate the three indexes (from left to right): LPS-induced inflammation index-1(composed of IL-10, IFN-γ, IL-2, IL-6, MMP-2, TNF-α, and TNF-β), LPS-induced inflammation index-2 (composed of IL-8, IL-18, MCP-1, MIP-1α, and MIP-1β), and basal inflammation index (composed of basal levels of CRP, IL-6, and TNF-α). Because LPS-induced markers were available for a subset of *n* = 1229 out of *n* = 2904 participants, the basal inflammation index was incomplete in the present figure.


Sickness behavior
Basal inflammation index
Non-sickness behavior
Basal inflammation index
Sickness behavior
LPS-induced inflammation index-1
Non-sickness behavior
LPS-induced inflammation index-1
Sickness behavior
LPS-induced inflammation index-2
Non-sickness behavior
LPS-induced inflammation index-2
IDS item
Beta-estimate (SE)

SI Figure 2. Associations of the basal inflammation index (*n* = 908), LPS-induced inflammation index-1 (*n* = 338), and LPS-induced inflammation index-2 (*n* = 364) with individual depressive symptoms during 9 years within a subsample of MDD patients. Standardized beta coefficients with error bars representing standard errors of the predictive values of inflammatory indexes in relation to individual depressive symptoms during 9 years of follow up. Assessed using linear mixed models with repeated measures, adjusted for gender, age, use of anti-inflammatory drugs, and sickness prior to interview.


Sickness behavior
LPS-induced inflammation index
25. Aches and pains
6. Feeling irritable
17. Pessimism
30. Leaden paralysis
20. Energy level
8. Responsiveness of mood
15. Concentration
28. Constipation/diarrhea
19. General interest
11. Decrease in appetite
23. Psychomotor retardation
21. Capacity for pleasure
13. Decrease in weight
1. Falling asleep
4. Sleeping too much
22. Interest in sex
p=0.005
p=0.011
p=0.023
p=0.064
p=0.052
p=0.058
p=0.082
p=0.080
p=0.094
p=0.064
p=0.163
p=0.195
p=0.198
p=0.552
p=0.569
p=0.586
Non-sickness behavior
LPS-induced inflammation index
27. Panic/Phobic
7. Anxious or tense
26. Sympathetic arousal
10. The quality mood
18. Death or suicide
29. Interpersonal sensitivity
16. View of myself
2. Sleep during the night
5. Feeling sad
24. Psychomotor agitation
14. Increase in weight
9. Mood in time of day
3. Waking up too early
12. Increase in appetite
p=0.010
p=0.033
p=0.026
p=0.045
p=0.098
p=0.141
p=0.208
p=0.186
p=0.315
p=0.381
p=0.342
p=0.894
p=0.989
p=0.993
IDS item
0.00
0.03
0.06
0.09
Beta-estimate (SE)

SI Figure 3. Associations of the LPS-induced inflammation index with individual depressive symptoms during 9 years for the whole sample (*n* = 1147). Standardized beta coefficients with error bars representing standard errors of the predictive values of inflammatory indexes in relation to individual depressive symptoms during 9 years follow-up. Assessed using linear mixed models with repeated measures adjusted for gender, age, use of anti-inflammatory drugs, and sickness prior to interview.


Sickness behavior
Basal inflammation index
Sickness behavior
LPS-induced inflammation index-1
Sickness behavior
LPS-induced inflammation index-2
Non sickness behavior
Basal inflammation index
Non sickness behavior
LPS-induced inflammation index-1
Non sickness behavior
LPS-induced inflammation index-2
IDS item
Beta-estimate (SE)
0.00
0.05
0.10

SI Figure 4. Associations of the basal inflammation index (*n* = 2872), LPS-induced inflammation index-1 (*n* = 1147), and LPS-induced inflammation index-2 (*n* = 1229) with individual depressive symptoms during 9 years. Standardized beta coefficients with error bars representing standard errors of the predictive values of inflammatory indexes in relation to individual depressive symptoms during 9 years of follow up. Assessed using linear mixed models with repeated measures, adjusted for chronic somatic diseases, antidepressants, gender, age, use of anti-inflammatory drugs, and sickness prior to interview.

* Wessel A. van Eeden and Ebtisam El Filali share first authorship.
[1] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands;
[2] Department of Mood disorders, PsyQ, Parnassia Group, The Hague, The Netherlands;
[3] Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam UMC, Vrije Universiteit, and GGZ inGeest, Amsterdam, The Netherlands;
[4] Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands;

Chapter 6

Basal and LPS-stimulated inflammatory markers and the course of anxiety symptoms

van Eeden, W.A. [1*] , El Filali, E.[1,2 *], van Hemert[1], A.M., Carlier, I.V.E[1], Penninx, B.W.J.H. [3], Lamers, F.[3], Schoevers, R.[4], Giltay, E.J[1]

(2021). Brain, behavior, and immunity

Abstract

A cross-sectional relationship between low-grade inflammation –characterized by increased blood levels of C-reactive protein (CRP) and pro-inflammatory cytokines– and anxiety has been reported, but the potential longitudinal relationship has been less well studied. We aimed to examine whether basal and lipopolysaccharide (LPS-)induced levels of inflammatory markers are associated with anxiety symptom severity over the course of nine years.

We tested the association between basal and LPS-induced inflammatory markers with anxiety symptoms (measured with the Beck's Anxiety Inventory; BAI, Fear Questionnaire;FQ and Penn's State Worry Questionnaire; PSWQ) at 5 assessment waves over a period up nine years. We used multivariate-adjusted mixed models in up to 2867 participants of the Netherlands Study of Depression and Anxiety (NESDA).

At baseline, 43.6% of the participants had a current anxiety disorder, of which social phobia (18.5%) was most prevalent. Our results demonstrated that baseline inflammatory markers were significantly associated with several outcomes of anxiety at baseline over nine subsequent years. BAI subscale of somatic (arousal) symptoms of anxiety, and FQ subscale of agoraphobia demonstrated the strongest effects with standardized beta-coefficients of up to 0.14. The associations were attenuated by 25%-30% after adjusting for the presence of (comorbid) major depressive disorder (MDD), but remained statistically significant.

In conclusion, we found that participants with high levels of inflammatory markers have on average high levels of anxiety consisting of physical arousal and agoraphobia, which tended to persist over a period of nine years, albeit with small effect sizes. These associations were partly driven by co-morbid depression.

Key words: Anxiety severity, Anxiety disorder, Epidemiology, Inflammation, Longitudinal

Highlights

1. A wide array of inflammatory markers were assessed, including LPS-induced markers.
2. Anxiety symptoms were assessed over the course of 9 years.
3. Baseline Inflammatory markers were associated to anxiety over nine years follow-up.
4. Somatic symptoms of anxiety and symptoms of agoraphobia were related the strongest.
5. Associations were partly driven by (co-morbid) depression.

6.1 Introduction

Anxiety is regarded as a psychobiological state or reaction that, amongst others, consists of unpleasant subjective feelings of tension, nervousness and worry, often accompanied by physiological manifestations such as increased heart rate and blood pressure, and irregularity of breathing [1]. Earlier studies have suggested that inflammation could be involved in the pathophysiology of anxiety [2-6]. There are many pathways which may underlie this link. In laboratory conditions, anxiety can be induced by an external stressor (Trier social stress test), resulting in the characteristic physiological changes, as well as the biochemical response of cortisol and catecholamines release [7]. Interestingly, this also activated inflammatory pathways in peripheral mononuclear cells through the transcription factor-kB (NF-kB), leading to increased levels of circulating pro-inflammatory cytokines such as interleukin-6 [8, 9]. Similarly, chronic psychosocial distress, which goes hand in hand with symptoms of anxiety [10], has been linked to dysregulation of the hypothalamic-pituitary-adrenal axis, which has been shown to impact immune regulation [11, 12]. In reverse, following administration of the cytokine interferon alpha (IFN-α), significant anxiety as well as depressive symptoms may arise [13, 14]. These symptoms could be prevented when patients were pretreated with selective serotonin reuptake inhibitors (SSRIs) before the start of IFN-α administration, indicating that these inflammation-related symptoms may in part be mediated through serotonin [15].

There is increasing evidence for higher circulating concentrations of acute-phase proteins and pro-inflammatory cytokines in anxiety patients versus healthy subjects. Specifically C-reactive protein (CRP) as well as the pro-inflammatory cytokine interleukin-6 (IL-6) appear to have been repeatedly associated with symptoms and disorders of anxiety, such as panic disorders [16], generalized anxiety disorders [17], agoraphobia [18, 19] and anxiety symptoms in general [1, 20]. However, other studies did not find significant associations or even found reduced levels of inflammatory markers in subjects with anxiety symptoms [3, 21, 22]. Almost all previous studies had cross-sectional designs. One large longitudinal study that included 3,113 participants from the general population found that anxiety disorders, of which particularly agoraphobia, were associated with a steeper increase in CRP over time (not with

IL-6 and Tumor Necrosis Factor-alpha; TNF-a), but baseline inflammatory markers did not predict anxiety disorders the other way around during up to 5.5 years of follow-up [19].

Lipopolysaccharide (LPS) stimulated cytokine levels may better reflect physiological immune system functioning *in vivo* than basal levels of inflammation markers [23]. After *ex vivo* exposure of whole blood samples to LPS (the cell membrane of Gram-negative bacteria that strongly induce immunological responses), a wide array of pro- and anti-inflammatory cytokines are released that can be measured in the supernatant [24-26]. Whereas basal serum levels of inflammatory mediators generally show low values with high variability between and within persons over time (partly due to circadian rhythmicity), LPS-stimulated cytokine levels may have less of these drawbacks [27].

Previous cross-sectional analyzes from the NESDA cohort, that we used, have shown that basal inflammatory markers [3, 28], as well as LPS-induced inflammatory markers [29, 30], were positively associated with anxiety and major depressive disorders at baseline (MDD). Vogelzangs et al. (2016) showed that LPS-stimulated inflammation was associated with increased odds of anxiety disorders, whereas Gaspersz et al. (2017) found that LPS-induced inflammatory markers were especially elevated among MDD patients with the DSM-5 'anxious distress'-specifier. Although several analyzes within the NESDA cohort have focused on the prospective relationship of inflammation and depression, the longitudinal relation with anxiety symptoms has not been analyzed [31, 32]. Prospective studies regarding anxiety symptom severity remain scarce.

The aim of the present study is to examine whether basal as well as LPS-induced inflammatory markers determined at baseline are associated with the course of anxiety symptoms in the large Netherlands Study of Depression and Anxiety (NESDA) cohort. For this purpose, we chose three often-used self-reported measures of anxiety symptoms as outcome variables. Together this gives a broad spectrum of anxiety symptomatology containing subjective and somatic experienced anxiety, avoidance and worry. We hypothesize that markers of (low-grade) inflammation are associated with elevated levels of anxiety over the course of nine years, measured at baseline and up to five following time-points. In order to study whether the relationship with anxiety was independent of that with depression, we adjusted for the presence of MDD in a sensitivity analysis.

6.2 Materials and methods

6.2.1 Study sample and procedure

We evaluated baseline and follow-up data from participants from the Netherlands Study of Depression and Anxiety (NESDA) cohort. A detailed description of the NESDA design and sampling procedures have been published elsewhere [33]; its aim was to investigate the course and consequences of depressive and anxiety disorders. The first wave (baseline) started in 2004 and ended in September 2007, and the 6th wave of measurement at 9-year follow-up finished in October 2016. The baseline measurement (n=2,981) consisted of demographic and personal characteristics, a standardized diagnostic psychiatric interview, medical assessment (e.g. BMI, blood sampling, etc.), and self-report questionnaires. The 1-year follow-up consisted of self-report questionnaires and was completed by 2,445 participants (82.0%). Face-to-face follow-up assessments with standardized diagnostic psychiatric interviewing and self-report questionnaires were conducted at 2 years (n=2,596, 87.1%), 4 years (n=2,402, %), 6 years (n=2,256, 75.7%) and 9 years post-baseline (n=2,069, 69.4% of the baseline sample).

This cohort was recruited from the community (n=564, 18.9%), general practice (n=1,610, 54.0%), and secondary mental healthcare [n=807, 27.1%; 33]. Basal serum levels of inflammation were collected from 2,867 of 2,981 participants (96.2%). LPS induction in blood was only assessed during the last year of baseline data collection, due to logistical reasons. As a consequence, inflammatory markers after in vitro LPS induction of whole blood samples was therefore available for the subgroup of 1,229 out of 2,981 participants (41.2%). A general inclusion criterion was an age of 18 through 65 years. Only two exclusion criteria existed: 1) a primary clinical diagnosis of a psychiatric disorder not subject of NESDA which will largely affect course trajectories, including a psychotic disorder, obsessive compulsive disorder, bipolar disorder, or severe addiction disorder; and 2) not being fluent in Dutch, since language problems would harm the validity and reliability of collected data [34]. The study protocol was approved centrally by the Ethical Review Board of the VU University Medical Centre and subsequently by local review boards of each participating center. After full verbal and written information about the study, written informed consent was obtained from all participants at the start of baseline assessment [34].

6.2.2 Measures

6.2.2.1 Demographics and clinical features

The Composite International Diagnostic Interview (CIDI WHO version 2.1) was used to assess the presence of depressive- and anxiety disorders according to the Diagnostic and Statistical Manual of Mental Disorders – Fourth Edition (DSM-IV) at baseline, after 2-, 4-, 6- and 9 years. These included dysthymia, MDD, social phobia, panic disorder, agoraphobia, generalized anxiety disorder, and lifetime anxiety disorder. The CIDI is a fully standardized diagnostic interview with validated psychometric characteristics [33, 35].

Baseline demographic variables included gender, age, ethnicity (yes/no from north European heritage), level of education (i.e., elementary or less; general intermediate or secondary education; college or university), BMI, illness prior to interview, chronic somatic diseases, and anti-inflammatory medication. BMI was calculated by dividing weight (kg) by squared height (m^2). Patients were asked about illness (e.g., a mild cold or fever) prior to interview. A wide variety of diseases were assessed through a self-report questionnaire, asking for the presence of 20 common chronic diseases including asthma, chronic bronchitis or pulmonary emphysema, heart diseases or infarct, diabetes, stroke or CVA, arthritis or arthrosis, rheumatic complaints, tumor and/or metastasis, stomach or intestinal disorders, liver disease or liver cirrhosis, epilepsy, thyroid gland disease, or another chronic disease for which the patient receives treatment. A count was made of the chronic diseases for which a person reported receiving treatment. More details regarding this variable can be found elsewhere [36]. Anti-inflammatory medication use (ATC codes M01A, M01B, A07EB, A07EC) was based on inspection of medication containers (further referred to as anti-inflammatory medication).

6.2.2.2 Basal and LPS-induced inflammatory markers

Inflammatory markers C-reactive protein (CRP), IL-6 and TNF-α were determined from fasting morning blood plasma at baseline. After an overnight fast, 50 ml blood was drawn which was immediately transferred to a local laboratory and kept frozen at -80 °C. High-sensitivity plasma levels of CRP were measured in duplicate by an in-house high-sensitivity enzyme linked immunosorbent assay (ELISA) based on purified protein and polyclonal anti-CRP antibodies (Dako, Glostrup, Denmark). The lower detection limit of CRP is 0.1 mg/l and the sensitivity is 0.05 mg/l. Intra- and interassay coefficients of variation were 5% and 10%, respectively. Plasma IL-6 levels were measured in duplicate by a high-sensitivity ELISA

(PeliKine CompactTM ELISA, Sanquin, Amsterdam, the Netherlands). The lower detection limit of IL-6 is 0.35 pg/ml and the sensitivity is 0.10 pg/ml. Intra- and interassay coefficients of variation were 8% and 12%, respectively. Plasma TNF-α levels were assayed in duplicate using a high-sensitivity solid phase ELISA (Quantikine HS Human TNF-α Immunoassay, R&D systems, Minneapolis, MN, USA). The lower detection limit of TNF-α is 0.10 pg/ml and the sensitivity is 0.11 pg/ml. Intra- and interassay coefficients of variation were 10% and 15%, respectively. As done before [32], we created an overall basal inflammation index, as we assumed that high inflammatory marker levels in multiple markers are the best indication of general low-grade inflammation. The basal inflammation index consisted out of the mean value of all 3 $\log_e$-transformed (due to their positively skewed distributions) and standardized markers.

The innate immune response of 12 cytokines and inflammatory markers was assessed in blood that was *ex vivo* stimulated with LPS at basline. Serial venous whole blood samples were obtained at baseline in a 7-ml heparin-coated tube (Greiner Bio-one, Monroe, NC, USA). Between 10 and 60 minutes after blood draw, 2.5 ml of blood was transferred into a PAXgene tube (Qiagen, Valencia, CA, USA). Remaining blood (4.5 ml) was stimulated by addition of LPS (10 ng ml − 1 blood; Escherichia coli, Sigma, St. Louis, MO, USA), as done by others [26]. LPS-stimulated samples were laid flat and incubated at a slow rotation for 5–6 h at 37 °C. A 2.5-ml sample of this LPS-stimulated blood was transferred into a PAXgene tube. This LPS procedure was carried out at four laboratories (Amsterdam, Leiden, Groningen, Heerenveen). Remaining plasma (±0.5 ml) was kept frozen at − 80 °C for later analysis.

Levels of interferon (IFN)-γ, IL-2, IL-6, IL-8, IL-10, IL-18, monocyte chemotactic protein-1 (MCP-1), macrophage inflammatory protein (MIP)-1α, MIP-1β, matrix metalloproteinase-2 (MMP2), and TNF-α were assayed simultaneously for all available samples, using a multi-analyte profile (Human CytokineMAP A v 1.0; Myriad RBM, Austin, TX, USA). This commercial platform adheres to stringent guidelines of quality control and has Clinical Laboratory Improvement Amendments (CLIA) approval, which means that the platform is validated and calibrated on a continuous basis. Cytokines were $\log_e$-transformed to normalize their positively skewed distributions.

In order to reduce the number of statistical tests and because we did not have specific hypotheses about individual inflammation markers, we used exploratory factor analysis (EFA)

with Principal Axis Factoring and Oblimin rotation to examine dimensionality of the 12 inflammatory markers, that yielded two LPS-induced inflammation indexes, as previously described [32, 37]. The two LPS-induced inflammation indexes are further referred to as LPS-induced inflammation index-1 and LPS-induced inflammation index-2. Markers IFN-γ, IL-10, IL-2, IL-6, MMP-2, TNF-α, and TNF-β loaded on LPS-induced inflammation index-1 with factor loadings between 0.41 and 0.88 and a raw alpha of 0.86. IL-8, IL-18, MCP-1, MIP-1α, and MIP-1β loaded on LPS-induced inflammation index-2 with factor loadings between 0.34 and 0.94 and a raw alpha of 0.89. Together with the basal inflammation index, these indexes were considered the main independent variables of interest.

6.2.2.3 Anxiety symptoms

The Beck's Anxiety inventory ([BAI; 38], the Fear Questionnaire [FQ; 39], and the Penn State Worry Questionnaire [PSWC; 40], as well as its subscale scores, were used as the outcome measures for severity of anxiety symptoms over time. These measures capture different aspects of, but is not exclusive for, anxiety disorders such as symptoms of arousal (BAI), avoidance (FQ), and worry (PWSQ). These constructs are common in panic disorders, common phobias, and generalized anxiety disorder among others.

The BAI is a self-report questionnaire which assesses common symptoms of anxiety such as fear of dying, fear of losing control and nervousness [38]. It consists of 21 equally weighted items, rated on a 4-point scale, ranging from 0 (not at all) to 3 (severely, "I could barely stand it"). The BAI is scored by summing the ratings for all of the 21 symptoms to obtain a total score that can range from 0 to 63. It contains a Somatic subscale (14 items) and a subjective subscale (7 items), representing physical- and cognitive symptoms of anxiety [41]. The reliability and validity of the BAI are well-established [38, 42]. Research has showed adequate reliability estimates for the BAI in a sample of psychiatric inpatients ($\alpha = 0.92$) and high school adolescents [$\alpha = 0.88$; 43]. In our study, the Cronbach's alpha coefficient was $\alpha = 0.93$ at baseline.

The 15-item Fear Questionnaire (FQ) is a self-report instrument that assesses the level of avoidance in relation to common phobias, including social phobia (five items), agoraphobia (five items), and hematophobia/traumatophobia [five items; 39]. It consists of 15 equally weighted items, rated on a 9-point scale, ranging from 0 ("Would not avoid it") to 8 ("Always

avoid it"). The sum-score ranges from 0 through 120. Three phobia subscales of five items can be derived, a blood phobia subscale, a social phobia subscale, and a agoraphobia subscale. The psychometric properties of the FQ has been researched in multiple studies among both non-clinical populations [44] and patients with an anxiety disorder [45, 46]. These studies conclude that the psychometric properties of the FQ are sufficient with moderate to high Cronbach's alpha coefficients per subscale, ranging from $\alpha = 0.71$ to $\alpha = 0.83$ [45, 46]. In our study, the Cronbach's alpha coefficient was $\alpha = 0.88$ at baseline.

The Penn State Worry Questionnaire (PSWQ) is also a self-report questionnaire which consists of 16 equally weighted items rated on a 5-point scale (1-5) with 1 meaning "not at all typical of me" to 5 "very typical of me". The total score ranges from 16 to 80. This 16-item instrument emerged from factor analysis of a large number of items, and was found to possess high internal consistency and good test-retest reliability [40]. The psychometric properties of the PSWQ were considered satisfactory in a community sample [47] and a sample of anxiety patients [48]. Cronbach's alpha coefficients of 0.94 were found in a community sample [47], and ranging from 0.86 to 0.93 in a clinical sample [48]. In our study, the Cronbach's alpha coefficient was $\alpha = 0.96$ at baseline.

6.2.3 Statistical analysis

A multivariate linear mixed model was used with BAI, FQ, PSWQ total- and subscale scores at baseline, and after 1, 2, 4, 6, and 9 yearsas outcome variables and baseline inflammatory indexes as the main independent variables. PSWQ was not assessed at 1 year. Because of the heterogeneity of our sample (both healthy participants as well as anxiety patients at baseline), random intercepts and slopes were added, as they resulted in a significantly better fit compared to model without random effects, as tested with -2LL ratio tests. Adding an interaction between a continuous modelled time variable and inflammatory markers resulted in a minimal increase of model fit and was therefore not included in the main analyzes, but instead was added as a sensitivity analysis of which the results were included in the supplementary material. This resulted in mixed models which assessed whether participants with elevated level of inflammation were more likely to have higher symptom-levels of anxiety at baseline and throughout a follow-up period of up to nine years. Models were adjusted for baseline variables of gender, age, reported sickness prior to interview, the use of anti-inflammatory medication, and BMI.

Analyzes were done separately for each of the three inflammatory index scores as main independent variables and as exploratory analysis for each of the individual markers. In sensitivity analyzes, we repeated the analysis in which we adjusted for the presence of (comorbid) MDD (about 35.4% of the total sample) as a dichotomous variable. Moreover, in a sensitivity analysis we repeated the analyzes in a subsample of participants who met DSM-IV criteria for an anxiety disorder (see supplementary material figure 1). For the main analyzes with the index scores, we adjusted the outcomes of the inflammation indexes for multiple testing with the Bonferroni-correction which resulted in p-values regarded as being significant at p = 0.001 [49]. In order to yield beta-coefficients, that can be compared among different tests, all outcome and independent variables were standardized (i.e., z-scores). For all analyzes, we used RStudio (R version 3.6.3; R Foundation for Statistical Computing, Vienna, Austria, 2016. URL: https://www.R-project.org/), with main packages 'lme4' (version 1.20.1), and 'emmeans' (version 1.4.6).

6.3 Results

6.3.1 Sociodemographic and clinical characteristics at baseline

The demographics of the study sample are shown in Table 1. Our study sample was 66.5% female (n=1,930), and the age ranged from 18 through 64 years at baseline (mean 41.9 years; SD 13.0; see also table 1A). As demonstrated in table 1B, at baseline a total of 1,299 (43.6%) of the participants had an anxiety disorder in the month prior to the baseline wave, of which social phobia (18.5%) was most common. There were also 27.1% patients with (comorbid) MDD (n=796). Of the total sample, 47.9% did not have a mood or anxiety diagnosis (n=1,368) of whom 54.2% never had had a psychiatric diagnosis before (n = 742). As a considerable percentage of the sample was recruited from general practice and secondary mental healthcare, percentages of patients meeting DSM criteria for anxiety or mood disorders were the highest at baseline and decreased at later follow-ups, most likely due to symptoms naturally resolving over time and by means of treatment, as well as due to regression to the mean effects.

Table 1. A. Sociodemographic and clinical characteristics		
	Whole sample n = 2867 Baseline	**LPS-induced sample n = 1227 Baseline**
Age in years (mean, SD)	41.9 (13.0)	42.8 (12.7)
Female (%)	66.5	65.6
North European ethnicity (%)	94.9	94.8
BMI (mean, SD)	25.6 (5.0)	25.67 (5.0)
Smoking status (%)		
Never smoker	28.0	29.0
Former smoker	33.6	34.2
Current smoker	38.4	36.8
Education level (%)		
Elementary or lower	6.5	6.4
Secondary education	58.2	56.7
College or university	35.4	36.9
Sickness prior to interview (%)	27.9	30.1
Chronic somatic disease, yes (%)	40.4	44.3
Anti-inflammatory med., yes (%)	4.9	3.1
Inflammatory markers (median, IQR)		
TNF-α (pg/ml)	0.80 (0.50)	
IL-6 (pg/ml)	0.80 (0.76)	
CRP (mg/L)	1.22 (2.48)	
Inflammatory markers after LPS induction (median, IQR)		
IFN-γ (pg/ml)		10.2 (7.44)
IL-10 (pg/ml)		205.5 (281.75)
IL-18 (pg/ml)		249.0 (104.0)
IL-2 (pg/ml)		9.07 (6.17)
IL-6 (ng/ml)		25800 (17875)
IL-8 (ng/ml)		10400 (8500)
MCP-1 (ng/ml)		1510 (1270)
MIP-1α (ng/ml)		17800 (12975)
MIP-1β (ng/ml)		234000 (146500)
MMP-2 (pg/ml)		73.0 (20.40)

Table 1. B. Sociodemographic and clinical characteristics

	Whole sample					LPS-induced sample				
	n = 2867	n = 2529	n = 2338	n = 2195	n = 2014	n = 1227	n = 1051	n = 955	n = 893	n = 818
	Baseline	T2	T4	T6	T9	Baseline	T2	T4	T6	T9
MDD, yes (%)	27.1	13.8	11.3	9.7	9.8	28.8	15.2	10.1	8.4	9.2
Dysthymia, yes (%)	9.3	8.3	6.1	6.1	4.3	10.4	8.2	6.4	6.1	4.0
Anxiety disorder, yes (%)	43.6	27.5	22.7	19.8	19.5	44.4	27.9	20.3	20.9	19.4
Social phobia, yes (%)	18.5	17.5	17.5	16.7	16.6	20.4	19.8	17.9	17.3	17.3
Panic disorder, yes (%)	17.0	15.9	15.7	14.7	14.9	17.3	16.5	14.6	14.5	13.3
Agoraphobia, yes (%)	17.1	16.0	155.6	14.7	14.8	17.0	16.1	13.6	13.0	13.1
General anxiety disorder, yes (%)	13.3	12.2	12..1	11.8	11.8	14.5	13.5	12.3	12.2	11.3
Comorbid mood and anxiety disorder (%)	19.9	10.1	8.0	6.7	6.4	21.4	10.8	8.3	6.8	5.9
No current anxiety or mood disorder (%)	47.9	66.2	71.7	74.7	75.6	46.9	70.4	66.5	68.8	66.7

Table 1. Sociodemographic and clinical characteristics. T1 (year 1; n = 2388) included only self-report measures , it was therefore included in the study but not included in the Table 1. Anti-inflammatory medication included ATC codes M01A, M01B, A07EB, A07EC. Chronic somatic diseases included: asthma, chronic bronchitis or pulmonary emphysema, heart diseases or infarct, diabetes, stroke or CVA, arthritis or arthrosis, rheumatic complaints, tumor and/or metastasis, stomach or intestinal disorders, liver disease or liver cirrhosis, epilepsy, thyroid gland disease, or another chronic disease for which the patient receives treatment. Tumor necrosis factor = TNF. Interleukin = IL. C-reactive protein = CRP. Interferon-y = IFN-y. Higher monocyte chemoattractant protein-1 = MCP-1. Macrophage inflammatory protein = MIP. Matrix metallopeptidase-2 = MMP-2.

