Data-driven subtypes of major depressive disorder
van Loo, Hanna Maria

DOI:
10.1016/j.jad.2015.03.045

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
CHAPTER 10

General discussion
Preface

This thesis started with the challenge of the heterogeneity of major depressive disorder (MDD): patients with MDD differ considerably in terms of clinical presentation, course of illness, and etiology, which complicates ‘one size fits all’ solutions in research and treatment assignments. Clinical subtypes of depression have not resolved this problem, and this motivated our search for data-driven subtypes of MDD as an alternative, i.e., subtypes informed by similarity patterns in data. The aim of this thesis was to explore the possibility of such empirical subtypes, and to formulate starting-points to identify them (as opposed to finding them, that aim would have been overoptimistic). We investigated empirical, methodological and theoretical issues relevant to the search of data-driven subtypes of depression, by focusing on three questions: what sort of categories are we looking for, what methods are suited to identify them and what do the data show us?

In this discussion, we will put our findings in perspective and discuss their significance and implications for clinical practice, policy decisions and future research. We start by summarizing our main findings (section 1). Then, based on these main findings, we will answer the three questions described above and formulate starting-points for future research to data-driven subtypes of MDD (section 2). Subsequently, we describe the strengths and limitations of our contribution to the literature (section 3). Finally, we indicate the implications of our findings for clinical practice and policy decisions (section 4) and end with some concluding remarks (section 5).

Summary main findings

Part 1: Methodological and empirical findings
The first part of this thesis (chapters 2-6) focused on the empirical and methodological aspects in search of data-driven subtypes of depression.

No conclusive evidence for subtypes with distinctive depressive symptoms based on unsupervised statistical learning methods
In the first study we assessed the available evidence for data-driven subtypes of depression based on depressive symptoms with a systematic review of the literature (chapter 2). Relatively few studies have investigated data-driven subtypes of depression, and all have applied unsupervised statistical learning methods such as latent class analyses and exploratory factor analyses (i.e., 20 studies including 34 analyses). These studies did not provide conclusive evidence for subtypes of depression with qualitatively different depressive symptoms because results diverged considerably. The most consistent finding was a clustering of patients with different degrees of overall symptom severity. We considered two explanations for this inconsistency: either clear symptomatic subtypes of depression do not exist, or such subtypes exist but different methods are needed to discover them. This latter explanation motivated the strategy in our second study.

Supervised statistical learning methods can be applied to search for subtypes predicting long-term course of illness
In the second study we applied supervised statistical learning techniques to find symptomatic subtypes that directly predict long-term course (chapter 3). The study was innovative as it aimed to relate MDD-subtypes directly to multiple clinically relevant course characteristics such as
persistence (years in episode and years in episode most of the year) and severity (disability and hospitalization). Instead of clustering patients based on similar symptom patterns (chapter 2), we clustered patients based on similar course patterns and thereby increased the predictive power of the resulting subtype distinctions. To achieve this, a new combination of statistical learning techniques (recursive partitioning, lasso regression analyses and \( k \)-means cluster analyses) was used in a large general population sample of patients with lifetime MDD from the World Mental Health Survey (WMH). The main result of this study was a proof of concept: we could identify subgroups of MDD-patients with a different course of illness pattern predicted by a selection of psychiatric symptoms during the index depressive episode. Analyses showed three preliminary clusters of patients with a severe, moderate and mild course of illness, mainly differentiated by a combination of anxiety, suicidality and a young age of onset during the index episode (area under curves (AUCs) ranging from 0.61-0.64 for persistence; 0.70-0.72 for severity; training data). Results were preliminary, since retrospective data were used and a limited set of predictors was investigated.

**Subtypes predict long-term course more accurately using comorbidity in addition to psychiatric symptoms as predictors**

To refine the identified subtype distinctions, a follow-up study was performed with an augmented set of predictors including information on lifetime psychiatric comorbidity in the same sample of the WMH Survey (chapter 4). Information on comorbidity indeed resulted in subtype distinctions that were more accurate in distinguishing between patients with a more and less severe course of illness (AUCs ranging from 0.62-0.68 for persistence; 0.70-0.73 for severity; training data). Especially a pre-adult onset of one or more anxiety disorders predicted a severe and persistent course of illness.