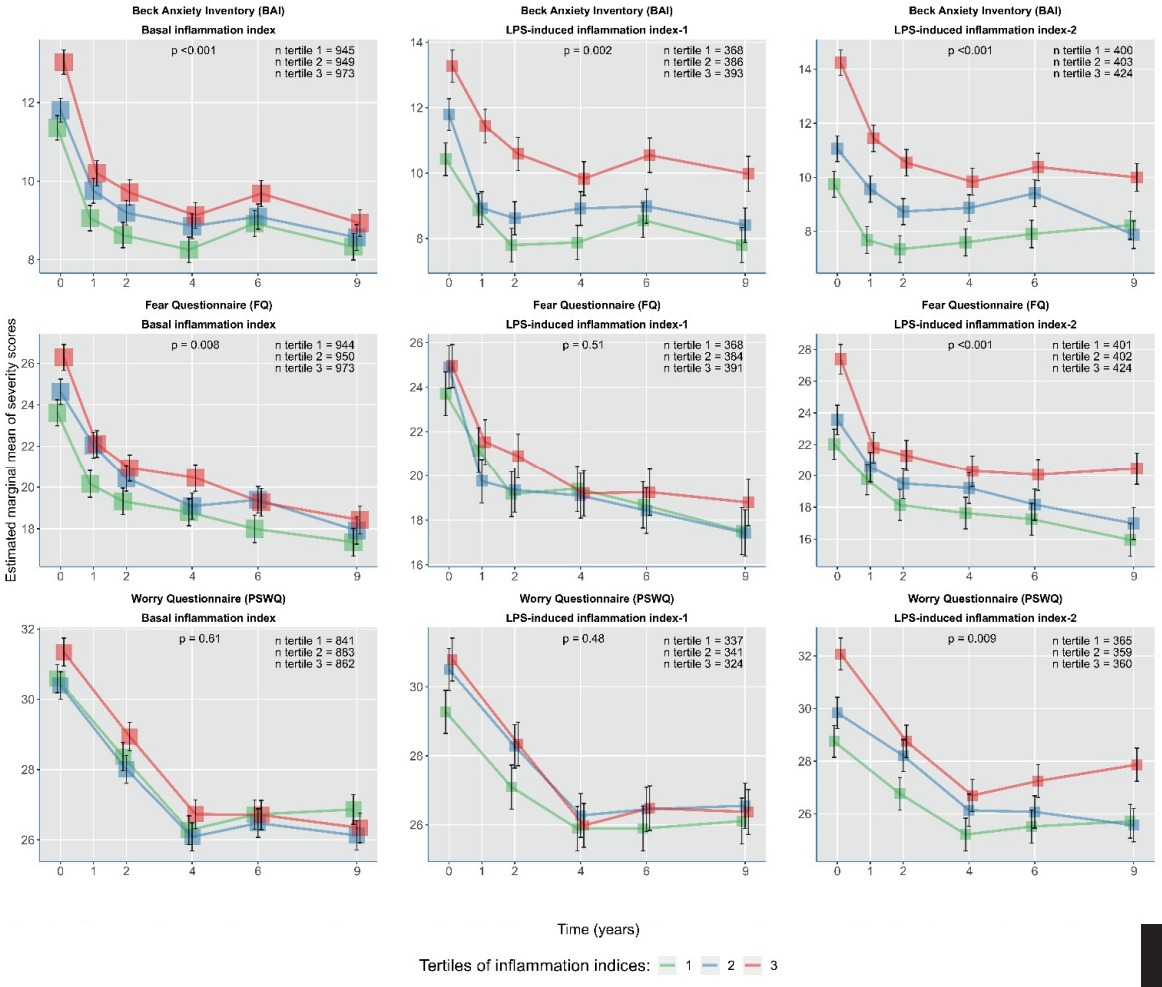


Figure 1. Tertiles of the basal inflammation index, LPS-induced inflammation index-1, and LPS-induced inflammation index-2 related to BAI, FQ, and PSWQ total scores over the course of 9 years. Inflammation indexes are divided into tertiles of equal proportions of the sample distribution at baseline (1. lowest inflammatory markers: 0.0 – 0.33; 2. middle: 0.33 – 0.66; 3. highest: 0.66 – 1.0). Sample sizes for each tertile at baseline are presented in the graphs. Sample sizes can vary due to missing individual variables of inflammatory markers and anxiety totals scores. Y-axis represents estimated marginal mean values of total scores adjusted for gender, age, reported sickness prior to interview, the use of anti-inflammatory medication, and BMI. Error bars represent standard errors.

6.3.2 Basal inflammation

The associations between basal inflammation index score in relation to anxiety symptom severity over the course of 9 years are shown in Figure 1 and Figure 2 (first column). Basal level of inflammation was significantly positively associated to BAI total score (β = 0.057, p = < 0.001) and its somatic subscale (β = 0.070, p = <0.001). This translates as a 0.057 SD increase of (BAI) anxiety severity with each SD increase of the basal inflammation index. Basal inflammation was also significantly associated to the FQ agoraphobia subscale (β = 0.074, p < 0.001). Additionally, significant associations were found for the FQ total score (β = 0.048, p = 0.008), although this was no longer significant after adjusting for multiple testing. Similar effects were found when only a subsample of participants who met DSM-IV criteria for an anxiety disorder were included (see supplementary material figure 1). Significant associations were present at baseline and tended to persist over the course of nine years, as shown in Figure 1. This was further confirmed by small effect sizes of the interaction terms with time (with a maximum β = -0.006 ;p = 0.009), which was not statistically significant when adjusted for multiple testing (Supplementary material table 1). No significant associations were found between basal inflammation and the BAI subjective subscale (β = 0.029, p = 0.084), the FQ social phobia subscale (β = 0.019, p = 0.2882), and the PSWQ scale (β = 0.009, p = 0.610).

After adjustment for the presence of MDD, we found that the effect estimates of basal inflammation with anxiety severity were attenuated by 25-30%, but remained statistically significant. When assessing the individual inflammatory markers of the basal index score, we found that TNF-α, IL-6, and CRP were related to anxiety with roughly equal effect sizes, although no longer statistically significant (see Figure 3).

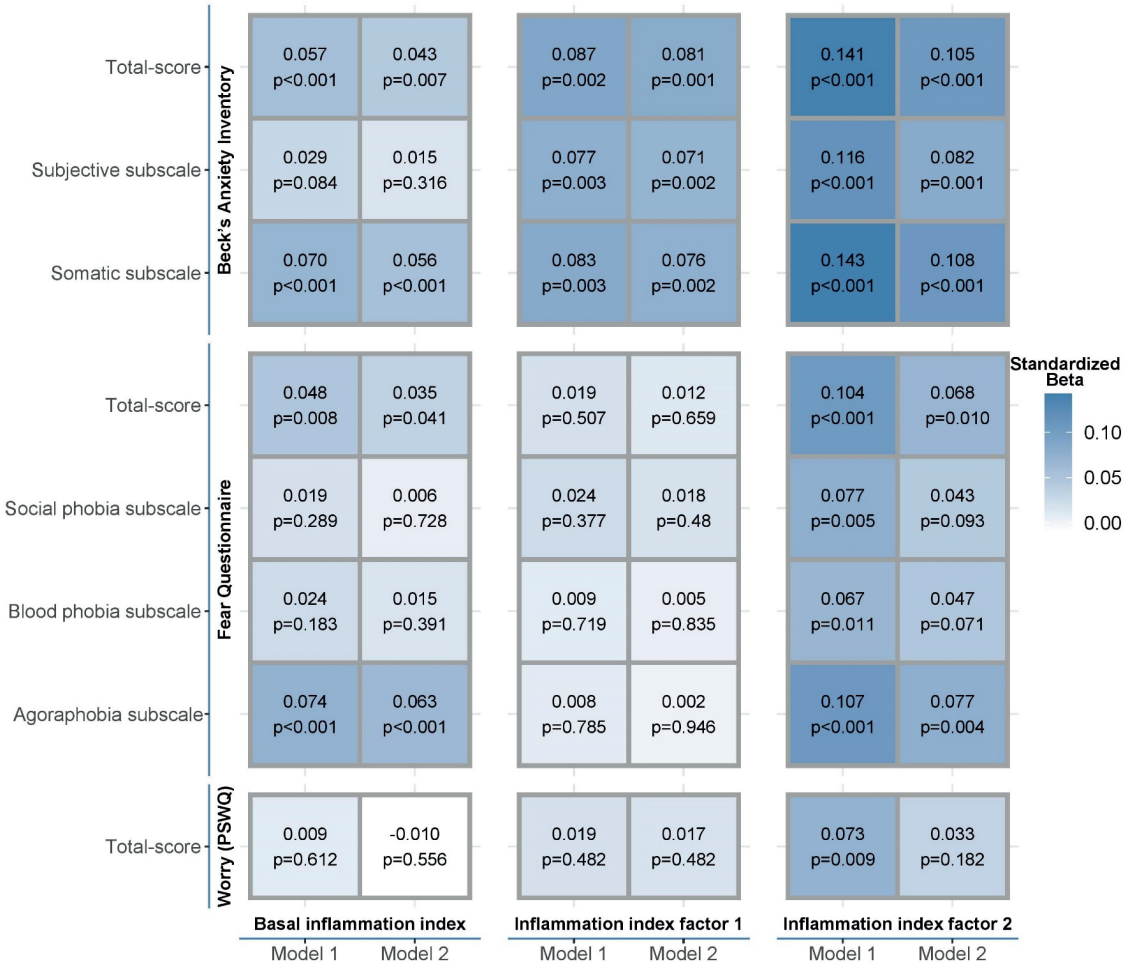


Figure 2. **Standardized beta coefficients of the association between basal and LPS-induced inflammation indexes and anxiety symptoms.** Linear mixed models fitted with repeated measures, which were assessed up to six times over 9 years of follow-up. P-values remain statistically significant after adjusting for multiple testing at p = 0.001. Model 1 standardized beta coefficients for basal inflammation index and LPS-induced inflammation index-1 and -2 were adjusted for baseline variables of gender, age, sickness prior to interview, the use of anti-inflammatory medication, and BMI. Model 2 beta-coefficients were additionally adjusted for the presence of MDD.

6.3.3 LPS-induced inflammation

The associations between LPS-induced inflammation index – 1 in relation to anxiety over the course of 9 years are shown in Figure 1 and Figure 2 (middle column – index 1; last column index 2). LPS-induced inflammation index – 1 was significantly positively associated to the BAI total score (β = 0.087, p = 0.002), its somatic subscale (β = 0.083, p = 0.003) and subjective subscale (β = 0.077, p = 0.003). However, none of these associations with the BAI remained significant (p's > 0.001) after adjustment for multiple testing. LPS-induced inflammation index – 1 was not significantly associated to the FQ and PSWQ (sub)scales. When we adjusted these analyzes for the presence of MDD (comorbidity), we found that the (lack of) association of LPS-induced inflammation index – 1 remained roughly similar. When assessing the individual components of biomarkers of LPS-induced inflammation index – 1 , we found that there were significant positive associations between TNF-β, IL-2, IL-6 and MMP-2 and BAI total score, BAI somatic subscale, and BAI subjective subscale (see Figure 3).

Contrary to LPS-induced inflammation index – 1, LPS-induced inflammation index – 2 demonstrated significant associations with all BAI, FQ, and PWSQ (sub) scales. Standardized beta's ranged from β = 0.067, p = 0.011 (for FQ blood phobia) to β = 0.1, p <0.001 (for BAI somatic subscale). When adjusting for multiple testing, associations remained statistically significant for the BAI (sub) scales, and the FQ total score and agoraphobia subscale. Similar effects were found when only a subsample of participants who met DSM-IV criteria for an anxiety disorder were included (see supplementary material figure 1). As is demonstrated in Figure 1, these statistical associations were strongest at baseline, but persisted over time. We found a significant negative interaction term of up to β = -0.014 (p = < 0.001), between a continuous modelled time variable and LPS-induced inflammation index – 2 (Supplementary material table 1). This suggests that the relationship with baseline LPS-induced inflammation index – 2 tended to attenuate somewhat over time, although to a small degree.

Similar to basal inflammation, the association between LPS-induced inflammation index – 2 and the anxiety (sub) scales were attenuated by approximately 30%, when adjusted for the presence of MDD (comorbidity). When assessing the individual biomarkers that LPS-induced inflammation index – 2 consisted of, we found that all 5 markers were significantly related to these anxiety scales. However, the estimated associations of IL-8, IL-18, and MCP-1 were substantially stronger compared to those of MIP-1α and MIP-1β (see Figure 3).

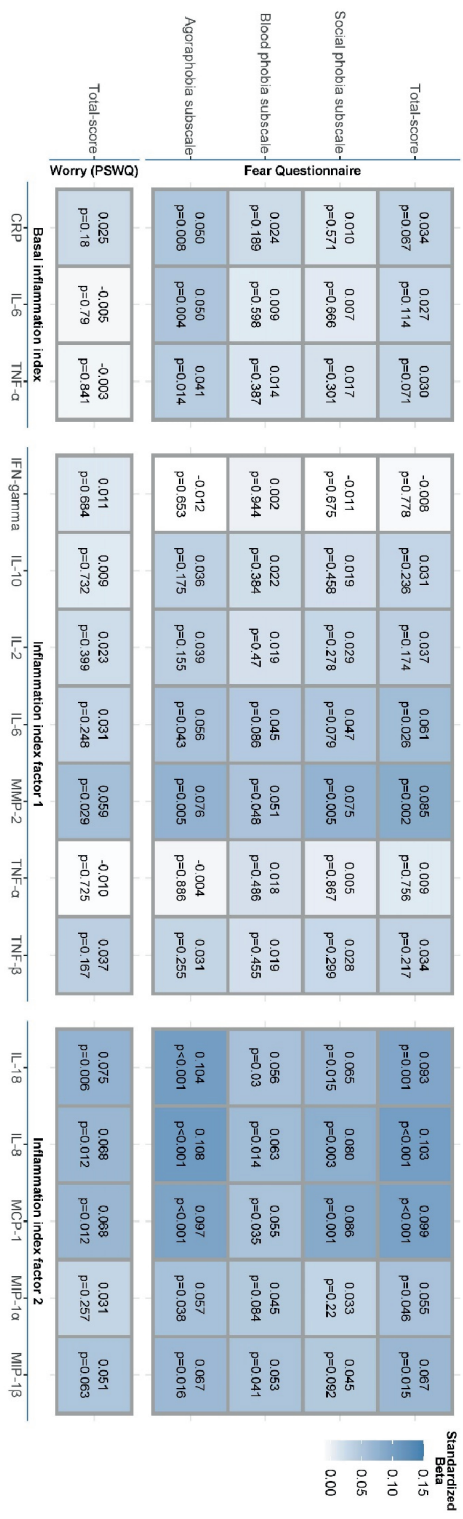

Beck's Anxiety Inventory
Somatic subscale
Subjective subscale
Total-score
Fear Questionnaire
Agoraphobia subscale
Blood phobia subscale
Social phobia subscale
Total-score
Worry (PSWQ)
Total score
Basal inflammation index
CRP
IL-6
TNF-α
Inflammation index factor 1
IFN-gamma
IL-10
IL-2
IL-6
MMP-2
TNF-α
TNF-β
Inflammation index factor 2
IL-18
IL-8
MCP-1
MIP-1α
MIP-1β
Standardized Beta
0.15
0.10
0.05
0.00

Figure 3. Standardized beta coefficients of the association between individual basal serum, and LPS-induced inflammatory markers and anxiety symptoms. Linear mixed models fitted with repeated measures, which were assessed up to six times over 9 years of follow-up. Standardized beta coefficients for basal inflammation index and LPS-induced inflammation index -1 and -2 were adjusted for baseline variables of gender, age, sickness prior to interview, the use of anti-inflammatory medication and BMI.

6.4 Discussion

Our study is the first to examine the relationship between basal as well as LPS-induced inflammatory markers with longitudinal measures of anxiety symptom severity over a period of up to nine years. Our results demonstrated that participants with elevated inflammatory markers at baseline had on average higher levels of anxiety at baseline, which persisted during the course of nine years follow-up. However, the effect sizes of these associations were small. Inflammatory markers were especially associated with somatic symptoms of anxiety (e.g., sensations of physical arousal), and symptoms of agoraphobia.

Thus far, most prospective studies examining the relationship between inflammation and anxiety used basal inflammatory markers such as CRP, TNF-α and IL-6 [18, 19, 22, 50]. We found stronger association for LPS-induced inflammatory markers index -2 with anxiety compared to the basal inflammatory index, which were assessed through distinct methods (ELISA versus multiplex). Earlier studies have shown that basal circulating levels of inflammatory markers (assessed by using Elisa method) are typically low and show a high degree of intra-individual variability [23]. The expression of inflammatory markers in response to ex vivo stimulation of LPS (using multiplex method) mimics the natural environment more closely and induces an inflammatory reaction reflecting the innate production of inflammatory markers [24, 26]. Our results underline the idea that basal inflammation levels and stimulated levels are a reflection of two different aspects of the immune system. LPS-induced inflammatory markers may show less (within person) variability compared to basal inflammatory markers serum level [27]. That being said, LPS-induced inflammatory index -1 demonstrated smaller effect sizes than LPS-induced inflammatory index-2, suggesting that LPS-induced inflammatory index -2 is made up of cytokines that may better reflect the innate immune response that is associated with anxious mood states than markers from index-1. Previously, similar results were found for this index score in relation to the course of symptoms of depression (van Eeden et al., 2020). Within LPS-induced inflammatory index -2, especially MCP-1, IL-8 and IL-18 demonstrated strong associations with anxiety. Cytokines are believed to play an important role in immune homeostasis and can display heterogenic, pleiotropic and overlapping functional properties as is illustrated by their ability to act in both a pro- and an anti-inflammatory manner in complex interactions with one another (Jones &

Jenkins, 2018). It appears that pro-inflammatory markers (e.g., IL-8 and IL-18) may contribute more than anti-inflammatory markers (e.g., IL-10; as shown in Figure 3), but such findings need to be replicated as our prospective results contrast with those by Vogelzangs et al., (2016) showing that both pro- and anti-inflammatory markers were positively associated with anxiety and depression in a cross-sectional analysis.

According to the "pathogen-host defence theory" (PATHOS-D), across evolutionary time, heavy pathogen load induced significant pressure on human survival [51]. This has led to adaptations which shaped interactions between the immune system and the brain, resulting in a set of behaviors such as anhedonia and fatigue (commonly referred to as *sickness behavior*), but also anxiety arousal and alarm [51]. First, due to these processes, modern humans may have inherited a genomic bias towards inflammation, because this response - and the symptoms it promotes - enhanced protective behaviors, host survival, and reproduction in the highly pathogenic environment in which humans evolved [52]. Second, stress perception by the brain may serve as an early warning signal to activate the immune system in preparation of subsequent wounding [52, 53], in which case symptoms of anxiety would lead to an increase of inflammatory markers. Finally, our findings could also be explained in light of the "sickness behavior theory" [54, 55], which is part of the PATHOS-D theory. The sickness behavior theory postulates that crosstalk between several inflammatory pathways and neurocircuits of the hypothalamic–pituitary–adrenal (HPA) axis could lead to sickness behavior—a set of motivational and behavioral changes including both somatic symptoms (low energy, malaise, etc.) and reward sensitivity related symptoms [anhedonia, and withdrawal; 52, 56, 57]. We found relative strong associations with agoraphobia, which supports this idea.

Alongside the PATHOS-D and sickness behavior theories (Dhabhar, 2009; Miller & Raison, 2016), the associations found between inflammation and anxiety symptoms, in particular the arousal anxiety symptoms may be explained by activation of the HPA axis. Replicated studies have demonstrated that following acute stress, cytokines such as IL-6 and TNF-α activate the HPA-axis, increasing levels of corticotrophin releasing hormone (CRH), adrenocorticotropic hormone (ACTH) and cortisol (Beishuizen & Thijs, 2003). Repeated activation of the inflammatory system due to chronic stress has been shown to disproportionally increase HPA-axis activity compared to the usual response (Grinevich et al., 2001), which in turn have been

shown to induce mood and anxiety symptoms (Leonard & Myint, 2009). Although our results may partially be explained by HPA-axis activation this seems not to be the most important mechanism as IL-6 and TNF-α were relatively weakly related to anxiety.

An additional finding of the present study was that a substantial part of this association was driven by MDD comorbidity, as the strength of the relationship between inflammatory markers and anxiety symptomatology attenuate by about 25% to 30% when adjusted for the presence of MDD. Although comorbid MDD may also be an indicator for overall severity, the findings of this study seem to replicate that the link found between anxiety and inflammation is partly driven by depression [18, 19, 22, 32, 50].

Considering the positive association of several inflammatory markers with anxiety, opportunities may arise for developing treatment options. Several meta-analyzes have found predominantly positive effects of anti-inflammatory medication (NSAIDs, fatty acids, statins and cytokine inhibitors amongst others) on depression [58-62]. Anti-inflammatory treatment may result in a decrease of depressive symptoms, but likely only for a subset of patients with chronic low-grade inflammation [32]. For example, there is some evidence for efficacy of add-on treatment with minocycline for treatment resistant depression, but only among those with low-grade inflammation defined as CRP ≥3 mg/L [63]. Perhaps anti-inflammatory drugs can also be used for treating some patients with anxiety, especially those with elevated (LPS-induced) inflammatory markers and who suffer from somatic anxiety symptoms or agoraphobia. It could be promising to devise strategies to identify such a subgroup of patients with anxiety disorders that may benefit from a (personalized) treatment with anti-inflammatory drugs.

Our study has several strengths. With a substantial sample size, we analyzed individual symptom domains of anxiety over a follow-up period of nine years. A wide array of inflammatory markers was assessed at baseline, including more costly and laborious LPS-induced markers. Moreover, we had a heterogenic sample containing patients with anxiety disorders as well as healthy controls recruited from multiple settings and with only few exclusion criteria, making this sample easier to generalize to other populations.

A number of limitations of our study need to be discussed. Firstly, we found no strong effects of interaction terms with time, but rather that baseline associations persisted over a long

follow-up period. Therefore, our findings cannot disentangle the relationships in time, whether inflammation predated anxiety or vice versa. Moreover, an earlier study demonstrated that comorbid depressive and anxiety disorders and higher symptom severity were associated to attrition, which could have been a potential bias in our analyzes. However, we do not expect large confounding effects with regard to our findings, when doing a sensitivity analysis with a subset of complete cases (n = 1713), the relationships between basal inflammation index and the BAI total score did somewhat increase in effect size, and remained statistically significant (β=0.057; p=0.011). Second, we focused on a dimensional approach of anxiety symptoms based on self-report severity scales, which differs from clinician-rated categorical DSM diagnoses of anxiety disorders. Anxiety DSM-diagnoses can be viewed as discrete categorical syndromes imposed on a continuum of anxiety symptoms of varying severity and duration (Kendler & Gardner Jr, 1998). Future research could assess whether inflammatory markers are also related to onset and remission of diagnoses over several years. Moreover, NESDA focussed on depression and anxiety and patients with other diagnoses have not been invited for the NESDA project. Although clinically overt diagnoses, such as bipolar disorder and severe PTSD were excluded, our sample was not diagnostically homogeneous. Future research with homogenic samples and clinician-rated DSM criteria are needed. Third, a large proportion of our sample had a prevalent chronic somatic condition, although detailed information of the nature of these conditions was lacking. We choose not to adjust our analyzes for somatic comorbidity, because the consequent pro-inflammatory state could be part of the causal pathway between inflammation and anxiety. However, when we adjusted the effects of basal inflammatory markers on BAI total score for the presence of a chronic somatic disorder (yes/no), the effect was only slightly reduced and remained significant (β = 0.049, p=0.004). Future research should examine if inflammation is a mediating factor for the relationship between many chronic somatic diseases and anxiety (Costello, Gould, Abrol, & Howard, 2019; Renna, O'Toole, Spaeth, Lekander, & Mennin, 2018). Fourth, due to logistical reasons, LPS-stimulated markers were only added to the study, after the inclusion was well underway, resulting in a smaller sample of 1,229 participants. Fortunately, the sample size was still reasonably large and was not substantially different with regard to baseline characteristics. Fifth, as more LPS-induced markers compared to basal serum markers were assessed, the results could be biased toward identifying relationships with one methods over the other. Sixth, the two LPS indued inflammatory indexes were

calculated based on data driven methods [Factor analysis; 37], as was done in our earlier research [32]. An alternative option would have been grouping of these individual markers based on underlying pro- and anti-inflammatory properties. Finally, our inflammatory markers were based on a single blood sample only. Sequential day-to-day measures of inflammatory markers would have increased the precision of the markers.

In conclusion, we found that participants with high levels of inflammatory markers have on average high levels of somatic symptoms of anxiety (arousal) and agoraphobia, which tended to persist over a period of nine years, albeit with small effect sizes. These associations were partly driven by co-morbid depression. These findings suggest that some of these patients could benefit from anti-inflammatory agents. Future studies are needed to develop strategies in order to select these patients and to test treatment effectiveness. The small effect sizes found in this study suggest that a large impact on group level may not be feasible.

Acknowledgements and financial disclosures

The infrastructure for the NESDA study (*www.nesda.nl*) is funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental health care organizations (Amsterdam University Medical Centers (location VUmc), GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum). BP has received (non-related) research funding from Boehringer Ingelheim and Jansen Research. All remaining authors report no biomedical financial interests or potential conflicts of interest.

References

1. Pitsavos, C., et al., *Anxiety in relation to inflammation and coagulation markers, among healthy adults: the ATTICA study.* Atherosclerosis, 2006. **185**(2): p. 320-6.
2. Salim, S., G. Chugh, and M. Asghar, *Inflammation in anxiety.* Adv Protein Chem Struct Biol, 2012. **88**: p. 1-25.
3. Vogelzangs, N., et al., *Anxiety disorders and inflammation in a large adult cohort.* Transl Psychiatry, 2013. **3**: p. e249.
4. Naude, P.J.W., et al., *Anxiety disorders and CRP in a population cohort study with 54,326 participants: The LifeLines study.* World J Biol Psychiatry, 2018. **19**(6): p. 461-470.
5. Renna, M.E., et al., *The association between anxiety, traumatic stress, and obsessive–compulsive disorders and chronic inflammation: A systematic review and meta-analysis.* Depression and anxiety, 2018. **35**(11): p. 1081-1094.
6. Costello, H., et al., *Systematic review and meta-analysis of the association between peripheral inflammatory cytokines and generalised anxiety disorder.* BMJ open, 2019. **9**(7): p. e027925.
7. Foley, P. and C. Kirschbaum, *Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings.* Neuroscience & Biobehavioral Reviews, 2010. **35**(1): p. 91-96.
8. Pace, T.W., et al., *Increased stress-induced inflammatory responses in male patients with major depression and increased early life stress.* Am J Psychiatry, 2006. **163**(9): p. 1630-3.
9. Bierhaus, A., et al., *A mechanism converting psychosocial stress into mononuclear cell activation.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1920-5.
10. Leonard, B.E., *The HPA and immune axes in stress: the involvement of the serotonergic system.* Eur Psychiatry, 2005. **20 Suppl 3**: p. S302-6.
11. de Kloet, C.S., et al., *Assessment of HPA-axis function in posttraumatic stress disorder: pharmacological and non-pharmacological challenge tests, a review.* J Psychiatr Res, 2006. **40**(6): p. 550-67.
12. Michopoulos, V., et al., *Inflammation in Fear- and Anxiety-Based Disorders: PTSD, GAD, and Beyond.* Neuropsychopharmacology, 2017. **42**(1): p. 254-270.
13. Capuron, L., et al., *Neurobehavioral effects of interferon-alpha in cancer patients: phenomenology and paroxetine responsiveness of symptom dimensions.* Neuropsychopharmacology, 2002. **26**(5): p. 643-52.
14. Roest, A.M., et al., *Anxiety and risk of incident coronary heart disease: a meta-analysis.* J Am Coll Cardiol, 2010. **56**(1): p. 38-46.
15. Musselman, D.L., et al., *Paroxetine for the prevention of depression induced by high-dose interferon alfa.* N Engl J Med, 2001. **344**(13): p. 961-6.
16. Belem da Silva, C.T., et al., *Cytokine Levels in Panic Disorder: Evidence for a Dose-Response Relationship.* Psychosom Med, 2017. **79**(2): p. 126-132.
17. Khandaker, G.M., et al., *Association between serum C-reactive protein and DSM-IV generalized anxiety disorder in adolescence: Findings from the ALSPAC cohort.* Neurobiol Stress, 2016. **4**: p. 55-61.
18. Wagner, E.-Y.N., et al., *Evidence for chronic low-grade systemic inflammation in individuals with agoraphobia from a population-based prospective study.* PloS one, 2015. **10**(4).
19. Glaus, et al., *The bidirectional relationship between anxiety disorders and circulating levels of inflammatory markers: Results from a large longitudinal population-based study.* Depression and anxiety, 2018. **35**(4): p. 360-371.
20. Liukkonen, T., et al., *The association between anxiety and C-reactive protein (CRP) levels: results from the Northern Finland 1966 birth cohort study.* Eur Psychiatry, 2011. **26**(6): p. 363-9.
21. Song, Y., et al., *Disturbance of serum interleukin-2 and interleukin-8 levels in posttraumatic and non-posttraumatic stress disorder earthquake survivors in northern China.* Neuroimmunomodulation, 2007. **14**(5): p. 248-54.
22. Baune, B.T., et al., *Inflammatory biomarkers predict depressive, but not anxiety symptoms during aging: the prospective Sydney Memory and Aging Study.* Psychoneuroendocrinology, 2012. **37**(9): p. 1521-1530.
23. van den Biggelaar, A.H., et al., *Inflammation and interleukin-1 signaling network contribute to depressive symptoms but not cognitive decline in old age.* Exp Gerontol, 2007. **42**(7): p. 693-701.

24. van der Linden, M.W., et al., *Determination of tumour necrosis factor-alpha and interleukin-10 production in a whole blood stimulation system: assessment of laboratory error and individual variation.* J Immunol Methods, 1998. **218**(1-2): p. 63-71.
25. Westendorp, R., et al., *Genetic influence on cytokine production in meningococcal disease.* The Lancet, 1997. **349**(9069): p. 1912-1913.
26. van Exel, E., et al., *Vascular factors and markers of inflammation in offspring with a parental history of late-onset Alzheimer disease.* Archives of general psychiatry, 2009. **66**(11): p. 1263-1270.
27. Üçeyler, N., W. Häuser, and C. Sommer, *Systematic review with meta-analysis: cytokines in fibromyalgia syndrome.* BMC musculoskeletal disorders, 2011. **12**(1): p. 245.
28. Duivis, H.E., et al., *Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: findings from the Netherlands Study of Depression and Anxiety (NESDA).* Psychoneuroendocrinology, 2013. **38**(9): p. 1573-85.
29. Vogelzangs, N., et al., *Cytokine production capacity in depression and anxiety.* Transl Psychiatry, 2016. **6**(5): p. e825.
30. Gaspersz, R., et al., *The role of anxious distress in immune dysregulation in patients with major depressive disorder.* Transl Psychiatry, 2017. **7**(12): p. 1268.
31. Lamers, F., et al., *Longitudinal association between depression and inflammatory markers: results from the Netherlands study of depression and anxiety.* Biological psychiatry, 2019. **85**(10): p. 829-837.
32. van Eeden, W.A., et al., *Basal and LPS-stimulated inflammatory markers and the course of individual symptoms of depression.* Translational Psychiatry, 2020. **10**(1): p. 235.
33. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* International Journal of Methods in Psychiatric Research, 2008. **17**(3): p. 121-40.
34. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* Int J Methods Psychiatr Res, 2008. **17**(3): p. 121-40.
35. Wittchen, H.-U., *Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review.* Journal of Psychiatric Research, 1994. **28**(1): p. 57-84.
36. Gerrits, M.M., et al., *The impact of chronic somatic diseases on the course of depressive and anxiety disorders.* Psychother Psychosom, 2013. **82**(1): p. 64-6.
37. Bandalos, D.L. and S.J. Finney, *Factor analysis: Exploratory and confirmatory*, in *The reviewer's guide to quantitative methods in the social sciences*. 2018, Routledge. p. 98-122.
38. Beck, A.T., R.A. Steer, and M.G. Carbin, *Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation.* Clinical psychology review, 1988. **8**(1): p. 77-100.
39. Marks, I.M. and A.M. Mathews, *Brief standard self-rating for phobic patients.* Behaviour research and therapy, 1979. **17**(3): p. 263-267.
40. Meyer, T.J., et al., *Development and validation of the Penn State Worry Questionnaire.* Behav Res Ther, 1990. **28**(6): p. 487-95.
41. Kabacoff, R.I., et al., *Psychometric properties and diagnostic utility of the Beck Anxiety Inventory and the State-Trait Anxiety Inventory with older adult psychiatric outpatients.* Journal of anxiety disorders, 1997. **11**(1): p. 33-47.
42. Steer, R.A., et al., *Structure of the computer-assisted Beck Anxiety Inventory with psychiatric inpatients.* Journal of personality assessment, 1993. **60**(3): p. 532-542.
43. Osman, A., et al., *Factor structure, reliability, and validity of the Beck Anxiety Inventory in adolescent psychiatric inpatients.* Journal of clinical psychology, 2002. **58**(4): p. 443-456.
44. Gillis, M.M., D.A. Haaga, and G.T. Ford, *Normative values for the Beck Anxiety Inventory, Fear Questionnaire, Penn State Worry Questionnaire, and Social Phobia and Anxiety Inventory.* Psychological Assessment, 1995. **7**(4): p. 450.
45. Oei, T.P., A. Moylan, and L. Evans, *Validity and clinical utility of the Fear Questionnaire for anxiety-disorder patients.* Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1991. **3**(3): p. 391.
46. Mavissakalian, M., *The Fear Questionnaire: a validity study.* Behaviour Research and Therapy, 1986. **24**(1): p. 83-85.
47. van Rijsoort, S., P. Emmelkamp, and G. Vervaeke, *The Penn state worry questionnaire and the worry domains questionnaire: Structure, reliability and validity.* Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 1999. **6**(4): p. 297-307.
48. Brown, T.A., M.M. Antony, and D.H. Barlow, *Psychometric properties of the Penn State Worry Questionnaire in a clinical anxiety disorders sample.* Behaviour research and therapy, 1992. **30**(1): p. 33-37.

49. Bland, J.M. and D.G. Altman, *Multiple significance tests: the Bonferroni method.* Bmj, 1995. **310**(6973): p. 170.
50. Copeland, W.E., et al., *Generalized anxiety and C-reactive protein levels: a prospective, longitudinal analysis.* Psychol Med, 2012. **42**(12): p. 2641-50.
51. Raison, C.L. and A.H. Miller, *Pathogen-Host Defense in the Evolution of Depression: Insights into Epidemiology, Genetics, Bioregional Differences and Female Preponderance.* Neuropsychopharmacology, 2017. **42**(1): p. 5-27.
52. Miller, A.H. and C.L. Raison, *The role of inflammation in depression: from evolutionary imperative to modern treatment target.* Nat Rev Immunol, 2016. **16**(1): p. 22-34.
53. Dhabhar, F.S., *Enhancing versus suppressive effects of stress on immune function: implications for immunoprotection and immunopathology.* Neuroimmunomodulation, 2009. **16**(5): p. 300-17.
54. Dantzer, R., et al., *From inflammation to sickness and depression: when the immune system subjugates the brain.* Nat Rev Neurosci, 2008. **9**(1): p. 46-56.
55. Rosenblat, J.D., et al., *Inflamed moods: a review of the interactions between inflammation and mood disorders.* Prog Neuropsychopharmacol Biol Psychiatry, 2014. **53**: p. 23-34.
56. Shattuck, E.C. and M.P. Muehlenbein, *Human sickness behavior: Ultimate and proximate explanations.* American Journal of Physical Anthropology, 2015. **157**(1): p. 1-18.
57. Maes, M., et al., *Depression and sickness behavior are Janus-faced responses to shared inflammatory pathways.* BMC Med, 2012. **10**(1): p. 66.
58. 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, 2014. **71**(12): p. 1381-1391.
59. 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.* Journal of Neurology, Neurosurgery & Psychiatry, 2020. **91**(1): p. 21-32.
60. Köhler-Forsberg, O., et al., *Efficacy of anti-inflammatory treatment on major depressive disorder or depressive symptoms: meta-analysis of clinical trials.* Acta Psychiatrica Scandinavica, 2019. **139**(5): p. 404-419.
61. Husain, M.I., et al., *Adjunctive simvastatin for treatment-resistant depression: study protocol of a 12-week randomised controlled trial.* BJPsych open, 2019. **5**(1).
62. Yatham, M.S., et al., *Do statins have an effect on depressive symptoms? A systematic review and meta-analysis.* Journal of affective disorders, 2019. **257**: p. 55-63.
63. Nettis, M.A., et al., *Augmentation therapy with minocycline in treatment-resistant depression patients with low-grade peripheral inflammation: results from a double-blind randomised clinical trial.* Neuropsychopharmacology, 2021: p. 1-10.

Supplementary material table 1. Interaction terms between a continuous modelled time variable and basal and LPS-induced inflammation indexes

		Basal inflammation index				LPS-induced inflammation index factor 1				LPS-induced inflammation index factor 2			
		Interaction term	95% confidence interval			Interaction term	95% confidence interval			Interaction term	95% confidence interval		
	Instrument	**Beta**	**Lower**	**Upper**	**P-value**	**Beta**	**Lower**	**Upper**	**P-value**	**Beta**	**Lower**	**Upper**	**P-value**
Beck's Anxiety Inventory	Total score	-0.005	-0.009	-0.001	0.012	-0.005	-0.012	0.001	0.098	-0.013	-0.018	-0.007	< 0.001
	Subjective subscale	-0.006	-0.010	-0.001	0.009	-0.004	-0.010	0.003	0.237	-0.009	-0.015	-0.003	0.005
	Somatic subscale	-0.004	-0.008	0.000	0.043	-0.005	-0.011	0.001	0.086	-0.014	-0.019	-0.008	< 0.001
Fear Questionnaire	Total score	-0.003	-0.007	0.000	0.063	-0.001	-0.006	0.005	0.821	-0.002	-0.008	0.003	0.458
	Social phobia subscale	-0.002	-0.005	0.002	0.291	-0.001	-0.007	0.004	0.603	-0.003	-0.008	0.002	0.280
	Blood phobia subscale	-0.003	-0.007	0.001	0.092	0.001	-0.005	0.007	0.792	-0.001	-0.007	0.004	0.631
	Agoraphobia subscale	-0.004	-0.008	0.000	0.060	-0.001	-0.007	0.005	0.757	-0.001	-0.007	0.005	0.769
Worry (PSWQ)	Total score	-0.003	-0.007	0.001	0.134	-0.004	-0.010	0.002	0.227	-0.006	-0.012	-0.000	0.033

Supplementary material table 1. Standardized beta coefficients of the interaction term between a continuous modelled time variable and basal and LPS-induced inflammation indexes and anxiety. Linear mixed models fitted with repeated measures, using standardized variables, which were assessed up to six times over 9 years of follow-up. P-values remain statistically significant after adjusting for multiple testing at p = 0.001. Models were adjusted for baseline variables of gender, age, sickness prior to interview, the use of anti-inflammatory medication, and BMI.

Chapter 7

Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression

van Eeden, W.A.,, Chuan, L., van Hemert, A.M., Carlier, I.V.E. , Penninx, B.W.c, Wardenaar, K.J., Hoos, H., Giltay, E.J.

(2021). Psychiatry Research, 299, 113823.

Abstract

Background: Predicting the onset and course of mood and anxiety disorders is of clinical importance but remains difficult. We compared the predictive performances of traditional logistic regression, basic probabilistic machine learning (ML) methods, and automated ML (Auto-sklearn).

Methods: Data were derived from the Netherlands Study of Depression and Anxiety. We compared how well multinomial logistic regression, a naïve Bayes classifier, and Auto-sklearn predicted depression and anxiety diagnoses at a 2-, 4-, 6-, and 9-year follow up, operationalized as binary or categorical variables. Predictor sets included demographic and self-report data, which can be easily collected in clinical practice at two initial time points (baseline and 1-year follow up).

Results: At baseline, participants were 42.2 years old, 66.5% were women, and 53.6% had a current mood or anxiety disorder. The three methods were similarly successful in predicting (mental) health status, with correct predictions for up to 79% (95% CI 75–81%). However, Auto-sklearn was superior when assessing a more complex dataset with individual item scores.

Conclusions: Automated ML methods added only limited value, compared to traditional data modelling when predicting the onset and course of depression and anxiety. However, they hold potential for automatization and may be better suited for complex datasets.

Keywords: Psychiatry, Depression, Anxiety disorder, Machine Learning, Logistic Models, Epidemiologic Methods, Regression Analysis

Highlights

- The predictive performances were compared between a automated machine learning algorithm, a basic probabilistic ML algorithm and more traditional multinominal logistic regression when predicting depression and anxiety at 2-, 4-, 6-, 9-year follow-up.
- In 96 models, we used multiple sets of demographic and self-report questionnaire data as predictor variables, which can be easily collected in clinical practice at two initial time points (baseline and 1-year follow up).
- Depression and anxiety could be predicted with correct predictions of up to 79%.
- None of the methods seemed to consistently outperform one another. Although, Auto-sklearn was superior when using a more complex data-set with individual item- scores.
- Clinical practice as may in time benefit from integrating next generation ML methods into clinical discussion making due to its potential for automatization and its adaptability for more complex datasets, rather than its increased predictive accuracy compared to more traditional data modelling methods.

7.1 Introduction

Despite a large body of epidemiological research, the course and onset of mood and anxiety disorders remain difficult to predict. Improving the ability to predict the onset and course of mood and anxiety disorders can be clinically relevant for prevention, early detection, staging, and personalized treatments [1]. In clinical settings, most decision making is based on clinical-care guidelines and experience [2]. However, even experienced clinicians may ignore relevant information or may put too much emphasis on clinically salient cues [3]. Information on demographic characteristics and clinician-rated and self-reported measures are increasingly collected as part of routine outcome monitoring [ROM; 4], but this information is underused in clinical decision making. Literature suggests that automated statistical prediction of current diagnoses and course may improve clinical decision making [2, 5], particularly through modern machine learning (ML) approaches [6].

ML may be more time efficient, better suited for large and complex datasets, and better able to detect complex patterns in the data than current data-modelling approaches that rely heavily on human decision making [7, 8]. Most clinical data thus far have been analyzed by selecting only specific putative predictors. It is possible that more complex (including nonlinear and higher dimensional) patterns exist in the data, which can efficiently be detected when analyzing all available data simultaneously using ML [9, 10]. These approaches are able to examine huge numbers of potential predictors in an unbiased manner while preventing overfitting [11].

Thus far, ML studies in the field of psychiatry have been promising. A recent meta-analysis, which included 20 studies that predicted the therapeutic outcome of depression using ML algorithms, found an overall accuracy of .82 [95% CI 0.77–0.87; 12]. Another ML study used an extensive set of baseline variables in a subset of 805 depressed patients from the Netherlands Study of Depression and Anxiety (NESDA) cohort, including biological and psychological variables [e.g., personality traits; 13]. The study achieved an accuracy significantly greater than chance of 66% for predicting persistent depression over the course of 2 years. A similar study, performed in a subset of the NESDA cohort of 887 anxiety patients, found an accuracy of predicting anxiety recovery of 62% ($p < .05$) and an accuracy of predicting recovery of all common mental disorders of 63% [$p < 0.05$; 14]. Clinical severity measures were the most important predictor variables, which is in line with previous reports [12-14]. Although these studies seem promising, recently published papers have demonstrated only limited added value of ML over traditional regression analyzes [15, 16].

Additionally, other studies found that when predicting suicide, ML did not outperform regression analysis and resulted in positive predictive values below 0.01, thus limiting the practical utility of these predictions [17, 18]. Despite the increasing number of publications in this field, ML has yet to move towards clinical application [19].

Although ML incorporates less human decision making than traditional methods, most ML methods are still not fully automated. Feature selection has been standardized as much as possible, but cut-off values that determine which features to include or exclude are somewhat arbitrarily selected. One solution would be to fully automate the selection of features, as is done in the Auto-sklearn system [20]. Auto-sklearn is a next generation ML system that automatically selects the learning algorithm that best suits the data and automatically optimizes the hyperparameter settings of this algorithm. It has proved effective when analyzing a diverse range of datasets and is considered to be an efficient and robust system for use by both ML novices and experts [21, 22].

We aimed to study and to compare the performance of traditional multinominal logistic regression, a basic probabilistic ML algorithm [naïve Bayesian classifier; 23] and a more advanced automated ML method (Auto-sklearn) to predict DSM-IV-TR psychiatric diagnoses at a 2-, 4-, 6-, and 9-year follow up with different sets of predictors. We incorporated predictor variables that can be easily and inexpensively collected in clinical practice, such as demographic variables, clinician-rated psychiatric diagnoses, and self-reported depression and anxiety. Our hypothesis was that Auto-sklearn would be better at detecting complex patterns in the data and therefore would outdo a naïve Bayesian classifier, which in turn would outdo traditional regression analysis techniques in achieved level of accuracy. Moreover, we hypothesized that Auto-sklearn would be particularly efficient when single items and follow-up measures were included.

7.2 Methods

7.2.1 Study sample and procedures

For the current study, we included participants from the NESDA cohort, which investigated the course and consequences of depressive and anxiety disorders. A detailed description of the NESDA design and sampling procedures are published elsewhere [24]. The first wave (baseline) lasted from 2004 to September 2007, and the sixth wave of measurement at the 9-year follow up finished in October 2016. NESDA is a cohort study that recruited from the community (*n* = 564; 18.9%), general practice (*n* = 1,610; 54.0%), and secondary mental healthcare [n=807; 27.1%; 24] and included patients with a current or lifetime depressive or anxiety disorder as well as healthy controls (see supplementary Table 1). A limited number of exclusion criteria were applied, namely not being fluent in Dutch and the presence of other clinically overt psychiatric disorders (e.g., addiction, psychotic, bipolar). With this method, NESDA aimed for a cohort that is representative for diverse populations of healthy controls and patients with depression and anxiety [24]. Due to missing outcome data (mainly due to attrition), we included 2,596 (87.1%) participants to predict 2-year outcomes, 2,402 (80.6%) to predict 4-year outcomes, 2,256 (75.7%) to predict 6-year outcomes, and 2,068 (69.4%) to predict 9-year outcomes.

7.2.2 Measures

7.2.2.1 Independent variables

An overview of the independent variables within each predictor set can be found in Table 1 in the supplementary material. Independent variables comprised baseline demographics, lifetime and baseline DSM-IV-TR diagnoses, self-reported depression, and anxiety symptomatology. Demographic variables included gender, age, ethnicity (North European heritage: yes/no), level of education (1 = elementary or less; 2 = general intermediate/secondary education; 3 = college/university), partner status (no partner, with partner [not married], married, living apart/no partner, divorced/no partner, widowed/no partner), and working status (employed/unemployed). The Composite International Diagnostic Interview (CIDI WHO, version 2.1) was used to assess the presence of mood and anxiety disorders according to the DSM-IV-TR. This included current dysthymia, major depressive disorder (MDD), lifetime depressive disorder, social phobia, panic with agoraphobia, panic without agoraphobia, agoraphobia without panic, generalized anxiety disorder, and lifetime anxiety disorder. Future CIDI-based diagnoses were used as outcome variables at 2-, 4-, 6-, and 9-year follow up, and past and current CIDI-based diagnoses were used as independent

variables. Thus, diagnoses at baseline and at Years 2, 4, and 6 were used to predict the diagnosis at the 9-year follow up (see Section 2.2.2).

Anxiety and depressive severity as well as symptoms at baseline and 1-year follow up were assessed using the Fear Questionnaire [FQ; 25], the Beck's Anxiety Inventory [BAI; 26], and the Inventory of Depressive Symptomatology [IDS-SR; 27]. These measures were entered into the models as either sum scores only or as a combination of sum scores and individual items. Detailed (psychometric) information about the measures can be found in the supplementary material.

7.2.2.2 Outcome variable: Clinical diagnoses

The CIDI WHO, version 2.1 was used to assess clinical diagnoses according to the DSM-IV-TR. The CIDI is a fully standardized diagnostic interview with extensively validated psychometric characteristics [24, 28] and may be considered a gold standard for psychiatric diagnostic classification [29, 30].

At the 2-, 4-, 6-, and 9-year follow up, CIDI-based outcomes were coded both as a binary variable (psychiatric disorder absent vs. present) and as a categorical variable with four categories: healthy, mood disorder (i.e., major depression and/or dysthymia), anxiety disorder (i.e., general anxiety, social phobia, panic with agoraphobia, panic without agoraphobia, and/or agoraphobia without a panic disorder), and comorbid mood and anxiety disorders.

7.2.3 Statistical analysis

A total of 96 models were tested. We compared three methods, over four sets of predictor variables, over two outcome sets, and over four follow-up waves. The three methods were multinomial logistic regression [31], naïve Bayes classifier [23], and Auto-sklearn [21]. The four sets of predictor variables (all including sociodemographic variables and baseline diagnoses) were (a) baseline sum scores only; (b) baseline sum scores and 1-year follow up sum scores; (c) baseline sum scores, 1-year follow up sum scores, and individual items at baseline; and (d) sum scores and individual items at baseline and 1-year follow up. For an overview of the predictor Sets A–D, see Table 1 in the supplementary material. Missing item values (0.54% –13.1%) were replaced by the mean of the available cases. The two outcomes were binary (healthy/mood or anxiety disorder) and multinomial (healthy [A], mood disorder [B], anxiety [C], or comorbid mood- and anxiety disorder [D]). The follow-up waves occurred at 2, 4, 6, and 9 years.

Auto-sklearn is an automated ML system that addresses both the problem of choosing which ML algorithm is best suited to analyze a specific application scenario (i.e., the model/algorithm selection problem) and the problem of determining which parameter setting leads to high performance (i.e., the hyperparameter optimization problem). Auto-sklearn considers a wide range of feature selection methods including all classification approaches implemented within the Python `scikit-learn` package, spanning 15 classifiers (e.g., random forests, decision tree, gradient boosting, etc.), 14 feature preprocessing methods (e.g., feature agglomeration, polynomial, nystroem sampler, etc.), and four data preprocessing methods (i.e., one-hot encoding, imputation, balancing, and rescaling), giving rise to a structured hypothesis space with 110 hyperparameters. Auto-sklearn features preprocessing methods that can be mainly categorized into feature selection, kernel approximation, matrix decomposition, embeddings, feature clustering, polynomial feature expansion, and methods that use a classifier for feature selection [for more details see; 22]. Previous research shows that the classification performance is often much better than using standard selection/hyperparameter optimization methods [21], and researchers believe Auto-sklearn to be a promising system for use by both ML novices and experts [22]. Auto-sklearn won six out of 10 phases of the first ChaLearn AutoML challenge. Furthermore, a comprehensive analysis of over 100 diverse datasets, while taking into account time and computational resource constraints, demonstrated that Auto-sklearn outperformed the previous state of the art in AutoML [22]. More details about Auto-sklearn can be found elsewhere [21, 22; https://automl.github.io/auto-sklearn/master/api.html, accessed at 2019-12-10].

Naïve Bayes classifier is a basic ML method that can predict class membership probabilities, such as the probability that a given MDD patient is still depressed after 2 years, with the underlying assumption that the effect of an attribute value on a given class is independent of the values of the other attributes. It aims to simplify the computation involved and, in this sense, is considered naïve [23]. For the present study, we used the Gaussian Naïve Bayes Classifier provided in the `scikit-learn` package with the var_smoothing hyper-parameter. According to the `scikit-learn` manual, by using this implementation a researcher need not choose the probability cut off. Several hyper-parameter settings were tried in the preliminary analysis, resulting in no significant differences. Therefore, the default hyper-parameter setting was used (i.e., setting the value of var_smoothing to `1e-9`). More details about the `scikit-learn` can be found elsewhere (https://scikit-learn.org/stable/modules/generated/sklearn.naive_Bayes.GaussianNB.html #sklearn.naive_Bayes.GaussianNB, accessed at 2019-12-10).