**Long-term course is a multifactorial phenomenon**

The fourth study expanded the set of baseline predictors further to improve prediction models for recurrence of MDD in prospective data from a general population sample of female twins with last-year MDD (VATSPSUD data, chapter 5). A wide variety of risk factors predicted recurrence of MDD: specific depressive and anxiety symptoms during the index episode, the level of internalizing symptoms at the moment of interview, psychiatric and family history, personality, early and recent adverse life events, marital status, and problems with friends and finances (AUC estimated in the range of 0.61-0.79; test/training data).

**Interactions improved understanding and prediction of mortality after myocardial infarction**

The fifth study concerned a systematic investigation of interactions of risk factors on prediction of mortality after myocardial infarction in individual patient data meta-analysis (chapter 5). This different patient category enabled us to study the effect of interactions of risk factors on well-established prediction models, which are presently lacking for MDD. Interactions showed to be relevant to improve understanding and prediction of mortality after myocardial infarction: certain risk factors were relevant only in the presence of other risk factors. For instance, depression increased the risk of mortality in men, but not in women. Interactions accounted for these subgroup differences, and therefore, those interactions could be of relevance for prediction models applied to heterogeneous populations such as patients with MDD.
Part 2: Theoretical findings
High rates of comorbidity have raised extensive debate in psychiatry. How should we interpret the high rates of comorbidity? Do comorbidity patterns indicate something about the nature of classification of psychiatric disorders in the DSM? And if so, what? In the second part of this thesis, we tried to answer these questions by studying the debate on comorbidity and its implications regarding the nature of psychiatric disease classification.

Preference for causal disease models underlies a part of comorbidity debate
First, we focused on the discussion whether comorbidity is a problem or a validator for classifications in psychiatry (chapter 7). We found that the two different views on comorbidity leaned on similar assumptions about causality, namely that: i) the high rates of comorbidity indicate that current psychiatric disorders do not correspond with underlying causes and that ii) this is unfortunate because a classification based on causes is preferred to a classification not based on causes, as causal disease models benefit disease understanding and treatment. We nuanced these assumptions by showing that i) high rates of comorbidity also occur in case of causal disease models, and that ii) generally, diseases are defined in a variety of ways and causality is not a necessary condition for a useful classification. A clarification of this debate showed that the limited predictive value of current psychiatric diagnoses is regarded as more problematic than comorbidity patterns.

Rates of comorbidity depend on classification choices and symptom distributions
Second, we investigated rates of comorbidity between MDD and generalized anxiety disorder (GAD) in a large general population sample, in order to clarify the origins of comorbidity: is comorbidity a real phenomenon existing independent of our categorization attempts or is it an artificial by-product of classification choices in the DSM instead? Psychiatric disorders are subject to all sorts of classification choices that possibly influence rates of comorbidity, leading some to conclude that comorbidity is an artificial phenomenon. For instance, MDD is defined in terms of a certain number of necessary symptoms (≥5 of 9 symptoms). The intuition is that a lower ‘threshold’ (e.g., ≥4 of 9 symptoms) results in more patients with MDD and consequently to more comorbidity between MDD and GAD. Data from more than 74,000 subjects from the LifeLines cohort showed that this is not always true: comorbidity rates increased for lower thresholds of MDD, but remained relatively stable for lower thresholds of GAD, owing to specific symptom patterns in the general population (chapter 8). This study illustrated that rates of comorbidity depend on the interplay between classification choices and symptom distributions in the population.

Psychiatric disorders are coordinative definitions serving as bases for empirical knowledge
In chapter 9, we elaborated on this latter point – that measurements in psychiatry are the result of both chosen definitions and reality – and its consequences for the interpretation of psychiatric disorders. By relating the debate on psychiatric comorbidity to other debates in the philosophy of science, we found that definitions of psychiatric disorders are comparable to so-called coordinative definitions. Coordinative definitions are bridge principles that coordinate concepts (e.g., depression) to empirical reality (members of the population with a certain symptom profile). They are necessary conditions to organize empirical facts, and thus serve as bases for measurement and empirical knowledge about psychiatric disorders. Measurement results are
indeed relative to these definitions, yet informative about reality. It means also that definitions like psychiatric disorders are not true or false (they are not empirical facts), but that they can be more or less successful in increasing knowledge about psychiatric diseases, for instance about etiology, treatment, or course of illness.

**Main findings in perspective**

The aim of this thesis was to provide starting-points relevant for the search of data-driven subtypes of MDD, by focusing on three questions: what sort of categories are we looking for, what methods can we use to identify them and what clues do we get from the data? In this section we will try to answer these three overarching questions by combining the main findings, and thereby indicate some starting-points for future research.