Logistic regression is a classification method used for binary or multinomial outcome variables. Multinomial logistic regression is a classification method that generalizes logistic regression to multiclass problems [31]. We used the `R` package `nnet` [R Foundation for Statistical Computing, Vienna, Austria, 2016. https://www.R-project.org/; 32].

We computed all models by randomly splitting (50:50) the dataset into a training and a test dataset using `Scikit-learn` data split [33]. The training dataset was used to select the best fitting regression model or ML algorithm. For the present study, models were optimized for overall accuracy. Auto-sklearn feature selection and preprocessing were based on the training data. Auto-sklearn selected "multinomial_nb" as its classifier for the binary outcome analysis and "random forest" for the multinomial outcome analyzes. Subsequently, we tested and compared the accuracy of how well these models/algorithms predicted outcomes in the test data with a 95% CI (i.e., percentage of correctly predicted individuals). We also tested and compared their balanced accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. For the multinominal outcomes, this was computed using a one-versus-all approach. For each model, we tested the significance of accuracy related to the no-information rate. The no-information rate contains the accuracy if the model were to choose the most frequent outcome group: healthy, that is, the proportion of correct predictions when all patients are predicted to be healthy. Auto-sklearn and naïve Bayes classifier were implemented using the Python programming language [34]. For logistic regression, `R` was used [R Foundation for Statistical Computing, Vienna, Austria, 2016. https://www.R-project.org/; 32].

7.3 Results

7.3.1 Sociodemographic and clinical characteristics at baseline

Characteristics of the study population are presented in supplementary Table 2. Age at baseline ranged from 18 to 64 years (*M* = 42.2, *SD* = 13.1), and 1,975 (66.5%) participants were women. At baseline, 26.8% of the sample suffered from MDD (*n* = 796), 9.3% of the sample from dysthymia (*n* = 241), and 43.7% from a (comorbid) anxiety disorder (*n* = 1,299), of which social anxiety disorder was the most common (18.6%; *n* = 483). Of the participants in our sample, 46.1% did not meet DSM-IV-TR criteria for a mood or anxiety diagnosis within the preceding 6 months (*n* = 1,368), of whom 54.2% had never been diagnosed with a psychiatric disorder (*n* = 742).

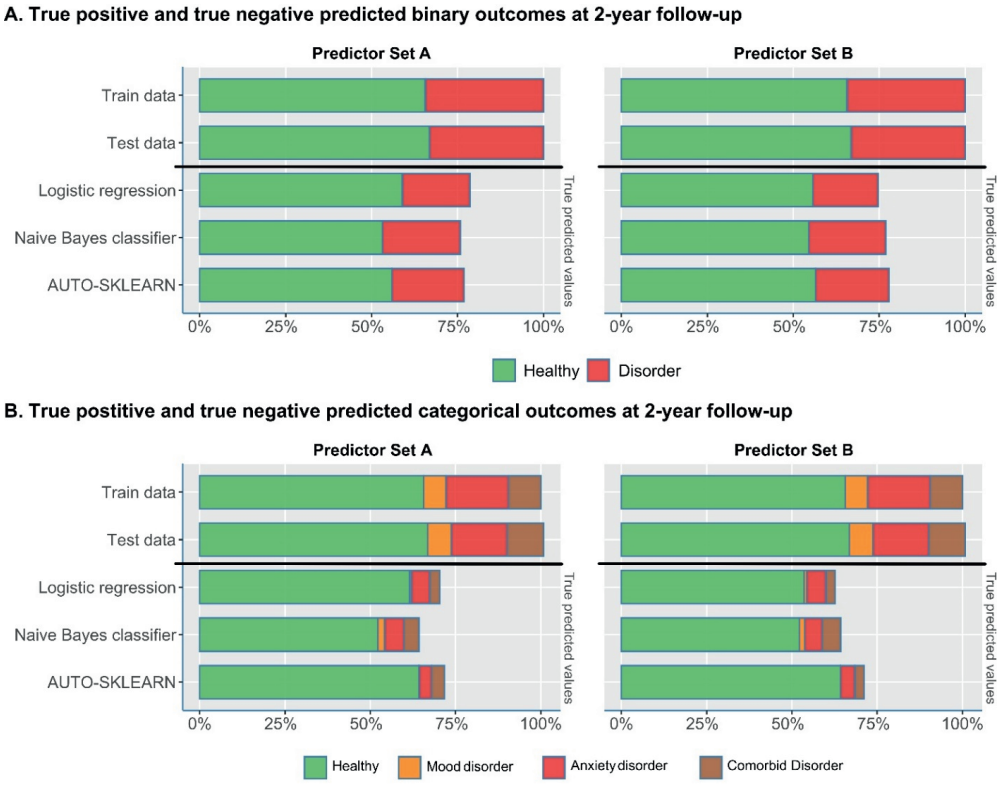


Figure 1. Percentages of train and test dataset values, as well as those correctly predicted at 2-year follow up, using the three data models. All predictor sets included baseline psychiatric diagnoses and demographic variables. Predictor Set A further includes baseline and 1-year follow-up sum scores. Predictor Set B additionally includes baseline and 1-year follow-up individual items.

7.3.2 Prediction of health status as binary outcome

Figures 1 and 2 and supplementary material Figure 1 and Table 3 contain the prediction of health status as a binary outcome (i.e., mentally healthy vs. any anxiety or mood disorder) at the 2-, 4-, 6-, and 9-year follow up using either logistic regression, naïve Bayes classifier, or Auto-sklearn. Figure 1 demonstrates the correctly predicted health status at the 2-year follow up (true negatives and true positives). With optimized overall accuracy, the three methods had different sensitivity and specificity levels. As demonstrated in Figure 2, Auto-sklearn had the highest specificity, with values between .84 and .90, but it had poor sensitivity values (.54–.75), predicting more disorders at the expense of correctly predicting a healthy health status (see also supplementary Table 1). The naïve Bayes classifier had specificity values between .76 and .88 and sensitivity values between .60 and .69. Logistic regression models had the lowest specificity values (.35–.59) but performed better regarding sensitivity values (.82–.93). Together this resulted in balanced accuracy levels ranging from .60–.75, .68–.75, and .63–.74 for Auto-sklearn, naïve Bayes classifier, and logistic regression, respectively.

As further demonstrated in Figure 2, the accuracy values ranged from .75 through .79. Logistic regression, naïve Bayes classifier, and Auto-sklearn were all significantly ($p < .001$) more accurate than the no-information rate (level of accuracy when only predicting a healthy status). Regarding logistic regression, the level of accuracy was significantly higher when only sum scores, and not individual item scores, were included as predictor variables (predictor Set A; acc .79 [95% CI .76–.81]), compared to logistic regression predictor Set B (acc .75 [95% CI .72–.77). The level of accuracy of naïve Bayes classifier and Auto-sklearn did not significantly decrease or improve when individual items were added as predictor variables. At 4-, 6-, and 9-year follow up, accuracy values ranged between .73–.78, .71–.77, and .76–.79 for logistic regression, naïve Bayes classifier, and Auto-sklearn, respectively. Of 16 tests per method (of which eight are presented in Figure 2 and eight in supplementary Table 3), Auto-sklearn had significantly higher accuracy levels than the no-information rate for all tests, compared to eight out of 16 for naïve Bayes classifier and eight out of 16 for logistic regression. Auto-sklearn thus performed adequately within each of the different datasets four different datasets.

	Predictor sets	Sensitivity	Specificity	PPV	NPV	Balanced accuracy	Accuracy
Outcome at year 2:							
- Logistic regression	A.	0.882	0.592	0.813	0.714	0.702	0.786***
	B.	0.834	0.571	0.796	0.631	0.737	0.747***
- Naive Bayes classifier	A.	0.684	0.795	0.624	0.835	0.740	0.758***
	B.	0.677	0.815	0.646	0.836	0.770	0.747***
- AUTO-SKLEARN	A.	0.633	0.836	0.658	0.821	0.735	0.769***
	B.	0.645	0.845	0.675	0.827	0.745	0.779***
Outcome at year 4:							
- Logistic regression	A.	0.911	0.446	0.805	0.668	0.679	0.779***
	B.	0.824	0.490	0.802	0.527	0.657	0.729
- Naive Bayes classifier	A.	0.647	0.808	0.574	0.851	0.728	0.762***
	B.	0.662	0.789	0.556	0.854	0.753	0.753**
- AUTO-SKLEARN	A.	0.414	0.921	0.676	0.797	0.667	0.776***
	B.	0.539	0.885	0.651	0.828	0.712	0.786***
Outcome at year 6:							
- Logistic regression	A.	0.919	0.361	0.808	0.605	0.640	0.777**
	B.	0.858	0.410	0.809	0.498	0.634	0.744
- Naive Bayes classifier	A.	0.639	0.808	0.533	0.867	0.724	0.765
	B.	0.604	0.799	0.507	0.855	0.702	0.749
- AUTO-SKLEARN	A.	0.375	0.930	0.647	0.813	0.652	0.788***
	B.	0.444	0.050	0.615	0.826	0.675	0.787***
Outcome at year 9:							
- Logistic regression	A.	0.928	0.345	0.803	0.626	0.637	0.778**
	B.	0.928	0.472	0.835	0.696	0.700	0.811***
- Naive Bayes classifier	A.	0.633	0.767	0.486	0.857	0.700	0.732
	B.	0.596	0.754	0.457	0.843	0.675	0.713
- AUTO-SKLEARN	A.	0.255	0.949	0.636	0.786	0.602	0.770*
	B.	0.626	0.943	0.614	0.786	0.602	0.767*

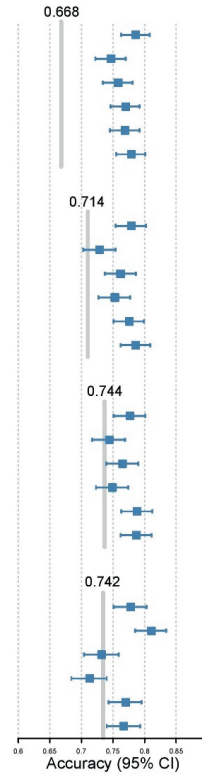


Figure 2. Predicting health status (binary outcome) at 2-, 4-, 6-, and 9-year follow up. All predictor sets included baseline psychiatric diagnoses and demographic variables. Predictor Set A further includes baseline and 1-year follow-up sum scores. Predictor Set B additionally includes baseline and 1-year follow-up individual items. The grey vertical line denotes as the no information rate for year 2-, 4-, 6-, and 9-year outcomes, respectively. Accuracy values were compared to the no-information rate by using a one way ANOVA test of which the *p* values are as follows:

* *p* value < .05

** *p* value < .01

*** *p* value < .001

7.3.3 Prediction of health status as categorical outcome

The results of predicting health status as a categorical outcome (i.e., healthy, mood disorder, anxiety disorder, or comorbid mood- and anxiety disorder) at the 2-, 4-, 6-, and 9-year follow up using either Auto-sklearn, naïve Bayes classifier, or logistic regression are shown in Figures 1, 3, and 4 and in the supplementary material Figure 1 and Tables 4 and 5. Figure 1 demonstrates the correctly predicted health status at 2-year follow up (true positives and true negatives). When the models were optimized for overall accuracy, their performance for predicting the disorder categories were low. When predicting with logistic regression, balanced accuracy values were .53 for mood disorders, .62 for anxiety disorders, and .61 for comorbidity. When predicting with Auto-sklearn, balanced accuracy values were .50 for mood disorders, .60 for anxiety disorders, and .61 for comorbidity. Comparatively, these figures were .70 and .66 when predicting a healthy health status with logistic regression and Auto-sklearn, respectively (see figure 3 outcome year 2). Mood disorder (n = 91 cases in the test data set) was predicted the least often, resulting in sensitivity values ranging from .00–.32 and specificity values ranging from .89–1.00. Further inspection of Figure 1 in the supplementary material demonstrates that both logistic regression and Auto-sklearn mostly predicted a healthy health status instead of mood disorders (n = 55 and n = 68, respectively).

As further demonstrated in Figures 3 and 4, the accuracy values when predicting health status at 2-year follow up ranged from .63 to .72. Both logistic regression (acc .70 [95% CI .68–.73]; p = .003) and Auto-sklearn (acc.72 [95% CI .69–.74]; p < .001) were significantly more accurate than the no-information rate, when predicting health status with sum scores at 2-year follow-up (see Figure 3), but only Auto-sklearn was significantly more accurate than the no-information rate when also individual item scores were included (acc .71 [95% CI .69–.74]; p < .001; see Figure 4). Again, the level of accuracy of logistic regression was significantly lower when individual item scores were included as predictor variables (predictor Set B; acc .63 [95% CI .60–.65]; p = >.99), compared to only sum scores (predictor Set A; acc .70 [95% CI .68–.73]; p = .003) when predicting health status at 2-year follow up. Auto-sklearn achieved demonstrated similar predictive performance when using sum scores as well as individual item scores (see Tables 4 and 5 in the supplementary material). Naïve Bayes classifier did not achieve levels of accuracy above the no-information rate. Achieving significantly accurate predictions became more difficult at later follow-ups. None of the models achieved accuracy levels that exceeded the no-information rate when predicting health status at 4-, 6-, and 9-years follow up.

	Health status	Sensitivity	Specificity	PPV	NPP	Balanced accuracy	Accuracy
Outcome at year 2:							
- Logistic regression	Healthy	0.922	0.480	0.781	0.753	0.701	0.703**
	Mood disorder	0.074	0.988	0.300	0.941	0.531	
	Anxiety disorder	0.333	0.909	0.418	0.874	0.621	
	Comorbidity	0.270	0.959	0.435	0.918	0.614	
- Naive Bayes classifier	Healthy	0.782	0.705	0.842	0.617	0.744	0.643
	Mood disorder	0.321	0.891	0.164	0.952	0.606	
	Anxiety disorder	0.343	0.881	0.361	0.872	0.612	
	Comorbidity	0.423	0.936	0.439	0.932	0.680	
- AUTO-SKLEARN	Healthy	0.963	0.364	0.753	0.831	0.656	0.718***
	Mood disorder	0.000	1.000	0.000	0.938	0.500	
	Anxiety disorder	0.225	0.959	0.516	0.863	0.603	
	Comorbidity	0.358	0.960	0.510	0.927	0.611	
Outcome at year 4:							
- Logistic regression	Healthy	0.944	0.309	0.774	0.688	0.627	0.718
	Mood disorder	0.035	0.986	0.111	0.954	0.511	
	Anxiety disorder	0.173	0.946	0.360	0.867	0.560	
	Comorbidity	0.178	0.972	0.380	0.924	0.575	
- Naive Bayes classifier	Healthy	0.787	0.671	0.857	0.557	0.729	0.668
	Mood disorder	0.193	0.936	0.131	0.959	0.565	
	Anxiety disorder	0.408	0.866	0.348	0.893	0.637	
	Comorbidity	0.402	0.931	0.361	0.941	0.666	
- AUTO-SKLEARN	Healthy	0.992	0.146	0.744	0.877	0.574	0.732
	Mood disorder	0.000	1.000	0.000	0.953	0.499	
	Anxiety disorder	0.089	0.983	0.485	0.860	0.531	
	Comorbidity	0.112	0.989	0.500	0.919	0.559	
Outcome at year 6:							
- Logistic regression	Healthy	0.957	0.240	0.786	0.657	0.957	0.745
	Mood disorder	0.053	0.991	0.231	0.952	0.053	
	Anxiety disorder	0.168	0.963	0.419	0.879	0.168	
	Comorbidity	0.092	0.978	0.233	0.937	0.092	
- Naive Bayes classifier	Healthy	0.795	0.701	0.886	0.540	0.748	0.685
	Mood disorder	0.228	0.950	0.194	0.959	0.589	
	Anxiety disorder	0.426	0.874	0.349	0.905	0.650	
	Comorbidity	0.342	0.913	0.220	0.950	0.627	
- AUTO-SKLEARN	Healthy	0.955	0.278	0.794	0.678	0.619	0.748
	Mood disorder	0.018	0.994	0.143	0.950	0.515	
	Anxiety disorder	0.187	0.950	0.372	0.880	0.567	
	Comorbidity	0.158	0.980	0.364	0.942	0.607	
Outcome at year 9:							
- Logistic regression	Healthy	0.954	0.217	0.778	0.624	0.586	0.737
	Mood disorder	0.019	0.989	0.083	0.950	0.504	
	Anxiety disorder	0.127	0.966	0.388	0.867	0.546	
	Comorbidity	0.154	0.977	0.313	0.945	0.566	
- Naive Bayes classifier	Healthy	0.753	0.648	0.860	0.477	0.700	0.633
	Mood disorder	0.192	0.942	0.149	0.957	0.567	
	Anxiety disorder	0.280	0.893	0.307	0.880	0.586	
	Comorbidity	0.385	0.862	0.157	0.954	0.623	
- AUTO-SKLEARN	Healthy	0.999	0.019	0.745	0.833	0.532	0.745
	Mood disorder	0.000	1.000	0.000	0.950	0.500	
	Anxiety disorder	0.020	0.999	0.750	0.857	0.521	
	Comorbidity	0.015	0.999	0.500	0.938	0.535	

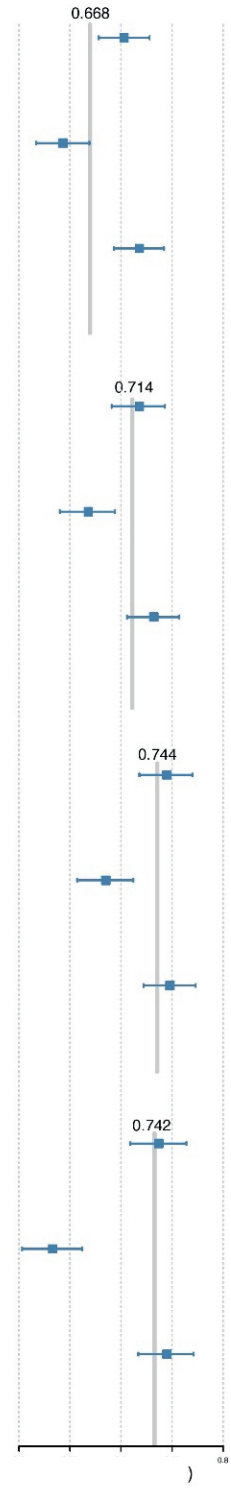


Figure 3. Predicting health status (multinominal outcome) at 2-, 4-, 6-, and 9-year follow up with baseline and 1-year sum scores (predictor Set A). All predictor sets included baseline psychiatric diagnoses and demographic variables. Predictor Set A further includes baseline and 1-year follow-up sum scores. Predictor Set B additionally includes baseline and 1-year follow-up individual items. PPV denotes as positive predictive value. NPV denotes as negative predictive value. The grey vertical line denotes as the no information for year 2-, 4-, 6-, and 9-year outcome, respectively. Accuracy values were compared to the no-information rate by using a one way ANOVA test of which the *p* values are as follows:

** *p* value < .01

*** *p* value < .001

	Health status	Sensitivity	Specificity	PPV	NPP	Balanced accuracy	Accuracy
Outcome at year 2:							
- Logistic regression	Healthy	0.802	0.596	0.800	0.599	0.699	0.627
	Mood disorder	0.136	0.926	0.109	0.942	0.531	
	Anxiety disorder	0.343	0.882	0.363	0.872	0.612	
	Comorbidity	0.255	0.921	0.276	0.913	0.588	
- Naive Bayes classifier	Healthy	0.781	0.717	0.847	0.619	0.749	0.643
	Mood disorder	0.259	0.915	0.168	0.949	0.587	
	Anxiety disorder	0.315	0.894	0.368	0.869	0.604	
	Comorbidity	0.511	0.895	0.365	0.939	0.703	
- AUTO-SKLEARN	Healthy	0.962	0.350	0.749	0.821	0.664	0.712***
	Mood disorder	0.000	1.000	0.000	0.938	0.500	
	Anxiety disorder	0.258	0.948	0.495	0.867	0.592	
	Comorbidity	0.255	0.967	0.479	0.917	0.659	
Outcome at year 4:							
- Logistic regression	Healthy	0.824	0.502	0.805	0.533	0.663	0.646
	Mood disorder	0.105	0.935	0.075	0.955	0.520	
	Anxiety disorder	0.207	0.903	0.272	0.867	0.555	
	Comorbidity	0.243	0.926	0.243	0.926	0.584	
- Naive Bayes classifier	Healthy	0.767	0.685	0.859	0.540	0.726	0.645
	Mood disorder	0.140	0.928	0.089	0.956	0.534	
	Anxiety disorder	0.313	0.880	0.313	0.880	0.596	
	Comorbidity	0.495	0.897	0.319	0.948	0.696	
- AUTO-SKLEARN	Healthy	0.987	0.160	0.746	0.833	0.569	0.729
	Mood disorder	0.000	0.998	0.000	0.952	0.500	
	Anxiety disorder	0.084	0.978	0.405	0.859	0.536	
	Comorbidity	0.131	0.988	0.519	0.921	0.551	
Outcome at year 6:							
- Logistic regression	Healthy	0.806	0.448	0.81	0.442	0.806	0.657
	Mood disorder	0.211	0.927	0.133	0.957	0.211	
	Anxiety disorder	0.226	0.921	0.313	0.882	0.226	
	Comorbidity	0.224	0.931	0.189	0.943	0.224	
- Naive Bayes classifier	Healthy	0.780	0.639	0.863	0.499	0.709	0.662
	Mood disorder	0.246	0.941	0.182	0.959	0.593	
	Anxiety disorder	0.271	0.897	0.296	0.885	0.584	
	Comorbidity	0.474	0.892	0.240	0.959	0.683	
- AUTO-SKLEARN	Healthy	0.957	0.281	0.795	0.692	0.616	0.754
	Mood disorder	0.035	0.994	0.250	0.951	0.506	
	Anxiety disorder	0.174	0.959	0.403	0.879	0.568	
	Comorbidity	0.237	0.977	0.429	0.947	0.569	
Outcome at year 9:							
- Logistic regression	Healthy	0.773	0.453	0.803	0.410	0.613	0.629
	Mood disorder	0.135	0.945	0.115	0.954	0.540	
	Anxiety disorder	0.287	0.882	0.293	0.880	0.585	
	Comorbidity	0.108	0.918	0.080	0.939	0.513	
- Naive Bayes classifier	Healthy	0.723	0.648	0.855	0.448	0.685	0.624
	Mood disorder	0.269	0.925	0.159	0.960	0.597	
	Anxiety disorder	0.313	0.882	0.311	0.883	0.598	
	Comorbidity	0.462	0.879	0.204	0.961	0.670	
- AUTO-SKLEARN	Healthy	0.993	0.071	0.755	0.792	0.509	0.749
	Mood disorder	0.000	1.000	0.000	0.950	0.500	
	Anxiety disorder	0.047	0.994	0.583	0.860	0.509	
	Comorbidity	0.077	0.993	0.417	0.941	0.507	

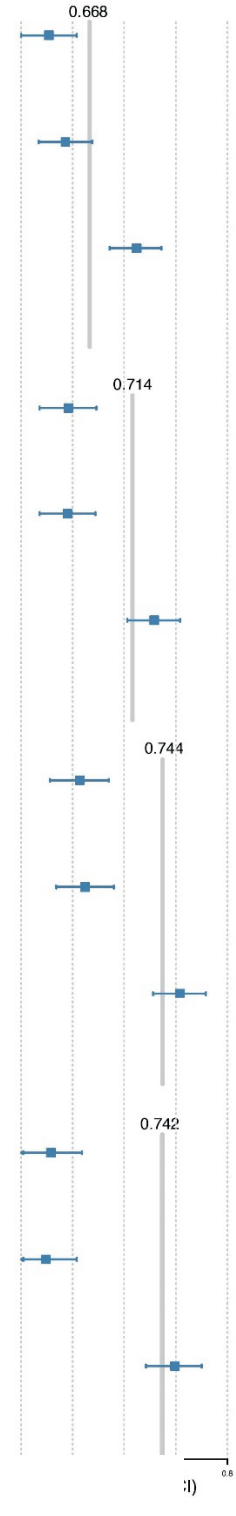


Figure 4. Predicting health status (multinominal outcome) at 2-, 4-, 6-, and 9-year follow up with baseline and 1-year sum scores and individual item-scores (predictor Set B). All predictor sets included baseline psychiatric diagnoses and demographic variables. Predictor Set B further includes baseline and 1-year follow-up sum scores and individual items. PPV denotes as positive predictive value. NPV denotes as negative predictive value. The grey vertical line denotes as the no information rate for year 2-, 4-, 6-, and 9-year outcome, respectively. Accuracy values were compared to the no-information rate by using a one way ANOVA test of which the p values are as follows: *** p value < .001

7.4 Discussion

Our aim was to assess and compare the predictive performances and clinical usefulness of Auto-sklearn, naïve Bayes classifier, and logistic regression to predict mood and anxiety disorders at follow up. Furthermore, we assessed the effects of different sets of predictors. Although we hypothesized that Auto-sklearn would outperform the two other data models, this could not be concluded unequivocally. In fact, only moderate levels of accuracy were found, with correct prediction percentages of up to 79% and 75% when using either binary or categorical outcomes, respectively. Yet, Auto-sklearn outperformed both logistic regression and naïve Bayes when predictor sets included individual item scores. Categorical outcomes were more difficult to predict than binary outcomes, compared to the no-information rate; in particular, mood disorders could not be distinguished well.

Our results support those of previous ML studies that reported 60% to 82% of correctly predicted mood and anxiety diagnoses when using a broad spectrum of predictor variables [10, 12-14, 35, 36]. One of these studies used a subset of the NESDA dataset that included patients with a depression at baseline and a more extensive set of clinical, behavioral, and biological baseline-only variables in order to predict the course of depression, resulting in accuracy levels of 62–66% [13]. A similar study, within a subset of anxiety patients in NESDA (again using an extensive set of predictors) found an accuracy for predicting anxiety recovery of 62% and a accuracy of predicting recovery of all common mental disorders of 63% [14]. In contrast to these prior studies, we only used data that could be easily collected in clinical practice, including 1-year follow-up data as predictor variables. Despite our dataset not being as rich and diverse, we achieved a higher overall accuracy which was significantly higher than the no-information rate [13, 14]. However, these results cannot be compared easily. Our often higher accuracy values were likely in part due to our inclusion of healthy participants. The predictive performance when predicting the disorder value were similar and the large proportion of the healthy health status outcomes resulted in unbalanced sensitivity and specificity values when models were optimized to maximum overall accuracy. Prior studies lacked thorough comparisons to (logistic) regression models, and thereby failed to address the additional value of ML methods over “traditional” data-modelling methods.