**What structures are we looking for?**

*Subtypes of depression are definitions and no empirical facts*

In chapter 9, we compared psychiatric classifications to coordinative definitions, which serve as bases to organize empirical facts and gain knowledge about psychiatry. Subtypes of MDD function in the same way: they are definitions that classify patients with major depressive disorder into subgroups, and provide some organization preceding empirical measurement. It follows that subtypes of MDD are not purely empirical facts and thus cannot be simply proved or falsified based on empirical evidence. Psychiatric disease classifications are ways to carve up the world, but no facts themselves. Classification systems may have truth conditions, but this is not the case for current classifications of psychiatric disorders. There are many different ways to classify psychiatric disorders, and not one is ‘true’. By way of comparison, biologists have different definitions to classify animals in species. Species can be 1) populations of individuals that can interbreed; 2) populations of individuals with common ancestors; or 3) populations of individuals that share phenotypic characteristics [1]. These are competing ways to classify animals into species, and can all serve as bases for increasing knowledge about different types of animals. Similarly, we can classify diseases according to different sorts of principles, such as infectious causes (tuberculosis, human immunodeficiency virus), genetic causes (Huntington, Down’s syndrome), physiological characteristics (diabetes mellitus, hypertension), and symptoms (migraine, tension headache) (chapter 7). Or, we can classify depressive patients according to symptoms, time-of-onset, etiology, or treatment-response [2]. We cannot simply look into the data to determine which of these classifications is true or false: classifications are not fully determined by the observations alone. If subtypes of MDD are no empirical facts, but ways to carve up the world, in what sense can they be data-driven? This has to do with the fact that subtypes can be more and less successful bases for knowledge, which we will discuss next.

*Successful subtypes increase grip on psychiatric disorders*

Some definitions can outperform others as bases for clinical knowledge. For instance, in medicine, we aim to understand, treat or prevent disease processes. Disease definitions such as subtypes should group patients that share important characteristics as a base for this type of generalizations (cf. [3, 4]). This aim drives not only the search for data-driven subtypes of MDD, but many of the other current developments in psychiatric disease classifications, including
clinical staging [5-7], individualized diagnosis based on momentary assessment methods [8], and the Research Domain Criteria [9]. Although classification approaches are different – either applying a staging model on existing DSM-classifications [5-7], or applying individualized diagnoses in addition to DSM-classifications [8], or replacing DSM-classifications with new categories based on neurobiological disturbances [9] – all are meant to provide classifications that improve understanding, prevention and treatment of psychiatric disorders.

Certain classifications might be more successful than others as bases for disease treatment, prevention and understanding. For instance, subtypes of patients with a similar course of illness (chapter 3, 4), similar etiology (chapter 7) or similar treatment reaction could all be more useful than subtypes that only share a certain set of symptoms. Such subtypes can guide decisions about long-term treatment and monitoring, or improve understanding of underlying disease mechanisms and possibilities to interfere. In other words, if we search for subtypes of MDD, we aim to find groups of MDD-patients who share important characteristics in order to maximize understanding, prevention and treatment opportunities. Given that subtypes can be chosen in many ways, but all have the aim of improving grip on psychiatric disorders, empirical studies do have a role. Data can be used in two ways: 1) to track down subgroups of patients with important similarities and 2) to compare the performance of these subgroups with alternative subtype distinctions.

Data-driven subtypes, step I: discover classes with important similarities

Statistical learning techniques might increase the chance of identifying subgroups of patients with relevant similarities, such as similar course of illness patterns. As mentioned in the introduction, statistical methods can be used to discern similarity patterns in data of MDD-patients that were not known beforehand, which could underpin data-driven subtypes of MDD. Thus, different, unexpected patterns and trends might emerge from the data and provide insight in relevant subgroups, which were unidentified by human pattern recognition [10, 11]. Because statistical learning techniques can combine and evaluate larger amounts of patient characteristics than human brains, they increase the chance of finding subgroups with many similarities. These data-driven subtypes might be good starting-points to increase our knowledge about psychiatry and to discover other similarities apart from the definitional characteristics. Note, however, that there is no a priori reason to assume that data-driven subtypes are useful, or that clinician-based subtypes are not. Some patterns that emerge from the data are not stable or not informative (chapter 2). The success of data-driven subtypes depends on all sorts of theoretical choices and data characteristics: what types of variables and statistical methods do we use, which study population, which validators? Therefore, the second step in search of data-driven subtypes includes empirical tests of the actual performance of suggested subtype distinctions.