Previous ML studies in the field of psychiatry used a wide variety of ML methods, ranging from regression trees to gradient boosting machines—methods that were included in Auto-sklearn [10, 35]. In line with an earlier study, we found that depending on the predictor set, more complex ML

methods do not necessarily result in higher similar levels of accuracy when predicting future outcomes of mood disorders [36]. Two previous studies found that when optimized on overall level of accuracy, ML methods were about 1–6% more accurate compared to regression analysis and needed fewer predictor variables when predicting the persistence of mood disorders at a 12-week follow up [10, 35]. Although level of accuracy was higher for ML, this difference was not found to be significant in either study [10, 35]. Several studies found that ML was of only limited added value in research (Belsher et al., 2019; Christodoulou et al., 2019; van Mens et al., 2020) and clinical usefulness [19]. Although we did not find any published reviews within the field of psychiatry, within other fields the added value of ML has been notably criticized [e.g., 16, 37, 38]. However, it is possible that ML does outperform traditional methods when more complex (large) datasets are used [7, 8]. More advanced ML methods have the capability to distinguish which variables in large datasets are relevant or irrelevant for prediction, whereas traditional (regression) models rely on the researcher or clinician to select variables of interest to a particular analysis. ML therefore requires less human input. Although regression models sequentially analyze the relationship between variables, ML approaches can iteratively and contemporaneously analyze multiple interacting associations between variables or variable sets. Indeed, ML approaches may potentially be better suited to complex datasets with a large amount of predictors, while limiting the risk of overfitting [12]. These advantages were confirmed by our findings. Auto-sklearn outperformed the other two models when our predictor sets included more variables, that is, they were more complex.

ML, especially when automated, has the potential for use in mental healthcare. Deciding what information to collect from patients and making predictions on the micro and macro level based on that information are important aspects of a clinician's skill set. This includes predictions regarding suicide risk, violence, the efficacy of treatment options, and the prognoses on the course of disorders [2]. The accuracy of these predictions is of vital importance for individual patients. Two major approaches to predict clinical outcomes can be identified: the clinical and the statistical method. The clinical approach refers to an informal and intuitive process in which the clinician combines and integrates patient data. A clinician's experience, interpersonal sensitivity, and theoretical perspective combined with a patient's characteristics and circumstances determine how that clinician recalls, synthesizes, and interprets all these bits of information [2]. With a statistical approach, statistical methods are applied on objectively measured variables in order to make predictions and prognoses based on probabilities [2]. Two meta-analyzes demonstrated that

statistical approaches were more accurate than clinical methods [2, 5]. Our study found that moderate levels of accuracy can be accomplished based on data that can be easily collected in clinical practice, confirming that integrating statistical methods into clinical decision making could provide added benefits. Current mental healthcare is already partly digitalized, and the development of automated digital tools to assist clinicians should be attainable, providing clinicians fast and cheap support in decision making. Automated ML can be developed into such a tool because its automated techniques can match or improve upon expert human performance in certain ML tasks—often in a shorter amount of time [20]. Moreover, Auto-sklearn demonstrated that it can perform even under rigid time and computational resource constraints [21]. Automated ML is already demonstrating its usefulness in healthcare practice [20].

There are several study limitations that need to be discussed. First, despite the marginal differences between DSM-IV-TR and DSM-5 criteria for mood and anxiety disorders, the diagnostic classifications used in this study were slightly outdated but were chosen to be kept constant during the follow-up waves [39]. Despite our relatively large sample size, our analyzes could not be carried out for each diagnosis separately (e.g., dysthymia, panic disorder, etc.) because the samples would have become too small. Second, in contrast with other studies, we did not replicate our findings with an independent dataset [10, 36]. Although we made use of a training and testing dataset, it is possible that the results from the ML methods and regression analyzes differed in generalizability to other datasets, which could not be assessed with our current study design. Third, NESDA is an observational cohort study, and different types of pharmacological and psychotherapeutic treatment were not taken into account as predictor variables. Fourth, we included both healthy participants and patients, testing concomitantly the prediction of the course and onset of depression and anxiety. The proportion of healthy controls may have influenced the predictive models because their homeostatic responses to internal or external stimuli do not represent that of psychopathologic disorders [40]. The large proportion of the healthy health status outcomes resulted in unbalanced sensitivity and specificity values when models were optimized to maximum overall accuracy. Fifth, differentiating depression, anxiety, and comorbid disorders as multinomial variables was especially poor and may have been unrealistic because anxiety disorders and depression have overlapping risk factors and high levels of (subclinical) comorbidity [41, 42]. Sixth, ML may have more added value when the dataset is more complex, such as imaging or genetic data [7, 8, 12]. Although our data was easy to collect in clinical practice, it may have lacked the complexity that is needed for ML methods to excel. Finally, because of its automated features, Auto-sklearn

acts like a black box, which made it difficult for us to examine which individual features were most predictive. Nevertheless, significant levels of accuracy were achieved when predictor sets included sociodemographic, baseline diagnoses, and self-reported sum scores, which did not significantly improve when variables were added, suggesting that these were the most important predictor variables.

In conclusion, we found that moderately high levels of accuracy could be achieved when predicting dichotomous outcomes with easy-to-collect data. Auto-sklearn did not achieve the highest level of accuracy in every set of predictors, compared to traditional logistic regression and a naïve Bayes classifier. However, it was most consistent regardless of the set of predictor variables, and it outperformed the other models when the predictor sets were more complex (i.e., individual item scores). In time, clinical practice may benefit from integrating next generation automated ML methods into clinical decision making.

Acknowledgement

The infrastructure for the NESDA study (www.nesda.nl) is funded through the Geestkracht program of the Netherlands Organization for Health Research and Development (ZonMw, grant number 10-000-1002) and financial contributions by participating universities and mental healthcare organizations (Amsterdam University Medical Centers (location VUmc), GGZ inGeest, Leiden University Medical Center, Leiden University, GGZ Rivierduinen, University Medical Center Groningen, University of Groningen, Lentis, GGZ Friesland, GGZ Drenthe, Rob Giel Onderzoekscentrum).

References

1. McGorry, P.D., *Risk syndromes, clinical staging and DSM V: new diagnostic infrastructure for early intervention in psychiatry.* Schizophrenia research, 2010. **120**(1): p. 49-53.
2. Ægisdóttir, S., et al., *The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction.* The Counseling Psychologist, 2006. **34**(3): p. 341-382.
3. Odeh, M.S., R.A. Zeiss, and M.T. Huss, *Cues they use: Clinicians' endorsement of risk cues in predictions of dangerousness.* Behavioral sciences & the law, 2006. **24**(2): p. 147-156.
4. Carlier, I.V., et al., *Routine outcome monitoring and feedback on physical or mental health status: evidence and theory.* Journal of Evaluation in Clinical Practice, 2012. **18**(1): p. 104-110.
5. Grove, W.M., et al., *Clinical versus mechanical prediction: a meta-analysis.* Psychological assessment, 2000. **12**(1): p. 19.
6. Johnson, A.E., et al., *Machine learning and decision support in critical care.* Proceedings of the IEEE. Institute of Electrical and Electronics Engineers, 2016. **104**(2): p. 444.
7. Iniesta, R., D. Stahl, and P. McGuffin, *Machine learning, statistical learning and the future of biological research in psychiatry.* Psychological medicine, 2016. **46**(12): p. 2455-2465.
8. Wang, Y., L. Kung, and T.A. Byrd, *Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations.* Technological Forecasting and Social Change, 2018. **126**: p. 3-13.
9. Hahn, T., A. Nierenberg, and S. Whitfield-Gabrieli, *Predictive analytics in mental health: applications, guidelines, challenges and perspectives.* Molecular psychiatry, 2016. **22**(1): p. 37.
10. Chekroud, A.M., et al., *Cross-trial prediction of treatment outcome in depression: a machine learning approach.* The Lancet Psychiatry, 2016. **3**(3): p. 243-250.
11. Hastie, T., R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction, Springer Series in Statistics*. 2009, Springer New York.
12. Lee, Y., et al., *Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review.* Journal of affective disorders, 2018.
13. Dinga, R., et al., *Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach.* Translational psychiatry, 2018. **8**(1): p. 241.
14. Bokma, W.A., et al., *Predicting the naturalistic course in anxiety disorders using clinical and biological markers: a machine learning approach.* Psychological Medicine, 2020: p. 1-11.
15. van Mens, K., et al., *Predicting future suicidal behaviour in young adults, with different machine learning techniques: a population-based longitudinal study.* Journal of affective disorders, 2020.
16. Christodoulou, E., et al., *A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models.* Journal of clinical epidemiology, 2019.
17. Kessler, R.C., et al., *Developing a practical suicide risk prediction model for targeting high-risk patients in the Veterans Health Administration.* International journal of methods in psychiatric research, 2017. **26**(3): p. e1575.
18. Belsher, B.E., et al., *Prediction models for suicide attempts and deaths: a systematic review and simulation.* JAMA psychiatry, 2019. **76**(6): p. 642-651.
19. Tran, B.X., et al., *The current research landscape on the artificial intelligence application in the management of depressive disorders: A bibliometric analysis.* International journal of environmental research and public health, 2019. **16**(12): p. 2150.
20. Waring, J., C. Lindvall, and R. Umeton, *Automated machine learning: Review of the state-of-the-art and opportunities for healthcare.* Artificial Intelligence in Medicine, 2020: p. 101822.
21. Feurer, M., et al. *Efficient and robust automated machine learning*. in *Advances in Neural Information Processing Systems*. 2015.
22. Feurer, M., et al., *Auto-sklearn: Efficient and Robust Automated Machine Learning*, in *Automated Machine Learning*. 2019, Springer. p. 113-134.
23. Jayant, A. and a.O.R.M.C. Safari, *Data Science and Machine Learning Series: Naive Bayes Classifier Advanced Concepts*. 2020: Technics Publications.
24. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* Int J Methods Psychiatr Res, 2008. **17**(3): p. 121-40.
25. Marks, I.M. and A.M. Mathews, *Brief standard self-rating for phobic patients.* Behaviour research and therapy, 1979. **17**(3): p. 263-267.
26. Beck, A.T., et al., *An inventory for measuring clinical anxiety: psychometric properties.* Journal of consulting and clinical psychology, 1988. **56**(6): p. 893.

27. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): psychometric properties.* Psychol Med, 1996. **26**(3): p. 477-86.
28. Wittchen, H.-U., *Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review.* Journal of Psychiatric Research, 1994. **28**(1): p. 57-84.
29. Haro, J.M., et al., *Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health surveys.* International journal of methods in psychiatric research, 2006. **15**(4): p. 167-180.
30. Kessler, R.C., et al., *National comorbidity survey replication adolescent supplement (NCS-A): III. Concordance of DSM-IV/CIDI diagnoses with clinical reassessments.* Journal of the American Academy of Child & Adolescent Psychiatry, 2009. **48**(4): p. 386-399.
31. Menard, S., *Applied logistic regression analysis*. Vol. 106. 2002: Sage.
32. Ripley, B., W. Venables, and M.B. Ripley, *Package 'nnet'.* R package version, 2016: p. 7-3.
33. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* the Journal of machine Learning research, 2011. **12**: p. 2825-2830.
34. Rossum, G.v., *Python tutorial, Technical Report CS-R9526*. 1995, The Netherlands, Amsterdam: Centrum voor Wiskunde en Informatica (CWI).
35. Kessler, R.C., et al., *Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports.* Molecular psychiatry, 2016. **21**(10): p. 1366.
36. Nie, Z., et al., *Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study.* PLOS ONE, 2018. **13**(6): p. e0197268.
37. Frizzell, J.D., et al., *Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches.* JAMA cardiology, 2017. **2**(2): p. 204-209.
38. Desai, R.J., et al., *Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes.* JAMA Network Open, 2020. **3**(1): p. e1918962-e1918962.
39. Regier, D.A., E.A. Kuhl, and D.J. Kupfer, *The DSM-5: Classification and criteria changes.* World psychiatry, 2013. **12**(2): p. 92-98.
40. Regier, D.A., et al., *Limitations of diagnostic criteria and assessment instruments for mental disorders: implications for research and policy.* Archives of general psychiatry, 1998. **55**(2): p. 109-115.
41. Jacobson, N.C. and M.G. Newman, *Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies.* Psychological bulletin, 2017. **143**(11): p. 1155.
42. Shorter, E. and P. Tyrer, *Separation of anxiety and depressive disorders: blind alley in psychopharmacology and classification of disease.* BMJ, 2003. **327**(7407): p. 158-160.

Supplementary Material

Measures

Composite International Diagnostic Interview

The Composite International Diagnostic Interview (CIDI WHO, version 2.1) was used to assess the presence of depressive and anxiety disorders according to the DSM-IV-TR. The CIDI is used worldwide and has been demonstrated to have high interrater reliability high test–retest reliability [1, 2] and high validity for depressive and anxiety disorders [3, 4]. Trained clinical research staff conducted the interviews [5].

Fear Questionnaire

The 15-item Fear Questionnaire (FQ) is a self-report instrument that assesses the level of avoidance in relation to common phobias, including social phobia (Items 2, 6, 8,10, 13), agoraphobia (Items 4, 5, 7, 11, 14), and hematophobia/traumatophobia [Items 1, 3, 9, 12, 15; 6]. It consists of 15 equally weighted items, rated on a 9-point scale, ranging from 0 (*Would not avoid it*) to 8 (*Always avoid it*). The sum score ranges from 0 to 120. The psychometric properties of the FQ have been researched in multiple studies among both nonclinical populations and patients with an anxiety disorder [7-9].

Beck's Anxiety Inventory

The 21-item Beck's Anxiety Inventory (BAI) is a self-report instrument that assesses the overall severity of anxiety.[10] The items consist of 21 anxiety symptoms, including physical symptoms (e.g., "Heart pounding/racing") and psychological symptoms (e.g., "Fear of the worst happening"). It consists of equally weighted items, rated on a 4-point scale, ranging from 0 (*not at all*) to 3 (*severely, I could barely stand it*). The BAI is scored by adding the ratings for all 21 symptoms to obtain a total score that can range from 0 to 63. The reliability and validity of the BAI are well established [10, 11].

Inventory of Depressive Symptomatology

The 30-item Inventory of Depressive Symptomatology (IDS-SR) was used to assess the severity of depression [12, 13]. The IDS-SR scale includes all symptoms of depression, including melancholic, atypical, and anxious symptoms. Moreover, several additional symptoms have been added, such as sympathetic arousal, pessimism, and interest in sex. It consists of 30 equally weighted items, rated on a 4-point scale (0–3). The IDS-SR is scored by adding the ratings of the 30 symptoms to obtain a total score that can range from 0 to 88. Items 11 and 12 ("increased/decreased appetite) and Items 13 and 14 (weight gain/weight loss) contain opposite features, so we combined each of them into

two ordinal items with both severe increase or decrease at Scale 3, yielding 28 items for the current analyzes [13].

References

1. Wittchen, H.-U., et al., *Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI).* The British Journal of Psychiatry, 1991. **159**(5): p. 645-653.
2. Wacker, H., et al., *Using the CIDI-C in the general population. Psychiatry: A world perspective. Edited by: Stefanis CN, Rabavilas AD, Soldatos CR. 2006*. Amsterdam: Elsevier Science Publishers.
3. Wittchen, H.-U., *Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review.* Journal of Psychiatric Research, 1994. **28**(1): p. 57-84.
4. Wittchen, H.-U., et al., *Recall and dating of psychiatric symptoms: test-retest reliability of time-related symptom questions in a standardized psychiatric interview.* Archives of General Psychiatry, 1989. **46**(5): p. 437-443.
5. Penninx, B.W., et al., *The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods.* Int J Methods Psychiatr Res, 2008. **17**(3): p. 121-40.
6. Marks, I.M. and A.M. Mathews, *Brief standard self-rating for phobic patients.* Behaviour research and therapy, 1979. **17**(3): p. 263-267.
7. Gillis, M.M., D.A. Haaga, and G.T. Ford, *Normative values for the Beck Anxiety Inventory, Fear Questionnaire, Penn State Worry Questionnaire, and Social Phobia and Anxiety Inventory.* Psychological Assessment, 1995. **7**(4): p. 450.
8. Oei, T.P., A. Moylan, and L. Evans, *Validity and clinical utility of the Fear Questionnaire for anxiety-disorder patients.* Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1991. **3**(3): p. 391.
9. Mavissakalian, M., *The Fear Questionnaire: a validity study.* Behaviour Research and Therapy, 1986. **24**(1): p. 83-85.
10. Beck, A.T., et al., *An inventory for measuring clinical anxiety: psychometric properties.* Journal of consulting and clinical psychology, 1988. **56**(6): p. 893.
11. Steer, R.A., et al., *Structure of the computer-assisted Beck Anxiety Inventory with psychiatric inpatients.* Journal of personality assessment, 1993. **60**(3): p. 532-542.
12. Trivedi, M.H., et al., *The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation.* Psychological Medicine, 2004. **34**(1): p. 73-82.
13. Rush, A.J., et al., *The Inventory of Depressive Symptomatology (IDS): Psychometric properties.* Psychological Medicine, 1996. **26**(3): p. 477-486.

Supplementary Table 1. The variables that were part of the four predictor sets (A through D) that were used in the analyzes with the 3 different data models

	Predictor set A	Predictor set B	Predictor set C	Predictor set D
1.	MDD (yes/no)	MDD (yes/no)	MDD (yes/no)	MDD (yes/no)
2.	Dysthymia (yes/no)	Dysthymia (yes/no)	Dysthymia (yes/no)	Dysthymia (yes/no)
3.	Minor depression (yes/no)	Minor depression (yes/no)	Minor depression (yes/no)	Minor depression (yes/no)
4.	History of MDD (yes/no)	History of MDD (yes/no)	History of MDD (yes/no)	History of MDD (yes/no)
5.	Social-phobia (yes/no)	Social-phobia (yes/no)	Social-phobia (yes/no)	Social-phobia (yes/no)
6.	Panic disorder with agoraphobia (yes/no)	Panic disorder with agoraphobia (yes/no)	Panic disorder with agoraphobia (yes/no)	Panic disorder with agoraphobia (yes/no)
7.	Panic disorder without agoraphobia (yes/no)	Panic disorder without agoraphobia (yes/no)	Panic disorder without agoraphobia (yes/no)	Panic disorder without agoraphobia (yes/no)
8.	Agoraphobia (yes/no)	Agoraphobia (yes/no)	Agoraphobia (yes/no)	Agoraphobia (yes/no)
9.	Generalized anxiety disorder (yes/no)	Generalized anxiety disorder (yes/no)	Generalized anxiety disorder (yes/no)	Generalized anxiety disorder (yes/no)
10.	History of anxiety disorder	History of anxiety disorder	History of anxiety disorder	History of anxiety disorder
11.	Gender	Gender	Gender	Gender
12.	Age	Age	Age	Age
13.	Level of education	Level of education	Level of education	Level of education
14.	North European ancestry (yes/no)	North European ancestry (yes/no)	North European ancestry (yes/no)	North European ancestry (yes/no)
15.	Partner status	Partner status	Partner status	Partner status
16.	Work status	Work status	Work status	Work status
17.	FQ sumscore baseline	FQ sumscore baseline	FQ sumscore baseline	FQ sumscore baseline
18.	BAI sumscore baseline	BAI sumscore baseline	BAI sumscore baseline	BAI sumscore baseline
19.	IDS-SR sumscore baseline	IDS-SR sumscore baseline	IDS-SR sumscore baseline	IDS-SR sumscore baseline
20.	FQ sumscore at 1-year follow-up	FQ item 1 through 15		FQ item 1 through 15
21.	BAI sumscore at 1-year follow-up	BAI item 1 through 21		BAI item 1 through 21
22.	IDS-SR sumscore at 1-year follow-up	IDS-SR item 1 through 28		IDS-SR item 1 through 28
23.	Delta FQ sumscore (1-year follow-up – baseline)	FQ sumscore at 1-year follow-up		
24.	Delta BAI sumscore (1-year follow-up – baseline)	BAI sumscore at 1-year follow-up		
25.	Delta IDS-SR sumscore (1-year follow-up – baseline)	IDS-SR sumscore at 1-year follow-up		
26.		Delta FQ sumscore (1-year follow-up – baseline)		
27.		Delta BAI sumscore (1-year follow-up – baseline)		
28.		Delta IDS-SR sumscore (1-year follow-up – baseline)		
29.		FQ item 1 through 15 at 1-year follow-up		
30.		BAI item 1 through 21 at 1-year 1 follow-up		
31.		IDS-SR item 1 through 28 at 1-year follow-up		

32.	Delta FQ item 1 through 15 (1-year follow up – baseline)
33.	Delta BAI item 1 through 21 (1-year follow up – baseline)
34.	Delta IDS-SR item 1 through 28 (1-year follow up – baseline)
35.	Delta IDS-SR item 28 - year 1 follow up[…]Delta BAI item 21 - year 1 follow up[…]

Note. MDD denotes Major depressive disorder. FQ denotes Fear Questionnaire. BAI denotes Beck's anxiety inventory. IDS-SR denotes as inventory of depressive symptomatology.

Supplementary Table 2. Baseline sociodemographic and clinical characteristics of the 2,596 NESDA participants.

	Cohort
Age in years (mean, *SD*)	42.2 (13.1)
Female (%)	65.5
North-European ethnicity (%)	94.8
Education level (%)	
Elementary or lower	38.1
Secondary education	58.0
College or university	3.9
Work status (%)	
Employed	53.4
Self-employed	6.3
Disability	9.1
Sick benefit	5.0
Early retirement	3.4
Unemployed	18.3
Partner status (%)	
Married	38.5
Partner but was not married	30.8
Divorced	7.3
Widowed	1.4
Mood disorder (%)	
Major depressive disorder	26.8
Minor depression	2.8
Dysthymia	9.3
Lifetime depression	66.2
Anxiety disorder (%)	
Panic disorder with agoraphobia	11.9
Panic disorder without agoraphobia	5.2
Agoraphobia without panic	5.1
Generalized anxiety disorder	13.3
Social anxiety disorder	18.6
Lifetime anxiety disorder	59.4
No Disorder (%)	46.1
No lifetime disorder	24.9
Self reports (mean, *SD*)	
Baseline totalscore IDS-SR	21.5 (14.1)
Baseline totalscore FQ	24.8 (19.9)
Baseline totalscore BAI	12.1 (10.7)
Year-1 totalscore IDS-SR	16.9 (12.4)
Year-1 totalscore FQ	20.8 (18.6)
Year-1 totalscore BAI	9.3 (9.2)

Note. *SD* denotes standard deviation. IDS-SR denotes Inventory of Depressive Symptomatology - Self Report. FQ denotes Fear Questionnaire. BAI denotes Beck Anxiety Inventory.

Supplementary Table 3 Predicting mental health status (binary outcome) at 2-, 4-, 6-, and 9-year follow up using baseline data as the independent variables (i.e., predictor set C, and D).

AUTO-SKLEARN	Outcome Year 2	Outcome Year 4	Outcome Year 6	Outcome Year 9
Baseline sum-scores				
accuracy	0.763	0.764	0.781	0.770
95% CI	0.739 - 0.786	0.738 - 0.787	0.756 - 0.805	0.743 - 0.795
p value [acc > NIR)	<0.001	<0.001	0.003	0.020
balanced accuracy	0.706	0.674	0.640	0.585
sensitivity	0.534	0.466	0.351	0.202
specificity	0.878	0.882	0.929	0.967
positive predictive value	0.685	0.613	0.627	0.684
negative predictive value	0.791	0.805	0.807	0.777
Baseline sum-scores and individual items				
accuracy	0.773	0.770	0.769	0.773
95% CI	0.749 - 0.795	0.745 - 0.794	0.743 - 0.793	0.746 - 0.798
p value [acc > NIR)	<0.001	<0.001	0.034	0.012
balanced accuracy	0.713	0.651	0.610	0.625
sensitivity	0.534	0.373	0.285	0.318
specificity	0.892	0.929	0.935	0.931
positive predictive value	0.710	0.677	0.599	0.616
negative predictive value	0.794	0.788	0.792	0.797
Naive Bayes classifier				
Baseline sum-scores				
accuracy	0.755	0.759	0.762	0.730
95% CI	0.732 - 0.778	0.733 - 0.782	0.736 - 0.786	0.702 - 0.757
p value [acc > NIR)	<0.001	<0.001	0.103	0.813
balanced accuracy	0.734	0.720	0.722	0.696
sensitivity	0.673	0.630	0.642	0.625
specificity	0.796	0.810	0.802	0.767
positive predictive value	0.621	0.570	0.527	0.483
negative predictive value	0.830	0.845	0.867	0.855
Baseline sum-scores and individual items				
accuracy	0.761	0.750	0.750	0.711
95% CI	0.737 - 0.784	0.725 - 0.774	0.724 - 0.775	0.682 - 0.739
p value [acc > NIR)	<0.001	0.003	0.355	0.989
balanced accuracy	0.746	0.725	0.714	0.683
sensitivity	0.701	0.668	0.639	0.625
specificity	0.791	0.783	0.788	0.741
positive predictive value	0.625	0.552	0.508	0.456
negative predictive value	0.842	0.855	0.864	0.851
Logistic regression				
Baseline sum-scores				
accuracy	0.766	0.769	0.762	0.769
95% CI	0.741 - 0.789	0.744 - 0.792	0.737 - 0.787	0.742 - 0.794
P value [acc > NIR)	0.000	0.000	0.091	0.625
balanced accuracy	0.713	0.661	0.619	0.625
sensitivity	0.870	0.911	0.912	0.923
specificity	0.557	0.411	0.326	0.326
positive predictive value	0.798	0.795	0.798	0.798
negative predictive value	0.680	0.650	0.560	0.596
Baseline sum-scores and individual items				
accuracy	0.743	0.748	0.748	0.782
95% CI	0.718 - 0.7663	0.722 - 0.772	0.722 - 0.773	0.755 - 0.807
p value [acc > NIR)	0.000	0.005	0.408	0.002
balanced accuracy	0.696	0.649	0.621	0.646
sensitivity	0.835	0.879	0.881	0.926
specificity	0.557	0.420	0.361	0.367
positive predictive value	0.791	0.791	0.801	0.808
negative predictive value	0.627	0.581	0.510	0.632

Note. The *p* value denotes the one-sided ANOVA statistic of accuracy (acc) compared with the No-Information Rate (NIR). NIR was 0.668, 0.714, 0.744, and 0.742 for year 2-, 4-, 6-, and 9-year outcome, respectively.