Data-driven subtypes, step II: test performance of proposed subtypes

Subtypes of MDD are ways to group patients with depression to improve the grip on this disease, and can be chosen more or less successfully with respect to that aim. Some subtype proposals might have more predictive power than others, with regard to etiology, course of illness or treatment. Empirical studies are needed to compare how well alternative subtype distinctions perform, either data-driven or clinician based. For instance, how does the specifier melancholic depression improve treatment or prognosis (DSM)? Does it contribute to clinical decision making in patients with MDD? Would an adjusted version based on data-driven methods (e.g., [12]) define a group of patients more predictive for treatment reaction or prognosis? If so, there
are good reasons to prefer the latter distinction to the first. The latter classification has so-called “comparative validity”: it outperforms competitors with respect to rationally articulated standards of validity [13]. Finally, if we have identified the ‘best’ subtype classification among different options, we should consider whether its implementation would increase clinical decision making enough to outbalance possible drawbacks, such as more complex classification systems. In sum, prior to introducing data-driven subtypes of MDD, a sound empirical evaluation is needed to map the classification’s benefits and drawbacks, and compare these to alternative classifications.

These theoretical conclusions have consequences for the methods used to search for subtypes of depression: some statistical methods are better suited to search for classes with important similarities and to test and compare their predictive value. We will discuss our methodological findings below.

**What methods can be used to identify data-driven subtypes of MDD?**

*Supervised learning methods likely outperform unsupervised learning methods*

As described in the introduction, there are two types of statistical learning methods: supervised and unsupervised. Supervised learning methods such as logistic regression aim to predict an outcome variable (e.g., hospitalization for MDD) based on a set of predictors (e.g. depressive symptoms). Unsupervised learning methods such as latent class analysis aim to describe the association or organization among a set of input measures (e.g., depressive symptoms), but there is no outcome variable [14]. In a systematic review of the literature we showed that, up to now, studies only used unsupervised learning methods in search for data-driven subtypes of MDD (chapter 2). Unfortunately, these studies did not provide conclusive evidence for symptomatic subtypes of depression because results diverged considerably. Therefore, we applied supervised learning methods as a novel strategy in search of MDD subtypes (chapters 3-5). Our results show that supervised learning methods can be used in search of subtypes. Moreover, we suggest that these might be more efficient in guaranteeing clinical value than unsupervised methods.

Given that categories are useful if they have predictive value (chapters 7-9) [4], the utility of subtypes based on unsupervised methods is a priori uncertain. Unsupervised methods ideally result in subgroups of patients with maximal internal consistency: clusters of patients that resemble each other and are considerably different from other subgroups based on certain initial characteristics. For example, imagine a subgroup of depressed patients with melancholic symptoms such as severe loss of interest, weight loss, and insomnia, versus a subgroup with atypical symptoms such as mood reactivity, weight gain and hypersomnia [15]. Classes of patients with similar characteristics at a given moment do not necessarily predict a specific course of illness or a similar etiological background. A certain symptomatic state at a given moment can be caused by multiple factors (multiple causes), and can also have multiple consequences (multiple courses). This is demonstrated by current clinical MDD-subtypes: patients with DSM-melancholic depression might share a specific symptom pattern, but the outcome of this symptomatic state is diverse. Some patients with melancholic depression will have only a single episode over time whereas others develop multiple episodes or chronic depression, and rates are similar to patients with atypical or other types of depression [16]. Of course, the assumption is that subgroups sharing certain characteristics at a given moment will be associated with a certain course of illness or etiology [12], but we can only test this afterwards. Even when stable symptom profiles can be discerned, the predictive value regarding course of illness or treatment reaction is not guaranteed.
The advantage of using supervised methods instead of unsupervised methods in search of MDD subtypes is that eventual subtypes are directly associated with one or more selected outcome variables. With supervised learning methods, we directly involve a clinically relevant outcome as the “learning” variable. This procedure guarantees that, if we are successful in finding accurate prediction models, subgroups based on these models for certain have predictive value. For instance, the subgroups we found in chapters 3 and 4 with supervised learning methods, based on lifetime comorbidity and symptoms of the index episode, have different risk profiles for a persistent and severe course of illness (Figure 1). The rationale of using supervised learning methods is thus to discern subgroups that are maximally predicting a certain clinically relevant outcome. Supervised learning methods might outperform unsupervised learning methods because models are driven by clinically relevant outcomes.