Supplementary Table 4 Predicting mental health status (categorical outcome) at 2-, 4-, 6-, and 9-year follow up using baseline sum scores as the independent variables (i.e., predictor Set C)

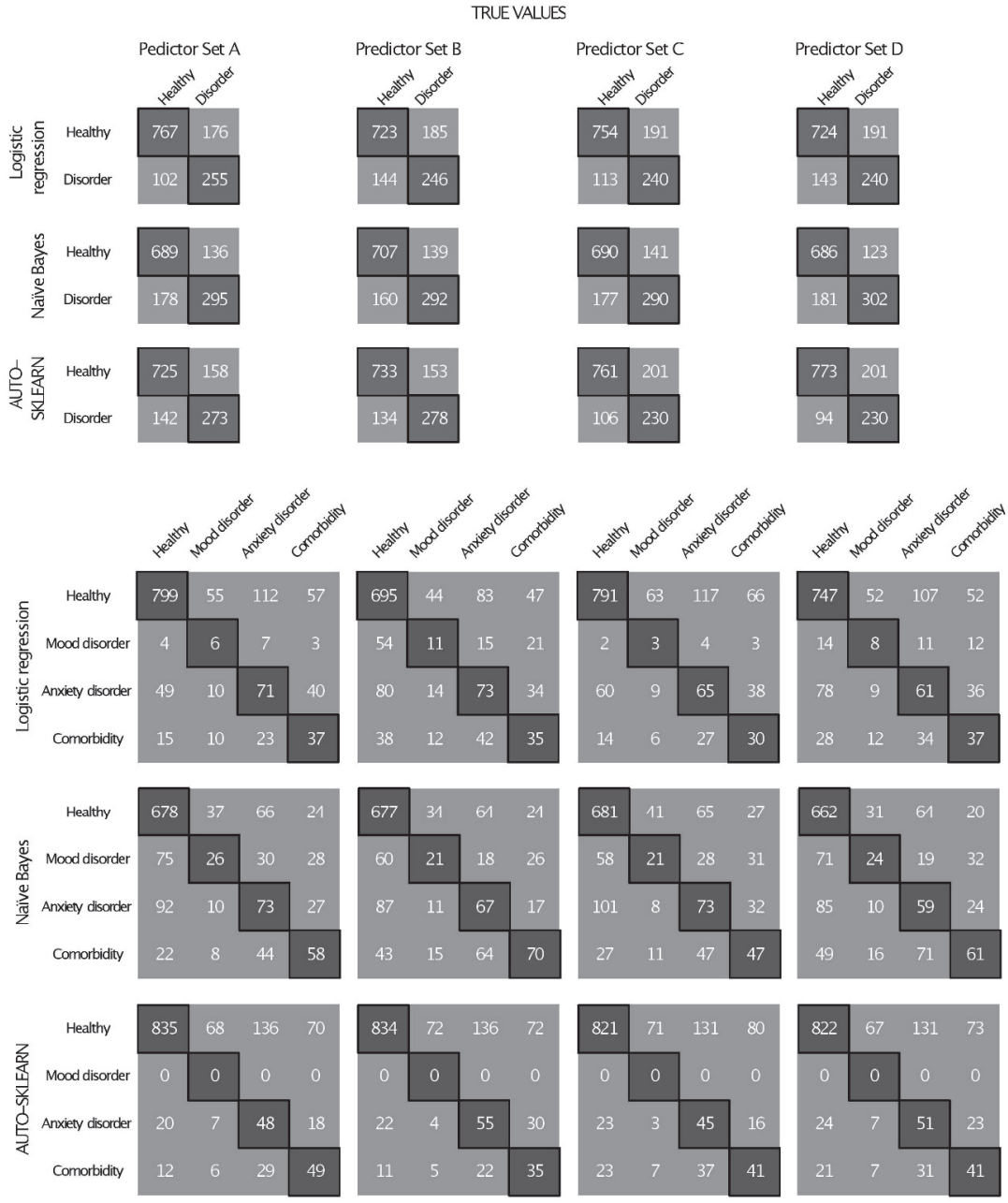
UTO-SKLEARN	Outcome at Year 2				Outcome at Year 4				Outcome at Year 6				Outcome at Year 9			
ealth status	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV
aseline sum scores																
accuracy	0.699				0.729				0.748				0.749			
95% CI	0.673 - 0.724				0.702 - 0.754				0.722 - 0.773				0.721 - 0.775			
p value [acc > NIR)	0.010				0.146				0.408				0.324			
balanced accuracy	0.646	0.500	0.586	0.621	0.569	0.500	0.535	0.570	0.537	0.500	0.527	0.517	0.527	0.500	0.521	0.528
sensitivity	0.947	0.000	0.211	0.299	0.983	0.000	0.084	0.159	0.988	0.000	0.071	0.039	0.995	0.000	0.047	0.062
specificity	0.346	1.000	0.961	0.942	0.155	0.999	0.985	0.982	0.087	1.000	0.984	0.995	0.060	1.000	0.995	0.995
positive predictive value	0.744	0.000	0.517	0.380	0.744	0.000	0.500	0.459	0.759	0.000	0.407	0.375	0.753	0.000	0.636	0.444
negative predictive value	0.764	0.938	0.861	0.919	0.779	0.953	0.860	0.923	0.714	0.949	0.869	0.935	0.800	0.950	0.860	0.941
aive Bayes classifier																
aseline sum scores																
accuracy	0.633				0.670				0.672				0.634			
95% CI	0.606 - 0.660				0.643 - 0.697				0.644 - 0.699				0.604 - 0.663			
p value [acc > NIR)	0.996				>0.999				1.000				1.000			
balanced accuracy	0.738	0.582	0.606	0.635	0.723	0.538	0.649	0.644	0.717	0.591	0.629	0.630	0.708	0.550	0.602	0.592
sensitivity	0.785	0.259	0.343	0.343	0.796	0.140	0.425	0.355	0.785	0.228	0.381	0.355	0.754	0.154	0.320	0.323
specificity	0.691	0.904	0.870	0.927	0.650	0.936	0.873	0.933	0.649	0.953	0.878	0.905	0.663	0.947	0.885	0.861
positive predictive value	0.837	0.152	0.341	0.356	0.851	0.099	0.369	0.342	0.867	0.206	0.331	0.213	0.865	0.133	0.320	0.135
negative predictive value	0.616	0.948	0.871	0.923	0.560	0.956	0.896	0.937	0.508	0.959	0.899	0.951	0.484	0.955	0.885	0.950
gistic regression																
aseline and sum scores																
accuracy	0.685				0.727				0.740				0.735			
95% CI	0.658 - 0.710				0.701 - 0.752				0.714 - 0.766				0.707 - 0.762			
P value [acc > NIR)	0.102				0.177				0.648				0.704			
balanced accuracy	0.671	0.515	0.603	0.589	0.613	0.501	0.575	0.567	0.575	0.516	0.549	0.549	0.584	0.504	0.539	0.533
sensitivity	0.912	0.037	0.305	0.219	0.958	0.018	0.190	0.150	0.956	0.035	0.135	0.118	0.961	0.019	0.107	0.092
specificity	0.429	0.993	0.901	0.960	0.268	0.984	0.960	0.984	0.194	0.997	0.962	0.980	0.206	0.989	0.972	0.973
positive predictive value	0.763	0.250	0.378	0.390	0.766	0.053	0.453	0.471	0.776	0.400	0.362	0.300	0.777	0.083	0.390	0.188
negative predictive value	0.709	0.939	0.869	0.912	0.719	0.953	0.871	0.922	0.602	0.951	0.875	0.939	0.647	0.950	0.865	0.941

ote. The *p* value denotes the one sided ANOVA statistic of accuracy (acc) compared with the No-Information Rate (NIR). When accuracy is smaller than NIR, the *p* value is 1.000. NIR was 0.668, 0.714, 744, and 0.742 for year 2-, 4-, 6-, and 9-year outcome, respectively. I denotes healthy. II denotes mood disorder. III denotes anxiety disorder. IV denotes comorbid mood/anxiety disorder.

pplementary Table 5 Predicting mental health status (categorical outcome) at 2-, 4-, 6-, and 9-year follow up using baseline and 1-year sum and item-scores as the independent variables (i.e., edictor Set D).

AUTO-SKLEARN	Outcome at Year 2				Outcome at Year 4				Outcome at Year 6				Outcome at Year 9			
ealth status	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**	**I**	**II**	**III**	**IV**
aseline sum scores and individual items																
accuracy	0.704				0.718				0.747				0.748			
95% CI	0.679 - 0.729				0.691 - 0.743				0.721 - 0.772				0.720 - 0.774			
p value [acc > NIR)	0.003				0.413				0.434				0.350			
balanced accuracy	0.660	0.500	0.595	0.624	0.540	0.500	0.524	0.538	0.561	0.500	0.523	0.534	0.539	0.500	0.519	0.534
sensitivity	0.948	0.000	0.239	0.299	0.980	0.000	0.067	0.084	0.983	0.000	0.071	0.079	0.992	0.000	0.047	0.077
specificity	0.371	1.000	0.950	0.949	0.099	0.999	0.980	0.992	0.139	1.000	0.974	0.989	0.086	1.000	0.991	0.991
positive predictive value	0.752	0.000	0.486	0.410	0.731	0.000	0.375	0.500	0.769	0.000	0.306	0.333	0.757	0.000	0.467	0.357
negative predictive value	0.780	0.938	0.864	0.920	0.667	0.953	0.857	0.917	0.741	0.949	0.868	0.937	0.793	0.950	0.860	0.941
aive Bayes classifier																
aseline sum scores and individual items																
accuracy	0.621				0.648				0.655				0.608			
95% CI	0.594 - 0.647				0.620 - 0.675				0.627 - 0.683				0.577 - 0.638			
p value [acc > NIR)	<0.999				1.000				1.000				1.000			
balanced accuracy	0.748	0.598	0.584	0.664	0.731	0.537	0.610	0.684	0.727	0.608	0.603	0.656	0.686	0.621	0.576	0.639
sensitivity	0.764	0.296	0.277	0.445	0.770	0.158	0.318	0.477	0.767	0.298	0.290	0.434	0.712	0.327	0.247	0.431
specificity	0.733	0.900	0.890	0.883	0.691	0.915	0.901	0.891	0.688	0.918	0.915	0.878	0.659	0.916	0.906	0.846
positive predictive value	0.852	0.164	0.331	0.310	0.862	0.085	0.361	0.300	0.877	0.162	0.352	0.205	0.857	0.170	0.308	0.158
negative predictive value	0.607	0.951	0.863	0.931	0.546	0.956	0.883	0.946	0.503	0.961	0.890	0.956	0.443	0.963	0.877	0.957
gistic regression																
aseline sum scores and individual items																
accuracy	0.657				0.694				0.708				0.696			
95% CI	0.631 - 0.683				0.668 - 0.720				0.681 - 0.735				0.667 - 0.724			
p value [acc > NIR)	0.804				0.941				0.997				1.000			
balanced accuracy	0.686	0.534	0.587	0.603	0.624	0.493	0.575	0.559	0.598	0.524	0.549	0.560	0.591	0.512	0.546	0.525
sensitivity	0.862	0.099	0.286	0.270	0.907	0.018	0.218	0.150	0.905	0.088	0.142	0.158	0.897	0.058	0.147	0.092
specificity	0.510	0.970	0.887	0.936	0.341	0.968	0.932	0.969	0.292	0.960	0.956	0.963	0.285	0.965	0.946	0.957
positive predictive value	0.780	0.178	0.332	0.333	0.775	0.026	0.358	0.320	0.788	0.104	0.338	0.235	0.783	0.081	0.314	0.125
negative predictive value	0.647	0.942	0.864	0.916	0.594	0.952	0.872	0.921	0.512	0.952	0.875	0.941	0.490	0.951	0.867	0.940

ote. The *p* value denotes the one sided ANOVA statistic of accuracy (acc) compared with the No-Information Rate (NIR). When accuracy is smaller than NIR, the *p* value is 1.000. NIR was 0.668, 714, 0.744, and 0.742 for year 2-, 4-, 6-, and 9-year outcome, respectively. I denotes healthy. II denotes mood disorder. III denotes anxiety disorder. IV denotes comorbid mood/anxiety disorder.


TRUE VALUES
Pedictor Set A
Predictor Set B
Predictor Set C
Predictor Set D
Healthy
Disorder
Logistic regression
Naïve Bayes
AUTO-SKLEARN
767 176 102 255
723 185 144 246
754 191 113 240
724 191 143 240
689 136 178 295
707 139 160 292
690 141 177 290
686 123 181 302
725 158 142 273
733 153 134 278
761 201 106 230
773 201 94 230
Healthy
Mood disorder
Anxiety disorder
Comorbidity
Logistic regression
Naïve Bayes
AUTO-SKLEARN
799 55 112 57 4 6 7 3 49 10 71 40 15 10 23 37
695 44 83 47 54 11 15 21 80 14 73 34 38 12 42 35
791 63 117 66 2 3 4 3 60 9 65 38 14 6 27 30
747 52 107 52 14 8 11 12 78 9 61 36 28 12 34 37
678 37 66 24 75 26 30 28 92 10 73 27 22 8 44 58
677 31 64 24 60 21 18 26 87 11 67 17 43 15 64 70
681 41 65 27 58 21 28 31 101 8 73 32 27 11 47 47
662 31 64 20 71 24 19 32 85 10 59 24 49 16 71 61
835 68 136 70 0 0 0 0 20 7 48 18 12 6 29 49
834 72 136 72 0 0 0 0 22 4 55 30 11 5 22 35
821 71 131 80 0 0 0 0 23 3 45 16 23 7 37 41
822 67 131 73 0 0 0 0 24 7 51 23 21 7 31 41

Supplementary Figure 1. Confusion Matrixes.

Upper confusion matrices depict the binary predictions, that is, (mentally) healthy or mood/anxiety disorder. The lower confusion matrices depict the categorical predictions, that is, (mentally) healthy, mood disorder, anxiety disorder, comorbid mood and anxiety disorder. The number in each cell describes the number of predicted diagnostic categories (*y*-axis) in relation to the true diagnostic categories (*x*-axis). The black borders depicts the correctly classified participants (i.e., true positive and true negative values). All predictor sets included baseline psychiatric diagnoses and demographic variables. Predictor Set A further includes baseline and 1-year follow-up sum scores. Predictor Set B additionally includes baseline and 1-year follow-up individual items. Predictor Set C includes baseline sum scores. Predictor Set D additionally includes individual items.

Chapter 8

Summary and General Discussion

The present dissertation aimed to expand our knowledge of depression by researching its symptom-specific longitudinal characteristics, its predictive factors, and methods for predicting depression and anxiety while taking individual symptoms into account. This dissertation mainly focused on depression, although anxiety has been studied as well, as anxiety is highly prevalent in patients with depression and share a common etiology. The following main research question was formulated: *Can major depressive disorder be characterized as a unified syndrome?* To answer this question we assessed the course of individual depressive symptoms over time (chapter 2), the relation between risk factors and the course of induvial symptoms and symptom domains of depression and anxiety (chapter 3-6), and examined if advanced statistical methods were more adequate to handle depression heterogeneity (chapter 7). We hypothesised that depression is a disorder with substantial within-person heterogeneity between symptoms in terms of intercepts, slopes, and variability. We expected that risk factors are associated with the course of specific symptoms, rather than depression as a homogeneous construct, with similar associations for each symptom. More specifically, we hypothesized that low-grade inflammation inflammatory markers demonstrate the strongest associations with symptoms that overlap with sickness behaviour. Lastly, we hypothesized that machine learning techniques are better in detecting complex patterns in the data and would outdo traditional regression analysis techniques and achieve higher levels of accuracy when predicting the course and onset of depression and anxiety, particularly when symptom-specific features of current depression and anxiety are included to predict future disorders. The first part of the present chapter will provide a summary of our findings. In the second half this chapter these findings will be discussed in light of the current literature, clinical implications and future research directions will be discussed.

8.1 Summary

Depression shows a large heterogeneity of symptoms between and within persons over time. However, most outcome studies have assessed depression as a single underlying latent construct, using the sum score on psychometric scales as a total indicator for depression severity. In **chapter 2,** we assessed the longitudinal symptom-specific trajectories and within-person variability of major depressive disorder over a 9-year period. The highest baseline severity scores were found for the items regarding energy and mood states. The core symptoms of depressed mood and anhedonia had the most favourable course, whereas sleeping problems and (psycho-) somatic symptoms were more persistent over 9-years follow-up. Within-person variability was highest for symptoms related to energy and lowest for suicidal ideation. The severity, course, and within-person variability differed remarkedly between depressive symptoms. Therefore, addressing depression at the syndrome level may obscure insights into both patient and symptom-specific characteristics. Our findings strengthen the idea that employing a symptom-focused approach in both clinical care and research is of value.

Individual symptoms demonstrate heterogeneity in their course over time, but this symptom-specific course is also related to different predictive factors. Preceding chronic depression and neuroticism are two of the most well established predictive factors for the course of depression. However, symptom-specific prospective studies are scarce. In **chapter 3,** we assessed if chronicity (i.e., being depressed for 24 months during a patient's preceding 48 months before baseline) and neuroticism at baseline could predict adverse course trajectories over 9 years of follow up with differential magnitudes for individual depressive symptoms. We found that patients with chronic depression or high levels of neuroticism showed similar absolute rates of decline over time compared to their counterparts. However, because symptoms had higher starting points for mood, cognitive, and somatic/vegetative symptoms (in that order), symptom severity remained higher over time. Findings for the effects of chronicity and neuroticism were remarkably similar, even when assessing the independent associations of both variables. Chronicity and neuroticism predict long-term persistence of diverse psychiatric symptoms, in particular low self-esteem and high interpersonal sensitivity.

Although neuroticism and chronicity are two of the most well-established predictor variables, current psychiatric symptoms are maybe the strongest predictor of all. Although this seems obvious, this is often ignored in scientific literature and previous studies have often failed to take baseline severity into account when assessing the effects of personality pathology. In **chapter 4,** we assessed the prognostic value of personality pathology (e.g. Emotional Dysregulation, measured with DAPP-SF) on treatment outcome among patients with depressive and/or anxiety disorders. Baseline symptom level (BSI-pretreatment) was considered as a mediator- or moderator variable. We found that personality pathology was strongly and significantly associated with treatment outcome. At first glance, this suggests that dimensional levels of personality pathology had a significant and seemingly clinically relevant effect on treatment outcome. However, when taking baseline symptom level into account, we found that patients with high symptom levels at baseline had substantially higher symptom levels after treatment, regardless of personality pathology level. These findings support our hypothesis that baseline symptom level was an important mediator. Furthermore, we found that the baseline symptom level also statistically moderated the predictive effects of Emotional Dysregulation and Inhibition, which were slightly more predictive of treatment outcome among participants with high baseline symptom level. However, the effect sizes of these interaction terms were small.

Besides psychological variables, we also assessed symptom-specific associations with biological variables. Multiple studies demonstrated an association between inflammatory markers and MDD. A cross-sectional relationship between low-grade inflammation and anxiety has also been reported, but the potential longitudinal relationship has been less well studied. People with chronic low-grade inflammation may be at an increased risk of MDD, often in the form of sickness behaviours. We hypothesized that inflammation is predictive of the severity and the course of a subset of MDD symptoms, especially symptoms that overlap with sickness behaviour, such as anhedonia, anorexia, low concentration, low energy, loss of libido, psychomotor slowness, irritability, and malaise. In **chapter 5** and **chapter 6** we tested the association between basal and lipopolysaccharide (LPS)-induced inflammatory markers with individual MDD symptoms and symptom domains of anxiety over a period of up to 9 years. We found that basal and LPS-stimulated inflammatory markers were more strongly associated with sickness behaviour symptoms over the course of 9-year follow up, compared

to non-sickness behaviour symptoms of depression. We also found associations with anxiety symptoms of somatic (arousal) symptoms and agoraphobia. However, the associations were attenuated by 25%-30% after adjusting for the presence of (comorbid) MDD, and the effect sizes of these associations were small. Inflammation was not related to depression as a unified syndrome but rather to the presence and the course of specific MDD symptoms, of which the majority were related to sickness behaviour. It is likely that many of the associations we found have to do with lifestyle and disease-related variables, as these factors are thought to be part of the causal pathway. Afterall, variables related to somatic diseases (e.g. obesity) may induce sickness behaviour, which includes (lifestyle) changes such as a decrease in physical activity. Another line of thought is that these somatic and lifestyle factors act as confounding variables as they are both related to inflammation and depression, though our conclusions remained when we adjusted our findings for the presence of chronic somatic diseases. Moreover, the fact that inflammation seems to be associated with symptoms related to sickness behaviour with the strongest magnitudes, suggests that the sickness behaviour theory is probable.

Due to the heterogeneity of depression and anxiety, predicting the onset and course of mood and anxiety disorders is of clinical importance but remains difficult. Perhaps more advanced statistical models are better suited to handle the complexity of mood and anxiety disorders and improve predictive accuracy. In **chapter 7,** we compared the predictive performances of traditional logistic regression, basic probabilistic machine learning methods, and advanced automated machine learning (Auto-sklearn). We compared how well multinomial logistic regression, a naïve Bayes classifier, and Auto-sklearn predicted depression and anxiety diagnoses at a 2-, 4-, 6-, and 9-year follow up, operationalized as binary or categorical variables. Predictor sets included demographic and self-report data, which can be easily collected in clinical practice at two initial time points (baseline and 1-year follow up). We additionally included predictor sets that took the current individual symptoms (item-scores) into account. The three methods were similarly successful in predicting (mental) health status, with correct predictions for up to 79% (95% CI 75–81%). When assessing a more complex dataset with individual item scores Auto-sklearn was superior but did not result in higher accuracy levels. Against our expectations, more advanced methods of automated machine learning added only limited value, compared to traditional data modelling, when predicting the onset and course of depression and anxiety.

8.2 General Discussion

8.2.1 Is the course of individual depressive symptoms uniform over time?

Although most studies approach depression as a unified construct, we found substantial heterogeneity between depressive symptoms in terms of symptom severity at baseline (i.e., intercepts), slopes over time, and within-person variability over time [1-3]. These findings are consistent with previous literature [4, 5], although in contrast with others [2, 6, 7].

Outcome measurements are generally based on a questionnaire sum score, in which the same weight is given to each of its items. This method would be valid in view of classic test theory; if MDD was a unified construct and all its symptoms contributed equally to its latent construct [8, 9]. However, MDD is unlikely to be a distinct illness with homogeneous symptomatology [8, 10, 11] and the symptom-specific severity, slopes and variability show that symptoms are not diagnostically equivalent and are not interchangeable [12]. Rather, MDD consists of individual symptoms that behave differently over time. These symptoms influence each other with different magnitudes on group level, but also may change within individuals over time [13].

The dynamic nature of these symptom profiles raise the question whether using a sum score of self-report questionnaires does justice to the heterogeneity between symptoms. The use of sum scores to estimate depression severity obscures insight into both patient- and symptom-specific characteristics and can lead to serious misinterpretations regarding depressive severity over time [8, 14]. For example, a patient who recovers by feeling less depressed will show a similar change in the depressive severity measure as a patient whose recovery takes place in another symptom domain, such as sleep. A clinically important change might be obscured by more trivial changes on other items.

In general, depression treatment focuses mainly on the core symptoms of depression. However, a more symptom-specific approach would reveal that other symptoms (e.g. sleeping problems) are more persistent. These residual symptoms are relevant, as they are known to form a risk factor for relapse and worse overall treatment outcome [15, 16]. Other techniques for measuring the course of depression symptomatology are needed and being developed, such as network analysis [17] and dynamic time warp analyses [18].

8.2.2 Are individual symptoms of depression related to the same risk factors?

If depression truly represents one unified latent disorder, all risk factors would have affected the individual symptoms with similar effect sizes. However, two comprehensive studies have demonstrated that individual symptoms have different risk factors [19, 20]. We extended these findings and demonstrated that history of chronic depression, neuroticism, and inflammation is not related to depression as a whole, but rather with specific symptoms with varying magnitudes. Our findings are discussed in the following paragraphs.

8.2.2.1 Preceding chronicity and neuroticism

Two of the most established prognostic factors for depression are a preceding chronic course and neuroticism. We found that a history of chronic depression at baseline was a predictor for the severity of most individual symptoms during 9 years of follow-up of MDD patients, albeit of varying magnitudes. Surprisingly, findings for the effects of chronicity and neuroticism were remarkably similar. Both baseline variables independently predicted an adverse course of symptoms of mood and cognitive symptom clusters, demonstrating the strongest link to 'low self-esteem' and 'interpersonal sensitivity'. The similar results for chronicity and neuroticism in relation to these two symptoms seem to suggest that either these symptoms might cause each other, or that a third dimension (e.g., general severity of MDD, chronic arousal and stress activation, or social isolation) underlies the reported relationships, or both. Although no longer in practice since the introduction of the DSM-III, our findings are relevant in light of a proposition to revive neurotic depression, a subtype of depression which is reactive to life events, persistent, and unlikely to benefit from antidepressants [21]. In light of one modern view of depression as a network of symptoms with between symptom causalities, it is likely that symptoms of low self-esteem and interpersonal sensitivity may be central in the network of patients with a neurotic-like expression of depression [12]. Low self-esteem and high levels of interpersonal sensitivity can play a role in the overall persistence and relapse of depression [22-25].

These findings are also interesting in light of an evolutionary approach of psychiatry. Within this approach, it is thought that the function of emotions is that they create a special state in an organism that allows it to cope effectively with adaptive challenges [26, 27]. In certain situations the effort of pursuing a goal does not match the potential benefits of success.

Feelings of low mood, anhedonia, and lack of energy may be beneficial in these circumstances, as they downregulate the tendency to put effort into the pursuit of unreachable goals, also known as the "*regulation of effort*" [26]. However, depression consists of more than these core symptoms, such as symptoms of increased interpersonal sensitivity and low self-esteem. Perhaps, specific symptoms have different functions for specific adaptive challenges. Price, among others, formulated the *social competition hypothesis of depression* [28, 29]. In this theory, symptoms of negative affect serve as signals in conflicts of hierarchy. In line with this theory, self-deception about one's abilities (low self-esteem) induces dominant others into thinking the individual is no threat. Perhaps symptoms of low self-esteem and feelings of worthlessness might specifically be induced by situations in which it is better to inhibit striving, signal submission and a wish for reconciliation [30-34]. In relation to our findings, perhaps in a subgroup of chronic patients with high levels of neuroticism, interpersonal relations are particularly problematic, therefore leading to symptoms of low self-esteem. Or, difficulties in interpersonal relationships are experienced as more stressful, which is in line with our findings of increased levels of interpersonal sensitivity. One can imagine that among patients with high level of neuroticism, symptoms of high interpersonal sensitivity and low self-esteem tend to bidirectionally influence each other, which could lead to a chronic course.

8.2.2.2 Personality pathology and symptom levels

Personality pathology and depression are two highly intercorrelated constructs. We demonstrated that dimensional personality pathology constructs had a significant and seemingly clinically relevant effect on treatment outcome of patients with a depression or anxiety disorder. Our results replicate findings from previous studies, in which personality pathology was found to have a negative impact on treatment outcome in patients with anxiety and depressive disorders [35-39]. However, high symptom levels at baseline resulted in substantially higher symptom levels after treatment, regardless of personality pathology levels. It is plausible that personality pathology has less prognostic value when researchers would adjust for baseline symptom levels [35-41].

The presentation and expression of personality pathology and depression/anxiety are known to bidirectionally influence each other [42, 43]. Personality pathology cause patients to respond to stress with (or relapse in) higher levels of depression and anxiety. Patients who

report lower (depression) symptom levels after treatment also display a decrease in levels of personality pathology [44]. Patients who are very anxious or depressed may fail to provide accurate self-descriptions [42, 45, 46]. Clearly, the depressive symptom of feeling worthless would influence self-descriptions of self-esteem and vice versa. Moreover, social anxiety symptoms would influence patients descriptions of interpersonal sensitivity and vice versa. In this regard, to some extent, personality pathology and depression/anxiety can in part be manifestations of one and the same underlying common spectrum [42].

8.2.2.3 Inflammation and mood states

We demonstrated that basal inflammatory markers and the LPS-induced inflammatory markers predicted specific depressive symptoms over the course of 9 years. Also associations with somatic (arousal) symptoms of anxiety and agoraphobia were found, although part of these relationships tended to be explained by MDD comorbidity. Our findings are largely consistent with previous findings; signs of low-grade inflammation at baseline were associated with the long-term symptomatology of sickness behavior [47], which may explain some of the symptoms in certain cases of depression [48-50]. The sickness-behaviour theory may (partly) explain the relation between inflammation and depression. More specifically, this theory states that somatic triggers induce an inflammatory response accompanied by sickness behaviour, which include reward oriented behavioural and motivational changes [47, 51-53]. These behavioural changes also are thought to hold some evolutionary advantages as they may protect the individual and facilitate recovery, by preserving energy resources needed for healing infection or other diseases and may help to prevent the transmission of its potential infectious agent to kin [47, 53]. Sickness behaviour (including lifestyle factors such as lower activity) is related to, and is part of, the depressive symptomatology [54-56]. However, when depression is approached on a syndrome level the relation is often rather weak or sometimes conflicting [47]. Inflammation may only be predominantly related to symptoms of sickness behaviour that overlap with those of mood disorders, which demonstrates the importance of symptom-specific research. This was recently confirmed with a pooled analysis in which 15 studies, of which ours, were included [57]. This demonstrated stronger associations between CRP and IL-6 and symptoms that were related to sickness behaviour, such as physical symptoms (e.g. loss of energy) and anhedonia.

Symptoms that were not related to sickness behaviour demonstrated smaller, or no associations with CRP and IL-6 [57].

8.2.3 Are advanced statistical methods more adequate to handle depression heterogeneity?

Besides assessing the added value of symptom-specific predictions of depression course, we also assessed whether improving statistical methods could improve predictive accuracy. Although we earlier approached individual symptoms as outcomes, a current symptom profile might as well predict depression or anxiety at follow-up [2]. In line with our increasing understanding of the complexity and heterogeneity of affective disorders, we expected that complex patterns exist in the data (including nonlinear and higher dimensional), which can be detected when analyzing all available data regarding individual symptoms and multiple variables simultaneously [58, 59]. Although we hypothesized using more advanced machine learning methods would be better suited for this task and would outperform simpler and more traditional data models, our research could not be concluded unequivocally. In fact, in line with an earlier study, we found that depending on the set of predictor variables, more complex machine learning methods do not necessarily result in higher levels of accuracy when predicting future outcomes of affective disorders [60].

Although expectations that machine learning methods will one day unravel the complex nature of psychiatry are still high, recent studies found that machine learning was only of limited added value in research compared to traditional regression models [61-63], and is limited in its clinical usefulness [64]. Within other fields, the proposed added value of machine learning is increasingly criticized [e.g., 62, 65, 66]. That aside, our findings as well as the literature suggest that machine learning might hold some benefits, especially when handling large and complex datasets [67]. Perhaps, the complexity and random chance effects, and therefore our inability to predict, is an inherent part of the nature of affective disorders, rather than a result of errors in our measuring and statistical methods. Although some progress in predicting psychiatry is still likely to be made, and might even be of some clinical usefulness (e.g. [68]), large accuracy levels are likely difficult to accomplish [69]. Small events could lead to dramatic changes in behaviour over time (also known as butterfly effects), such as certain childhood experiences or a treatment intervention in an early stage of the disease [70]. Moreover, the courses of psychiatric disorders are vastly influenced by factors outside

of mental healthcare such as individual choices and circumstances in social, economic, and lifestyles. More advanced models and more elaborate datasets might not be able to solve this. The field of psychiatry may benefit from acknowledging its chaos and complexity, while avoiding defeatism [71, 72].

8.2.4 Clinical implications and future research

Our results regarding symptom-specific associations with risk factors might contribute to the develoment of symptom-specific personalized treatments in the future. Moreover, we hope to have contributed to better understand the relation between inflammation and depression. However, as we made use of data from two cohortstudies without testing the use of certain interventions in clinical practice, we wish to be modest when it comes to giving advise for clinical implementations. Moreover, symptom-specific research on intervention level is only beginning to emerge and much more research is needed. Therefore, we integrated possible clinical implementations of our results with future research recommendations in the present paragraph.

8.2.4.1 Core symptoms

For clinical practice and research, more emphasis should be laid on the subjectively experienced phenomenology of symptoms instead of syndromes. When seeking help, patients do not describe that they experience a particular disorder, but instead they describe symptoms (e.g. “I feel depressed all the time”; “I can’t sleep”; “When I am in the supermarket, it feels like I am going to have a heart attack”). In theory, clinicians should then ask about DSM-5 criteria to classify patients. For example, when a patient is complaining about a depressed mood, clinicians should check if the patient has at least five out of nine symptoms. In practice, however, clinicians under time constraints want to provide care and not to categorize. Perhaps focusing on the reported core symptoms might be more important [73].

Research on personalized medicine in mental health care [74-76] and treatment of specific (residual) symptoms has highlighted that a symptom-specific approach may be beneficial [77-79]. Because a causal relationship exists between symptoms [80, 81], targeting the key symptoms (i.e., more central in the causal network of depressive symptoms) in clinical care may benefit a patient’s recovery [82]. Patients with similar DSM-5 classification may often have similar symptoms that are central in their symptomatology. For example, for most

patients with MDD, central symptoms would be a sad mood and anhedonia, although research also demonstrated that loss of energy is a highly central symptom [81]. For panic disorder, this might be “fear of internal sensations of physical arrousal” [83]. For generalized anxiety disorder, this often is “rumination”. For social anxiety, this often is “fear of social rejection”. However, most of these assumed “central” core symptoms are not researched suffiently with longitudinal network analyses. Most of these studies have used cross-sectional approaches, on the group level.

Although some stereotypical core symptoms per disorder could probably be identified on the group level, patients differ substantially on the individual level. Individual patients vary in the symptoms that are most central in their symptomatology. Only recently have idiographic analysis techniques been used more frequently to study time series of depressive symptoms in a single patient [84]. Especially when taking the vast comorbidity between depression and anxiety into account; patient A may experience a sad mood as a reaction to prolonged symptoms of panic, and patient B may experience panic after increasing levels of persistent sad mood. Patient A may thus benefit more from targeting panic in therapy than patient B. Moreover, other symptoms (e.g., sleeping problems) may be more persistent and can be a risk factor for relapse; therefore, it might be important to identify these symptoms in later stages of treatment [15, 16].

A new field of research is beginning to emerge in which patient-specific symptom networks are assessed [85]. In order to identify these networks, a patient is asked to report their symptoms over the course of several weeks, multiple times a day [86]. This method of intensive, acute, and real-life measurement is also known as ecological momentary assessment (EMA) [87]. This produces a rich dataset that allowes to assess which symptom potentially causes other symptoms, and therefore might be important to target with a personalized treatment. Altough this method is innovative and promissing, the vast effort that is needed by the patient makes it less likely to be implemented on a large scale in clinical practice. Novel analytical techniques are required to analyze panel data and time series data with a less intensive number of assessment [88], such as using Dynamic Time Warp [18, 89].

More research is also needed in order to assess other methods of determining patient-specific central symptoms. Paulhus and Vazire (2005, p. 227 [90]) stated that “no one else has access to more information than oneself”. Perhaps patients are able to assess their own central

symptoms when aided by professionals and a self-report questionnaire. Clinical practice may benefit from interview guidelines to identify patient-specific central symptoms through anamnesis and self-report. As is demonstrated with the Leiden Index of Depression Sensivitiy (LEIDS), patients are willing and able to self-report on their cognivity reactivity without mood induction [91]. Research is needed to assess if patients are able to report on the symptoms that are central in their depression.

8.2.4.2 Symptom-specific treatments

Although our current treatments often approach depression and anxiety on syndrome level when researched and implemented, in reality they are often already symptom-specific. The first-choice antidepressant (Selective Serotonin Reuptake Inhibitor) has demonstrated to have an effect on sadness and anhedonia that is more than twice as high, compared to the other symptoms of depression [92]. Furthermore, antidepressants even produce as negative side effects certain depression related symptoms, such as weight gain, sleeping problems, and psychomotor problems [8]. Symptom-specific cognitive behavioural therapy and pharmacological treatment, for instance, for insomnia appears to have a positive effect on depression as a whole [93, 94]. Multiple evidence-based treatments are available for the symptoms of low self-esteem, such as Competitive Memory Training [COMET; 95, 96] and mindfulness-based cognitive behavioural therapy [97, 98]. Interpersonal sensitivity is an important treatment target in interpersonal therapy [99]. Of course, keeping in mind depression as a network of symptoms, treating one symptom will likely effect other symptoms of depression, although not necessarily the full syndrome. Though, it might be beneficial to treat the person-specific “core symptom” first, before treating symptoms that are less central in the patients network [13]. More research is needed to assess the symptom-specific effects of these treatments, and whether a personalized symptom-specific treatment approach is indeed beneficial for the patient [79].

We found that inflammatory markers are related to specific depressive symptoms that overlap with sickness behaviour. Not all patients exhibit symptoms related to sickness behaviour, and only one third of MDD patients exhibit elevated inflammatory markers [100]. Our findings could have implications for anti-inflammatory treatment [101, 102] and personalized care [103-106]. Perhaps symptom-specific strategies could be developed in order to detect the subgroup of depressed patients for which anti-inflammatory treatments

could be valuable [107]. Instead of treating whole groups of patients with these interventions, only specific patients should be targeted that exhibit sickness related depressive symptoms [57]. Subsequently, inflammatory markers could be assessed before treating them with anti-inflammatory medication [108]. More research is needed in order to test the feasibility of this personalized medicine approach.

Based on our research, candidates for sickness-behaviour related depressive symptoms that also demonstrated significant association with most inflammatory markers are demonstrated in table 2. These symptoms include DSM-5 symptoms or IDS-SR symptoms that are often found in patients with MDD.

Table 2. Sickness-behaviour depressive symptoms that could be indicative elevated inflammatory markers *

1. Low energy
2. Psychomotor retardation
3. Anhedonia
4. Hyposomnia
5. Reduced libido
6. Leaden paralysis
7. Changes in appetite
8. Chances in weight
9. Somatic complaints, e.g. aches, pains and bowel problems

**More research is needed before clinical implication*

8.2.4.3 Using statistics in clinical practice

Deciding what information to collect from patients and making predictions on the micro level are important aspects of a clinician's skill set. This includes predictions regarding suicide risk, violence, the efficacy of treatment options, and the prognoses on the course of disorders [109]. The accuracy of these predictions is of vital importance for individual patients. Two major approaches to predict clinical outcomes can be identified: the clinical and the statistical method. The clinical approach refers to an informal and intuitive process. A clinician's experience, mentalization, and theoretical perspective combined with patient characteristics and circumstances determine how that clinician recalls and interprets these bits of information [109]. With a statistical approach, statistical methods are applied on objectively

measured variables in order to make predictions and prognoses based on probabilities [109]. Two meta-analyses demonstrated that statistical approaches were more accurate than clinical methods [109, 110]. In this dissertation, we demonstrated that moderate levels of accuracy can be accomplished based on data that can be easily collected in clinical practice, confirming that integrating statistical methods into clinical decision making could have an added benefit. Current mental healthcare is already partly digitalized, and the development of automated digital tools to assist clinicians should be attainable, providing clinicians with fast and cheap support in decision making. However, statistical reasoning may have certain ethical and clinical disadvantages, such as the inability to take into account patient specific circumstances. This could potentially lead to an inequality in access to care and stigmatisation [111]. Although Automated ML might be usefulness in healthcare practice [112], it should be used to assist and not to replace a clinicians decision-making.

A first step in the process toward statistically assisted clinical decision-making could be to bring more awareness about base rates into clinical practice. Research demonstrated that clinicians are often not aware of, or ignore, base rates and instead focus on patient-specific characteristics when making predictions. This is also known as the *base rate fallacy* [113, 114]. Using base rates when making clinical discissions is fundamental for clinical decision-making [114]. It can provide rough predictions for the prognosis of a disease, which could be important to take into account for both the clinician as the patient. Moreover, it could help to estimate the quality of care. For example, It would be important to notice when the percentage of successful CBT treatments goes down or is lower in one department compared with others [115]. However, calculating region-specific or clinic-specific base rates could be important.

8.2.5 Limitations

Some main limitations of our research need to be discussed.

- In both the NESDA and the Leiden Routine Outcome Monitoring study datasets, patients were selected at baseline when they met criteria for DSM disorders. Therefore, our data is subject to regression to the mean effects which resulted in a strong initial decrease in symptoms for most patients [116]. Although we tried to take baseline severity into account when assessing the course of symptoms over time, it is possible that patients were selected based on certain high (core) symptoms, which could therefore have coloured our findings.
- Individual symptoms of depression were assessed with items of the IDS-SR. Assessing individual symptoms based on single items presents psychometric hazards. Single items are more strongly affected by random error than sum scores of items [117]. Moreover, the ordinal scores per item are somewhat arbitrary and might differ in weight per item. For example, a score of "2. I think about about suicide or death several times a day", might be a more severe symptom than "2. I can feel the need to move and feel quite restless". Future research should preferable use multi-item measures per symptom such as – among others – the Inventory of Depression and Anxiety Symptoms, which incorporates multiple questions per symptom domain, for instance suicidal ideation is measured with six different items [118].
- Both NESDA and the Leiden Routine Outcome Monitoring Study have gathered limited data on the given treatments. Thus, we could not assess whether some types of treatment (pharmacological or psychological) were more effective with regard to certain variables (e.g., inflammation and neuroticism) than others.
- The time intervals between measures of the NESDA population ranged from a year to two years. We have no data on the course of symptoms between measurements. Therefore, it would be possible that patients remitted and relapsed between measurements.
- Most of our predictor and outcome variables relied on self-report. Self-report measures require patients to possess a certain level of insight, which may be lacking when levels of psychopathology are high, resulting in non-random errors of measurement.

8.2.6 General conclusion

The present dissertation aimed to expand our knowledge of depression by researching the symptom-specific longitudinal characteristics, its risk-factors, and methods for dealing with depression heterogeneity. The following main research question was formulated: *Can major depressive disorder be characterized as a unified syndrome?* Taken these findings together, our answer to this main research question is a resounding *no.* We demonstrated that individual depressive symptoms are not synchronized over time within patients and in groups of patients. We found that individual symptoms of depression are associated to different risk factors, as preceding chronicity, neuroticism, and inflammation were related to individual symptoms with vastly different magnitudes. With this dissertation, we hope to have contributed to the development of alternative ways to define and study depression and its symptoms. We are only at the beginning of a transition from one-fits-all syndromes to patient-specific symptoms. We hope to make a small contribution to the pavement of new ways of personalized symptom-specific treatments [79].

References

1. Fried, E.I., et al., *Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors.* Psychol. Med., 2014. **44**(10): p. 2067-2076.
2. Conradi, H.J., J. Ormel, and J.P. De, *Symptom profiles of DSM-IV-defined remission, recovery, relapse, and recurrence of depression: the role of the core symptoms.* Depress Anxiety, 2012. **29**(7): p. 638-645.
3. Sakurai, H., et al., *Trajectories of individual symptoms in remitters versus non-remitters with depression.* J. Affect. Disord., 2013. **151**(2): p. 506-513.
4. Minor, K.L., J.E. Champion, and I.H. Gotlib, *Stability of DSM-IV criterion symptoms for major depressive disorder.* Journal of Psychiatric Research, 2005. **39**(4): p. 415-420.
5. Conradi, H.J., J. Ormel, and J.P. De, *Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study.* Psychol Med, 2011. **41**(6): p. 1165-1174.
6. Romera, I., et al., *Residual symptoms and functioning in depression, does the type of residual symptom matter? A post-hoc analysis.* BMC Psychiatry, 2013. **13**: p. 51.
7. Harada, E., et al., *Residual symptoms in patients with partial versus complete remission of a major depressive disorder episode: patterns of painful physical symptoms in depression.* Neuropsychiatr Dis Treat, 2016. **12**: p. 1599-1607.
8. Fried, E.I. and R.M. Nesse, *Depression sum-scores don't add up: why analyzing specific depression symptoms is essential.* BMC medicine, 2015. **13**(1): p. 72.
9. Musliner, K.L., et al., *Heterogeneity in long-term trajectories of depressive symptoms: Patterns, predictors and outcomes.* Journal of Affective Disorders, 2016. **192**: p. 199-211.
10. Wardenaar, K.J., et al., *Symptom-specific course trajectories and their determinants in primary care patients with Major Depressive Disorder: Evidence for two etiologically distinct prototypes.* Journal of Affective Disorders, 2015. **179**: p. 38-46.
11. Sjöholm, L., C. Lavebratt, and Y. Forsell, *A multifactorial developmental model for the etiology of major depression in a population-based sample.* Journal of affective disorders, 2009. **113**(1): p. 66-76.
12. Cramer, A.O.J., et al., *Major depression as a complex dynamic system.* PLoS ONE, 2016. **11**(12).
13. McNally, R.J., *Network analysis of psychopathology: Controversies and challenges.* Annual Review of Clinical Psychology, 2021. **17**: p. 31-53.
14. Snaith, P., *What do depression rating scales measure?* The British Journal of Psychiatry, 1993. **163**(3): p. 293-298.
15. Buysse, D.J., et al., *Prevalence, course, and comorbidity of insomnia and depression in young adults.* SLEEP, 2008. **31**(4): p. 473.
16. Lustberg, L. and C.F. Reynolds, *Depression and insomnia: questions of cause and effect.* Sleep medicine reviews, 2000. **4**(3): p. 253-262.
17. Borsboom, D., et al., *Network analysis of multivariate data in psychological science.* Nature Reviews Methods Primers, 2021. **1**(1): p. 1-18.
18. Hebbrecht, K., et al., *Understanding personalized dynamics to inform precision medicine: a dynamic time warp analysis of 255 depressed inpatients.* BMC medicine, 2020. **18**(1): p. 1-15.
19. Fried, et al., *Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors.* Psychol. Med., 2014. **44**(10): p. 2067-2076.
20. Lux, V. and K.S. Kendler, *Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria.* Psychol Med, 2010. **40**(10): p. 1679-90.
21. Nassir Ghaemi, S., *Why antidepressants are not antidepressants: STEP-BD, STAR* D, and the return of neurotic depression.* Bipolar Disorders, 2008. **10**(8): p. 957-968.
22. Davidson, J., et al., *Symptoms of interpersonal sensitivity in depression.* Comprehensive Psychiatry, 1989. **30**(5): p. 357-368.
23. Sowislo, J.F. and U. Orth, *Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies*. 2013, American Psychological Association.
24. Verhoeven, F.E., et al., *Seeing the signs: Using the course of residual depressive symptomatology to predict patterns of relapse and recurrence of major depressive disorder.* Depression and anxiety, 2018. **35**(2): p. 148-159.
25. De Rubeis, J., et al., *Rejection sensitivity as a vulnerability marker for depressive symptom deterioration in men.* PloS one, 2017. **12**(10): p. e0185802.

26. Nesse, R.M., *Is depression an adaptation?* Archives of general psychiatry, 2000. **57**(1): p. 14-20.
27. Nesse, R.M., *Good reasons for bad feelings: insights from the frontier of evolutionary psychiatry*. 2019: Penguin.
28. Price, J. and R. Gardner Jr, *The paradoxical power of the depressed patient: A problem for the ranking theory of depression.* British Journal of Medical Psychology, 1995. **68**(3): p. 193-206.
29. Price, J., et al., *The social competition hypothesis of depression.* The British Journal of Psychiatry, 1994. **164**(3): p. 309-315.
30. Brown, G.W., T.O. Harris, and C. Hepworth, *Loss, humiliation and entrapment among women developing depression: a patient and non-patient comparison.* Psychological medicine, 1995. **25**(1): p. 7-21.
31. Gilbert, P., *Depression: The evolution of powerlessness*. 2016: Routledge.
32. Keller, M.C., M.C. Neale, and K.S. Kendler, *Association of different adverse life events with distinct patterns of depressive symptoms.* Am J Psychiatry, 2007. **164**(10): p. 1521-1529.
33. Keller, M.C. and R.M. Nesse, *The evolutionary significance of depressive symptoms: different adverse situations lead to different depressive symptom patterns.* Journal of personality and social psychology, 2006. **91**(2): p. 316.
34. Keller, M.C. and R.M. Nesse, *Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants.* Journal of affective disorders, 2005. **86**(1): p. 27-35.
35. Schat, A., et al., *Predictors of outcome in outpatients with anxiety disorders: the Leiden routine outcome monitoring study.* Journal of psychiatric research, 2013. **47**(12): p. 1876-1885.
36. Telch, M.J., J.H. Kamphuis, and N.B. Schmidt, *The effects of comorbid personality disorders on cognitive behavioral treatment for panic disorder.* Journal of Psychiatric Research, 2011. **45**(4): p. 469-474.
37. Goddard, E., J. Wingrove, and P. Moran, *The impact of comorbid personality difficulties on response to IAPT treatment for depression and anxiety.* 2015. **73**: p. 1-7.
38. Shea, M.T., et al., *Course of depressive symptoms over follow-up: findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program.* Archives of general psychiatry, 1992. **49**(10): p. 782-787.
39. van den Hout, M., C. Brouwers, and J. Oomen, *Clinically diagnosed axis II co-morbidity and the short term outcome of CBT for axis I disorders.* Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice, 2006. **13**(1): p. 56-63.
40. Mulder, R.T., *Personality pathology and treatment outcome in major depression: a review.* American Journal of Psychiatry, 2002. **159**(3): p. 359-371.
41. Kvarstein, E.H. and S. Karterud, *Large variation of severity and longitudinal change of symptom distress among patients with personality disorders.* Personality and Mental Health, 2013. **7**(4): p. 265-276.
42. Widiger, A.T., *Personality and psychopathology.* World Psychiatry, 2011. **10**(2): p. 103-106.
43. Widiger, T.A. and G.T. Smith, *Personality and psychopathology.* 2008.
44. Costa Jr, P.T., et al., *Personality self-reports are concurrently reliable and valid during acute depressive episodes.* Journal of affective disorders, 2005. **89**(1-3): p. 45-55.
45. Fava, M., et al., *Personality disorder comorbidity with major depression and response to fluoxetine treatment.* Psychotherapy and Psychosomatics, 1994. **62**(3-4): p. 160-167.
46. Gunderson, J.G., et al., *Plausibility and possible determinants of sudden "remissions" in borderline patients.* Psychiatry: Interpersonal and Biological Processes, 2003. **66**(2): p. 111-119.
47. Dantzer, R. and K.W. Kelley, *Twenty years of research on cytokine-induced sickness behavior.* Brain, behavior, and immunity, 2007. **21**(2): p. 153-160.
48. Haroon, E., et al., *Antidepressant Treatment Resistance is Associated with Increased Inflammatory Markers in Patients with Major Depressive Disorder.* Psychoneuroendocrinology, 2018.
49. Duivis, H.E., et al., *Differential association of somatic and cognitive symptoms of depression and anxiety with inflammation: findings from the Netherlands Study of Depression and Anxiety (NESDA).* Psychoneuroendocrinology, 2013. **38**(9): p. 1573-1585.
50. Smith, R.S., *The macrophage theory of depression.* Medical hypotheses, 1991. **35**(4): p. 298-306.
51. Shattuck, E.C. and M.P. Muehlenbein, *Human sickness behavior: Ultimate and proximate explanations.* American Journal of Physical Anthropology, 2015. **157**(1): p. 1-18.
52. Miller, A.H., L. Capuron, and C.L. Raison, *Immunologic influences on emotion regulation.* Clinical Neuroscience Research, 2005. **4**(5-6): p. 325-333.
53. Miller, A.H. and C.L. Raison, *The role of inflammation in depression: from evolutionary imperative to modern treatment target.* Nat Rev Immunol, 2016. **16**(1): p. 22-34.

54. Fried, E.I. and J.M. Haslbeck, *Using network analysis to examine links between individual depression symptoms, inflammatory markers, and covariates.* Psychological Medicine, 2019: p. 1-9.
55. Berk, M., et al., *So depression is an inflammatory disease, but where does the inflammation come from?* BMC medicine, 2013. **11**(1): p. 200.
56. Maes, M., et al., *Depression and sickness behavior are Janus-faced responses to shared inflammatory pathways.* BMC Med, 2012. **10**(1): p. 66.
57. Frank, P., et al., *Association Between Systemic Inflammation and Individual Symptoms of Depression: A Pooled Analysis of 15 Population-Based Cohort Studies.* American Journal of Psychiatry, 2021: p. appi. ajp. 2021.20121776.
58. Hahn, T., A. Nierenberg, and S. Whitfield-Gabrieli, *Predictive analytics in mental health: applications, guidelines, challenges and perspectives.* Molecular psychiatry, 2016. **22**(1): p. 37.
59. Chekroud, A.M., et al., *Cross-trial prediction of treatment outcome in depression: a machine learning approach.* The Lancet Psychiatry, 2016. **3**(3): p. 243-250.
60. Nie, Z., et al., *Predictive modeling of treatment resistant depression using data from STAR*D and an independent clinical study.* PLOS ONE, 2018. **13**(6): p. e0197268.
61. Belsher, B.E., et al., *Prediction models for suicide attempts and deaths: a systematic review and simulation.* JAMA psychiatry, 2019. **76**(6): p. 642-651.
62. Christodoulou, E., et al., *A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models.* Journal of clinical epidemiology, 2019.
63. Mens, K.v., et al., *Predicting Future Suicidal Behaviour with Different Machine Learning Techniques: A Population-Based Longitudinal Study.* OSF, 2019. **April 4.**
64. Tran, B.X., et al., *The current research landscape on the artificial intelligence application in the management of depressive disorders: A bibliometric analysis.* International journal of environmental research and public health, 2019. **16**(12): p. 2150.
65. Frizzell, J.D., et al., *Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure: comparison of machine learning and other statistical approaches.* JAMA cardiology, 2017. **2**(2): p. 204-209.
66. Desai, R.J., et al., *Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes.* JAMA Network Open, 2020. **3**(1): p. e1918962-e1918962.
67. Lee, Y., et al., *Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review.* Journal of affective disorders, 2018. **241**: p. 519-532.
68. van Vuuren, C., et al., *Comparing machine learning to a rule-based approach for predicting suicidal behavior among adolescents: results from a longitudinal population-based survey.* Journal of Affective Disorders, 2021.
69. McMahon, F.J., *Prediction of treatment outcomes in psychiatry—where do we stand?* Dialogues in clinical neuroscience, 2014. **16**(4): p. 455.
70. Nelson, B., et al., *Moving from static to dynamic models of the onset of mental disorder: a review.* JAMA psychiatry, 2017. **74**(5): p. 528-534.
71. Maj, M., *The need for a conceptual framework in psychiatry acknowledging complexity while avoiding defeatism.* World Psychiatry, 2016. **15**(1): p. 1.
72. Maier, T., *Chaos theory and complexity in psychiatry.* Psychotherapie, Psychosomatik, medizinische Psychologie, 1998. **48**(8): p. 314-317.
73. Chevance, A., et al., *Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey.* The Lancet Psychiatry, 2020. **7**(8): p. 692-702.
74. Arnow, B.A., et al., *Depression Subtypes in Predicting Antidepressant Response: A Report From the iSPOT-D Trial.* American Journal of Psychiatry, 2015. **172**(8): p. 743-750.
75. Menza, M., H. Marin, and R.S. Opper, *Residual symptoms in depression: can treatment be symptom-specific?* Journal of Clinical Psychiatry, 2003. **64**(5): p. 516-523.
76. Chekroud, A.M., et al., *Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach.* JAMA psychiatry, 2017. **74**(4): p. 370-378.
77. Fava, M., *Pharmacological approaches to the treatment of residual symptoms.* Journal of Psychopharmacology, 2006. **20**(3 Suppl): p. 29-34.
78. Rafanelli, C., S.K. Park, and G.A. Fava, *New psychotherapeutic approaches to residual symptoms and relapse prevention in unipolar depression.* Clinical Psychology & Psychotherapy, 1999. **6**(3): p. 194-201.