Figure 1. Identified clusters of MDD-patients

Association of identified risk clusters with course of illness after 10-12 years in the US National Comorbidity Survey (NCS; Kessler, Van Loo et al., forthcoming) [17]. The prediction models identified in chapters 3 and 4 were used to assign cluster scores to 1,056 subjects with lifetime MDD that participated in the NCS. This independent dataset made it possible to accurately test the predictive performance of the risk clusters identified in the World Mental Health Survey (WMH). Prospective associations between initial cluster scores (1990-1992) and subsequent persistence and severity of course of MDD (2001-2003) were examined and are presented above. The outcomes measured at follow-up concern percentages of years with MD episodes (persist.), the with MD episodes lasting most of the year (chronic.), hospitalization (hospital.) and suicide attempts since baseline (suic.), and current disability (disabl.), in the NCS data. Area under the curves (AUCs) for the three cluster classification varied between 0.60-0.69 for the outcomes indicating years with (chronic) episodes, and 0.70-0.73 for outcomes indicating severity (hospitalization, suicide attempts, and disability). These estimates of prediction accuracy were comparable to the estimates in the training data from the WMH Survey (AUCs 0.62-0.68 and 0.70-0.71, respectively) [18].
Model selection and model assessment

If our aim is to find data-driven subtype distinctions (chapters 3, 4) or algorithms (chapter 5) that predict some relevant outcome for MDD-patients, this has consequences for selection and assessment of statistical models informing these subtype distinctions. In this case, the problem of finding the right subtypes coincides with the problem of finding the right model. Model selection refers to the selection of the best of a batch of statistical models, and model assessment refers to estimating the actual performance and generalizability of the model [11]. Those two steps boil down to the following questions related to the search of subtypes: which statistical model predicts best in clinical practice (model selection)? How accurate does this model predict exactly (model assessment)?

Model selection: prediction as selection criterion

Prediction depends on the complexity of the statistical model: the model should not be too simple, neither too complex (Figure 2) [11]. Both situations will lead to prediction error: a too simple model does not pick up all the regularities in the data (i.e. high bias, underfitting) (left side, Figure 2), whereas a too complex model follows the data too closely and picks up random noise on top of regularities in the data (i.e. high variance) (right side, Figure 2). Think of all possible penalized regression models out of 80 predictors such as in chapter 5: a too simple model with only an intercept or few predictors will not predict well, but a too complex model including all 80 predictors in the model will follow the training data too closely. The latter phenomenon is called overfitting: the model fits very well to the training data, but performs poorly in a new dataset. Thus, we cannot simply rely on fit statistics in the training data: it might give the impression that more complex models perform better, which is not necessarily true.

Given our interest in subtypes with predictive power, we used prediction accuracy as main model selection criterion in this thesis. We involved many potential predictors and interactions and selected the right degree of model complexity by using penalized regression methods, such as lasso and elastic net regression, in combination with 10-fold cross-validation, to select the model with optimal prediction performance (chapters 3-6). We are unaware of any studies using this approach in research on MDD, but similar model selection strategies have for instance been successfully applied to predict alcohol misuse in adolescents [19].

Model assessment: test performance in new data

After model selection, the actual performance of the model in clinical practice should be estimated. How sensitive and specific is the model? How many false negatives and false positives are to be expected? Does the model improve on current tools for clinical decision-making or not? To estimate this, independent test samples are needed – new data sets that are not used for model fitting and selection – to get a reliable estimate of its performance. Independent test data are necessary to account for overfitting: in that case, the training data will give a too optimistic estimate of model accuracy (Figure 2). Although this way of model assessment is far from standard practice in psychiatry, several new studies have applied it successfully for models predicting MDD in the general population [20], alcohol misuse in adolescents [19], and suicide risk in emergency departments [21].

Also all prediction models found in this thesis have been tested in independent test samples. In two studies, we indeed found higher estimates of prediction accuracy in the training data then in the test data, despite we used techniques to prevent overfitting (penalized regression and cross-validation, chapters 5, 6). This was not the case for the prediction models identified in the
WMH Survey (chapters 3, 4). A recent follow-up study (not included in this thesis) tested the three-cluster solution in an independent sample of 1,056 adults with a lifetime diagnosis of MDD in the National Comorbidity Survey (NCS; Kessler, van Loo et al., forthcoming) [17]. The three risk clusters predicted course of illness after 10-12 years equally accurately in this new dataset and in the original training data, which indicated little overfitting. Differences between area under the curves (AUCs) for outcomes tested in WMH and NCS were minimal (on average 0.01). Some of the results of the three-cluster subtyping scheme in the NCS are shown in Figure 1.