79. Maj, M., et al., *The clinical characterization of the adult patient with depression aimed at personalization of management.* World Psychiatry, 2020. **19**(3): p. 269-293.
80. Beard, C., et al., *Network analysis of depression and anxiety symptom relationships in a psychiatric sample.* Psychol Med, 2016. **46**(16): p. 3359-3369.
81. Fried, E.I., et al., *What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis.* Journal of Affective Disorders, 2016. **189**: p. 314-320.
82. Borkulo, C.v., et al., *Association of symptom network structure with the course of longitudinal depression.(Report).* JAMA Psychiatry, 2015. **72**(12): p. 1219.
83. Robinaugh, D., et al., *Advancing the network theory of mental disorders: A computational model of panic disorder*. 2019.
84. Piccirillo, M.L. and T.L. Rodebaugh, *Foundations of idiographic methods in psychology and applications for psychotherapy.* Clinical Psychology Review, 2019. **71**: p. 90-100.
85. Wichers, M., et al., *De netwerkbenadering bij depressie. Veel noten op de zang of heilige graal?* Gedragstherapie, 2019.
86. Larson, R. and M. Csikszentmihalyi, *The experience sampling method*, in *Flow and the foundations of positive psychology*. 2014, Springer. p. 21-34.
87. Smyth, J.M. and A.A. Stone, *Ecological momentary assessment research in behavioral medicine.* Journal of Happiness studies, 2003. **4**(1): p. 35-52.
88. Hekler, E.B., et al., *Why we need a small data paradigm.* BMC medicine, 2019. **17**(1): p. 1-9.
89. Booij, M.M., et al., *Dynamic time warp analysis of individual symptom trajectories in depressed patients treated with electroconvulsive therapy.* Journal of Affective Disorders, 2021. **293**: p. 435-443.
90. Paulhus, D.L. and S. Vazire, *The self-report method.* Handbook of research methods in personality psychology, 2007. **1**(2007): p. 224-239.
91. Van der Does, W., *Cognitive reactivity to sad mood: structure and validity of a new measure.* Behaviour research and therapy, 2002. **40**(1): p. 105-119.
92. Boschloo, L., et al., *The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression: results from an individual patient data meta-analysis.* World Psychiatry, 2019. **18**(2): p. 183-191.
93. Manber, R., et al., *Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disor-der and insomnia.* Sleep, 2008. **31**(4): p. 489-495.
94. Britton, W.B., et al., *Mindfulness-based cognitive therapy improves polysomnographic and subjective sleep profiles in antidepressant users with sleep complaints.* Psychotherapy and Psychosomatics, 2012. **81**(5): p. 296-304.
95. Korrelboom, K., M. Maarsingh, and I. Huijbrechts, *Competitive memory training (COMET) for treating low self-esteem in patients with depressive disorders: a randomized clinical trail.* Depression and anxiety, 2012. **29**(2): p. 102-110.
96. Korrelboom, K., M. Marissen, and T. van Assendelft, *Competitive memory training (COMET) for low self-esteem in patients with personality disorders: A randomized effectiveness study.* Behavioural and Cognitive Psychotherapy, 2011. **39**(1): p. 1-19.
97. Fjorback, L.O., et al., *Mindfulness-Based Stress Reduction and Mindfulness-Based Cognitive Therapy–a systematic review of randomized controlled trials.* Acta Psychiatrica Scandinavica, 2011. **124**(2): p. 102-119.
98. Thompson, B.L. and J.A. Waltz, *Mindfulness, self-esteem, and unconditional self-acceptance.* Journal of Rational-Emotive & Cognitive-Behavior Therapy, 2008. **26**(2): p. 119-126.
99. Klerman, G.L. and M.M. Weissman, *Interpersonal psychotherapy of depression: A brief, focused, specific strategy*. 1994: Jason Aronson, Incorporated.
100. Raison, C.L. and A.H. Miller, *Is depression an inflammatory disorder?* Current psychiatry reports, 2011. **13**(6): p. 467-475.
101. 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, 2014. **71**(12): p. 1381-1391.
102. Jha, M. and M. Trivedi, *Personalized antidepressant selection and pathway to novel treatments: clinical utility of targeting inflammation.* International journal of molecular sciences, 2018. **19**(1): p. 233.
103. Gimeno, D., et al., *Associations of C-reactive protein and interleukin-6 with cognitive symptoms of depression: 12-year follow-up of the Whitehall II study.* Psychological medicine, 2009. **39**(3): p. 413-423.

104. Au, B., et al., *The longitudinal associations between C-reactive protein and depressive symptoms: evidence from the English Longitudinal Study of Ageing (ELSA).* International journal of geriatric psychiatry, 2015. **30**(9): p. 976-984.
105. Ridker, P.M., et al., *Prospective study of C-reactive protein and the risk of future cardiovascular events among apparently healthy women.* Circulation, 1998. **98**(8): p. 731-733.
106. Ridker, P.M., et al., *C-reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women.* New England journal of medicine, 2000. **342**(12): p. 836-843.
107. Kopschina Feltes, P., et al., *Anti-inflammatory treatment for major depressive disorder: implications for patients with an elevated immune profile and non-responders to standard antidepressant therapy.* Journal of Psychopharmacology, 2017. **31**(9): p. 1149-1165.
108. *Simvastatin add-on to Escitalopram in patients with comorbid obesity and major depression: A multicenter, randomized, double-blind, placebo-controlled trial; EudraCT number 2018-002947-27*. 2019, Charité – Universitätsmedizin Berlin: Germany.
109. Ægisdóttir, S., et al., *The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction.* The Counseling Psychologist, 2006. **34**(3): p. 341-382.
110. Grove, W.M., et al., *Clinical versus mechanical prediction: a meta-analysis.* Psychological assessment, 2000. **12**(1): p. 19.
111. O'neil, C., *Weapons of math destruction: How big data increases inequality and threatens democracy*. 2016: Crown.
112. Waring, J., C. Lindvall, and R. Umeton, *Automated machine learning: Review of the state-of-the-art and opportunities for healthcare.* Artificial Intelligence in Medicine, 2020: p. 101822.
113. Lavigne, J.V., M. Feldman, and K.M. Meyers, *Screening for mental health problems: addressing the base rate fallacy for a sustainable screening program in integrated primary care.* Journal of pediatric psychology, 2016. **41**(10): p. 1081-1090.
114. Andren, K.A.K., et al., *Population base rates and disease course of common psychiatric and neurodegenerative disorders.* Neuropsychology review, 2017. **27**(3): p. 284-301.
115. McKay, R., T. Coombs, and D. Duerden, *The art and science of using routine outcome measurement in mental health benchmarking.* Australasian Psychiatry, 2014. **22**(1): p. 13-18.
116. Streiner, D.v.L., *Regression toward the mean: its etiology, diagnosis, and treatment.* The Canadian Journal of Psychiatry, 2001. **46**(1): p. 72-76.
117. Fried, E.I. and A.O. Cramer, *Moving forward: challenges and directions for psychopathological network theory and methodology.* Perspectives on Psychological Science, 2017. **12**(6): p. 999-1020.
118. Watson, D., et al., *Development and validation of the Inventory of Depression and Anxiety Symptoms (IDAS).* Psychological assessment, 2007. **19**(3): p. 253.

Nederlandse Samenvatting

De symptomatologie van depressie en het beloop van haar symptomen vertonen een grote mate van heterogeniteit tussen patiënten; twee patiënten met beide een depressie kunnen soms maar enkele overlappende symptomen hebben. De meeste studies benaderen depressie echter als een eenduidige stoornis waarbij de somscore op vragenlijsten gebruikt wordt voor het meten van de ernst van de depressie. Dit proefschrift richt zich niet op depressie als syndroom maar op de individuele symptomen van depressie. We stellen ons de vraag: *"Kan een depressieve stoornis worden gekarakteriseerd als een eenduidig syndroom?"*. Wanneer depressie inderdaad een eenduidig homogeen syndroom is, verwachten we dat de individuele symptomen van depressie zich ongeveer hetzelfde gedragen; deze symptomen zijn immers een weerspiegeling van één onderliggende stoornis. Wanneer depressie echter helemaal geen eenduidige homogene stoornis is, verwachten we grotere heterogeniteit in symptomen, hun beloop en hun risicofactoren.

In de volgende 6 hoofdstukken worden de ernst en het beloop van individuele symptomen beschreven. We hebben onderzocht hoe verschillende factoren, zoals chroniciteit, persoonlijkheidsfactoren en inflammatie het beloop van affectieve symptomen voorspellen. Tenslotte hebben we onderzocht of machine learning technieken in staat kunnen zijn om tot een accuratere prognose van depressie en angst te komen in vergelijking met conventionele methodes. In hoofdstuk 2, 3, 5, 6, en 7 hebben we hiervoor gebruik gemaakt van data van de Nederlandste Studie naar Depressie en Angst (NESDA), een langlopende cohortstudie dat 2981 participanten die een depressie of een angststoornis hebben of hebben gehad, of gezonde controles langdurig volgt. Elke 2 tot 4 jaar worden de participanten gevraagd om verschillende vragenlijsten in te vullen, vindt er een interview plaats en wordt biologische data verzameld zoals bloed en speeksel. In dit proefschrift maken we gebruik van follow-up data tot 9 jaar. In hoofdstuk 4 hebben we gebruik gemaakt van data van de Leiden Routine Outcome Monitoring Study (ROM), een lopend cohortonderzoek van het Leids Universitair Medisch Centrum in samenwerking met GGZ Rivierduinen. ROM wordt gebruikt om de voortgang en behandeluitkomst van patiënten in de klinische praktijk te beoordelen. Bij intake, gedurende de behandeling en bij het afsluiten van

de behandeling worden verschillende (voornamelijk zelfrapportage) meetinstrumenten systematisch afgenomen. De patiënten die wij onderzocht hebben hadden allen een depressie en/of angststoornis. De bevindingen van deze studies worden hieronder samengevat.

In **hoofdstuk 2** hebben we het symptoom-specifieke beloop van depressie over een periode van 9 jaar onderzocht. We vonden dat een depressieve stemming en een laag energieniveau de symptomen waren die gemiddeld in de meest ernstige mate voorkwamen gedurende een depressieve episode. De kernsymptomen depressieve stemming en anhedonie hadden het meest gunstige verloop, terwijl slaapproblemen en (psycho-) somatische symptomen persisteerden gedurende 9 jaar follow-up. De variabiliteit van symptomen over tijd was het hoogst voor symptomen die gerelateerd waren aan energieniveau en het laagst voor gedachten aan zelfmoord. De ernst, het beloop en de variabiliteit van individuele symptomen verschilden aanzienlijk tussen patiënten en over tijd. Het benaderen van depressie op syndroomniveau kan belangrijke informatie over de patiënt en zijn of haar individuele symptomen vertroebelen. Onze bevindingen versterken het idee dat het toepassen van een symptoomgerichte benadering in zowel klinische zorg als wetenschappelijk onderzoek waardevol is.

Individuele symptomen kunnen verschillende risicofactoren hebben. Een voorgeschiedenis van chronische depressie en neuroticisme (een persoonlijkheidskenmerk dat wordt gekenmerkt door emotionele instabiliteit) zijn twee van de meest bekende risicofactoren voor depressie. Er is echter weinig onderzoek gedaan naar de symptoom-specifieke relaties van deze risicofactoren. In **hoofdstuk 3** hebben we de voorspellende waarde onderzocht van chroniciteit (d.w.z. 24 maanden depressief zijn tijdens de 48 maanden voorafgaand aan baseline) en neuroticisme op het symptoom-specifieke beloop gedurende 9 jaar follow-up. We vonden dat chroniciteit en neuroticisme met name gerelateerd waren aan de ernst van stemmings- en cognitieve symptomen van depressie en in mindere mate aan somatische/vegetatieve symptomen. Alle symptomen namen af in verloop van tijd maar stemmings- en cognitieve symptomen bleven van hogere ernst over de periode van negen jaar. De sterkste voorspellende relatie werd gevonden bij patiënten met symptomen van een laag zelfbeeld en verhoogde interpersoonlijke sensitiviteit. De bevindingen voor de effecten van chroniciteit en neuroticisme waren opmerkelijk

vergelijkbaar, zelfs wanneer het effect van beide variabelen onafhankelijk van elkaar werden onderzocht.

Hoewel neuroticisme en chroniciteit twee van de meest bekende voorspellende variabelen zijn, zijn de ernst van huidige psychiatrische symptomen misschien wel de sterkste voorspeller voor toekomstig beloop. Immers, een ernstige depressie is veelal moeilijker te behandelen dan milde depressieve klachten. Hoewel dit vanzelfsprekend lijkt, wordt dit vaak genegeerd in de wetenschappelijke literatuur. Eerdere studies hebben vaak geen rekening gehouden met de ernst van symptomen op baseline bij het beoordelen van de effecten van persoonlijkheidspathologie. In **hoofdstuk 4** hebben we de prognostische waarde van persoonlijkheidspathologie (gemeten met DAPP-SF) op het behandelresultaat bij patiënten met depressieve en/of angststoornissen onderzocht. Het baseline symptoomniveau hebben we benaderd als zowel een mediator- als een moderatorvariabele. We vonden dat persoonlijkheidspathologie sterk en significant geassocieerd was met het behandelresultaat. Op het eerste gezicht suggereert dit dat persoonlijkheidspathologie een significant en schijnbaar klinisch relevant effect had op het behandelresultaat. Wanneer we echter rekening hielden met het symptoomniveau op baseline, ontdekten we dat patiënten met hoge symptoomniveau's bij baseline aanzienlijk hogere symptoomniveau's hadden na de behandeling, ongeacht de mate van persoonlijkheidspathologie; patiënten met een hogere mate van persoonlijkheidspathologie rapporteerden een ernstigere depressie of angststoornis, zowel voor als na behandeling. Deze bevindingen ondersteunen onze hypothese dat het symptoomniveau bij aanvang een belangrijke mediator is. Verder vonden we dat het baseline symptoomniveau ook de voorspellende effecten van de persoonlijkheidstrekken van emotionele ontregeling en geremdheid modereerde. Deze trekken waren meer voorspellend voor de behandeluitkomst bij deelnemers met een hoog baseline symptoomniveau. De effectgroottes van deze interactietermen waren echter klein.

Naast psychologische variabelen hebben we ook symptoom-specifieke associaties met biologische variabelen onderzocht. Eerdere studies toonden een verband aan tussen inflammatoire markers en depressie. Een cross-sectionele relatie tussen inflammatoire markers en angst is ook gevonden, maar de mogelijke longitudinale relatie is minder goed bestudeerd. Mensen met chronische verhoogde inflammatoire waarden lopen mogelijk een verhoogd risico

op een depressieve stoornis. Dit hangt mogelijk samen met 'Sickness Behaviour', wat deels adaptieve gedragsveranderingen zijn die plaatsvinden bij lichamelijke inflammatoire ziekten. Onze hypothese was dat inflammatie voorspellend is voor de ernst en het verloop van een subset van depressieve symptomen, met name symptomen die overlappen met Sickness Behaviour, zoals anhedonie, eetgedrag, concentratie, energieniveau, verlies van libido, psychomotorische traagheid, prikkelbaarheid en malaise. In **hoofdstuk 5** en **hoofdstuk 6** hebben we de associatie tussen basale en lipopolysacharide (LPS)-geïnduceerde inflammatoire markers met individuele depressieve symptomen en symptoomdomeinen van angst onderzocht over een periode van 9 jaar. We vonden dat basale en LPS-gestimuleerde inflammatoire markers sterker geassocieerd waren met symptomen van Sickness Behaviour over een periode van 9 jaar follow-up, vergeleken met niet-Sickness Behaviour symptomen van depressie. We vonden ook associaties met angstsymptomen van somatische arousal en pleinvrees. De associaties werden echter met 25%-30% verzwakt na correctie voor de aanwezigheid van (comorbide) depressie. Inflammatie lijkt niet zozeer gerelateerd aan depressie op syndroomniveau, maar eerder aan de aanwezigheid en het beloop van specifieke symptomen, waarvan de meerderheid gerelateerd is aan Sickness Behaviour.

Vanwege de heterogeniteit van depressie en angst, is het voorspellen van het ontstaan en het beloop van depressieve- en angststoornissen erg moeilijk. Misschien zijn meer geavanceerde statistische modellen beter geschikt om met de complexiteit van stemmings- en angststoornissen om te gaan en de accuraatheid van voorspellende modellen te verbeteren. In **hoofdstuk 7** vergeleken we de voorspellende accuraatheid van traditionelere en relatief simpele methodes met geavanceerde geautomatiseerde machine learning (waarvoor we Auto-sklearn gebruikten). We vergeleken hoe goed multinomiale logistische regressie, een naïeve Bayes-classificator en Auto-sklearn depressie- en angstdiagnoses voorspelden bij een follow-up van 2, 4, 6 en 9 jaar, geoperationaliseerd als binaire of categorische variabelen. Predictorsets bevatten demografische en zelfrapportagevariabelen, die relatief gemakkelijk in de klinische praktijk kunnen worden verzameld op twee initiële tijdstippen (baseline en 1-jaars follow-up). Daarnaast hebben we predictorsets opgenomen waarbij huidige individuele symptomen (itemscores) werden meegenomen. Tegen onze verwachting in waren de drie methodes even succesvol in het

voorspellen van de depressie en angst, met een percentage van correcte voorspellingen tot 79% (95% betrouwbaarheidsintervallen: 75-81%). Bij het beoordelen van een complexere dataset met individuele itemscores was Auto-sklearn superieur, maar resulteerde niet in hogere accuraatheid. De geavanceerde methode van geautomatiseerd machine learning voegde slechts weinig toe vergeleken met conventionele datamodellering bij het voorspellen van het begin en het verloop van depressie en angst.

Aan het begin van dit proefschrift stelden we ons de vraag: "*Kan een depressieve stoornis worden gekarakteriseerd als een eenduidig syndroom?*". Wanneer we onze onderzoeksbevindingen samenvoegen, is ons antwoord op deze onderzoeksvraag een volmondig *nee*. Wanneer depressie daadwerkelijk een eenduidig syndroom zou zijn, en de individuele symptomen van depressie een reflectie zijn van een achterliggende eenduidige stoornis, zouden deze symptomen zich vergelijkbaar moeten gedragen. Wij toonden aan dat dit duidelijk niet het geval is. We hebben aangetoond dat individuele depressieve symptomen sterk verschillen tussen patiënten en in hun beloop over tijd. We ontdekten dat individuele symptomen van depressie geassocieerd zijn met verschillende risicofactoren; chroniciteit, persoonlijkheidsfactoren en inflammatie zijn verschillend geassocieerd met het beloop van individuele affectieve symptomen.

Met dit proefschrift hopen we een bijdrage te hebben geleverd aan de ontwikkeling van alternatieve manieren om depressie te definiëren en te bestuderen. We staan nog maar aan het begin van een transitie van one-fits-all syndromen naar patiënt-specifieke symptomen. We hopen dat ons onderzoek kan helpen in de ontwikkeling van meer gepersonaliseerde behandelmethoden.

Dankwoord

Het schrijven van dit proefschrift heeft mij ontwikkeld op zowel wetenschappelijk als persoonlijk vlak. Ik had van tevoren niet kunnen weten wat het schrijven van een proefschrift zou betekenen. Ik heb gedurende mijn promoveren waardevolle collega's en vrienden mogen ontmoeten.

Mijn voornaamste dankwoord gaat uit naar mijn promotor, Bert van Hemert en mijn twee copromotoren Erik Giltay en Ingrid Carlier. Bert, dankjewel voor het faciliteren van mijn promotietraject inclusief de mogelijkheid tot het volgen van cursussen en congressen. Jij liet mij vrij in de vormgeving van dit proefschrift maar ik kon ook altijd op je rekenen voor advies en raad. Erik, ik heb veel van je geleerd. Mijn poging om je werktempo bij te houden heeft de voortgang van mijn proefschrift goed gedaan. Je enthousiasme voor de wetenschap werkt aanstekelijk en inspirerend. Ingrid, jij hebt mij in eerste instantie binnengehaald als stagiair en vervolgens als promovendus. Ik ben je erg dankbaar dat je me hebt overtuigd om dit promotietraject aan te gaan. Een promotietraject is geen weg zonder kuilen, jij bekrachtigde mij door mij te wijzen op de successen en bood steun als dat nodig was.

Ik wil in het bijzonder mijn naaste collega en goede vriendin Stephanie bedanken. Precies 2 weken heb ik als promovendus gewerkt zonder jou als mijn collega. Dit waren twee saaie en lange weken. We hebben de afgelopen jaren veel gelachen en meegemaakt. Ik weet zeker dat nog veel borrels zullen volgen. Ik wil daarnaast mijn medepromovendi bedanken; Ericka, David, Erwin, Floor, Nienke, Rahele, Ikrame en Nancy. Gedurende Covid was de kantoortuin soms erg stil, maar als we samen waren was het nooit saai. Ik hoop dat we elkaar blijven zien en er nog veel dinertjes en drankjes zullen volgen

Ten slotte, Lisa. Zonder jou is mijn leven niet compleet. Dankjewel voor al je steun en voor het dogen van mijn vele avonden achter de laptop.

List of publications

Prognostic Value of Pathological Personality Traits for Treatment Outcome in Anxiety and Depressive Disorders: The Leiden Routine Outcome Monitoring Study. van Eeden, W.A., van Hemert, A.M., Giltay, E.J., Spinhoven, Philip, de Beurs, E., Carlier, I.V.E (2022). The *Journal of Nervous and Mental Disease, 10-1097*

Basal and LPS-stimulated inflammatory markers and the course of anxiety symptoms. van Eeden, W.A.* , El Filali, E.*, van Hemert, A.M., Carlier, I. V., Penninx, B. W., Lamers, F., Schoevers, R., E.J. Giltay (2021). *Brain, Behavior, and Immunity, 98, 378-387.*
*These authors share first authorship

Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression. van Eeden, W. A., Luo, C., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W., Wardenaar, K., Hoos, H. Giltay, E.J. (2021). *Psychiatry Research*, *299*, 113823

Basal and stimulated inflammation markers and the course of individual symptoms in depressed patients. van Eeden, W. A., van Hemert, A. M., Carlier, I. V., Penninx, B. W., Lamers, F., Fried, E., Schroevers, R., Giltay, E.J. (2020). *Translational Psychiatry*, *10*(1), 1-12.

Neuroticism and chronicity as predictors of 9-year course of individual depressive symptoms. van Eeden, W. A., van Hemert, A. M., Carlier, I. V., Penninx, B. W., Spinhoven, P., & Giltay, E. J. (2019). *Journal of Affective Disorders, 252, 484-492.*

Severity, course trajectory, and within-person variability of individual symptoms in patients with major depressive disorder. van Eeden, W. A., van Hemert, A. M., Carlier, I. V. E., Penninx, B. W., & Giltay, E. J. (2019). *Acta Psychiatrica Scandinavica*, *139*(2), 194-205.

Testing for response shift in treatment evaluation of change in self-reported psychopathology amongst secondary psychiatric care outpatients. Carlier, I. V., van Eeden, W. A., de Jong, K., Giltay, E. J., van Noorden, M. S., van der Feltz-Cornelis, C., ... & van Hemert, A. M. (2019). *International Journal of Methods in Psychiatric Research, 28(3), e1785.*
*These authors share first authorship

Routine outcome monitoring in mental health care and particularly in addiction treatment: Evidence-based clinical and research recommendations. Carlier, I. V. E., & van Eeden, W. A. (2017). *Journal of Addiction Research & Therapy, 8, 1-7.*

Curriculum Vitae

Wessel van Eeden is geboren op 1 januari 1992 te Delft. In 2009 behaalde hij zijn havo-diploma aan het Stanislascollege Westplantsoen Delft. Hierna studeerde hij Maatschappelijk Werk en Dienstverlening aan de Haagse Hogeschool waarvan hij in 2013 afstudeerde. Gedurende deze opleiding was hij als zorghulp werkzaam in het verzorgingstehuis Monica van de Pieter van Foreest. Na het behalen van zijn hbo diploma is hij door het volgen van extra vakken toegelaten tot de pre-master opleiding klinische psychologie aan Universiteit Leiden. Gedurende deze opleiding was hij als groepsleider en persoonlijk begeleider werkzaam bij Domus, een woonvoorziening van het Leger des Heil voor voormalig daklozen met dubbele verslaving en psychiatrische diagnose. Daarnaast werkte hij als psychologisch test-assistent bij Human Company. Zijn afstudeerthesis heeft hij volbracht binnen het LUMC, afdeling psychiatrie. Zijn klinische stage heeft hij volbracht bij GGZ-Delfland specialistische GGZ (SGGZ) poli volwassenen. In 2015 behaalde hij zijn master opleiding Klinische psychologie. In 2016 tot 2019 is hij als basispsycholoog werkzaam geweest bij GGZ-Delfland SGGZ poli volwassenen en het Mobiel Behandelteam. In 2016 startte hij het promotieonderzoek dat resulteerde in het huidige proefschrift. In 2019 tot 2021 heeft hij de opleiding tot gezondheidszorgpsycholoog gevolgd en afgerond binnen de SGGZ poli volwassenen Schiedam, generalistische basis GGZ poli Delft en het psychodiagnostisch centrum van GGZ Delfland. In 2020 is hij begonnen aan de opleiding tot Pro-Justitia rapporteur aan het Nederlands Instituut voor Forensische Psychiatrie en Psychologie. In 2021 is hij als gezondheidszorgpsycholoog werkzaam geweest bij het FACT-team voor psychotische stoornissen Delft en het Delft Eerste Psychose Team van GGZ-Delfland. Sinds 2022 is hij werkzaam als GZ-psycholoog in opleiding tot Klinisch-psycholoog specialist aan het LUMC afdeling psychiatrie.