Figure 2. Prediction error in training and test data as a function of model complexity

Behavior of test sample and training sample prediction error with increasing model complexity (in degrees of freedom, df) [11]. The light blue curves show training error, while the light red curves show the conditional test error for 100 training sets of size 50 (n=50) each, as the model complexity is increased. The solid curves show the expected test error and the expected training error. Both bias and variance influence prediction accuracy. Bias refers to approximating a real-life problem with a simplified prediction model. A model with high bias approximates the reality with a too simple model and does not pick up all regularities in the data, and therefore results in prediction error (it underfits the data). Variance refers to the amount to which a model would change if other data were used. A model with high variance means that the model would change considerably when a different dataset would be used to fit it: the model follows the training data too closely and picks up random noise in addition to regularities, and therefore results in prediction error in new data (it overfits the data). Figure reprinted from *The Elements of Statistical Learning* by Hastie et al. [11], with permission from the authors.
What do the data show?

No clear subtypes based on depressive symptoms

First, we did not find evidence for subtypes of depression with qualitatively different depressive symptoms. We focused on correlations among 12 disaggregated symptoms of depression representing DSM-criterion A (chapter 2). Our aim was to investigate if there was any evidence for clusters of depressive patients with distinctive depressive symptom profiles, such as atypical (mood reactivity, hypersomnia, appetite gain) and melancholic profiles (loss of interest, insomnia, appetite loss) [15]. The reviewed studies did not show clear symptomatic subtypes: most latent class analyses identified patient classes based on overall severity instead of distinctive symptom patterns, and exploratory factor analyses resulted in inconsistent factors. Interestingly, the same inconsistency was also observed in schizophrenia, suggesting that the finding is not unique for MDD. A systematic review of factor analytic studies of the Brief Psychiatric Rating Scale in patients with schizophrenia found that none of the 18 reported factor structures fitted well in new data [22].

Second, we did not find evidence that certain combinations of depressive symptoms had a more than additive effect on the risk of a persistent and severe course of illness (chapter 3). We focused on interactions between a broad range of symptoms during the depressive episode (i.e. symptoms of depression, anxiety, mixed episode), and investigated whether these were related to specific course of illness patterns (chapter 3). Although some symptoms predicted a more chronic and persistent course of illness, we did not find symptom combinations that synergistically increased risk, except for combinations of age of onset, suicidality and anxiety. In other words, we did not find specific combinations of symptoms strongly predicting course of illness patterns.

We can think of two main explanations. On the one hand, it could be that clusters of MDD patients with distinctive symptom profiles exist, but we have not applied the right methods to discern them. For instance, it could be that richer and prospective datasets, more advanced statistical techniques such as factor mixture models [10, 22], and accounting for the development of symptoms over time [18, 23, 24] are needed to uncover symptom-based subtypes of MDD. In addition, we related symptom patterns to persistence and severity of course of illness, but not to other clinically relevant outcomes. It could be that distinctive symptom profiles do predict other outcomes such as reaction to antidepressants, psychotherapy or suicide attempts. However, it could also be that substantially large clusters of MDD patients with distinctive profiles have not been identified because they simply do not exist. It could be that depression symptoms are highly correlated and more unidimensional than expected: maybe there is not so much symptomatic variation after all. This explanation fits well with studies reporting good fit indices for a single factor underlying depressive symptoms [25, 26]. It is also in accordance with recent network studies illustrating that depressive symptoms are highly correlated with each other, and the onset of one symptom triggers the onset of other depression symptoms, as a sort of ‘positive manifold’ [27-29].

Anxiety, age of onset and family history are hints for a subtype with severe course of illness

Although we did not find evidence for clusters of MDD-patients with distinctive depressive symptoms, we found certain patient characteristics that predicted a more severe and persistent course of illness (chapters 3-5). Three patient characteristics predicted a more severe course of illness in both general population samples we studied in this thesis, i.e. the WMH Survey
First, anxiety symptoms and comorbid anxiety disorders predicted a more persistent and recurrent course of illness, also when we adjusted for severity measures such as the number of symptoms. Patients at risk for a severe course of illness reported more anxiety symptoms during their depressive episode, such as panic, irritability, and feeling anxious, nervous, or tense. A broad range of anxiety disorders – generalized anxiety disorder, phobias, panic disorder, post-traumatic stress disorder and obsessive-compulsive disorder – predicted a more severe and persistent course of illness in the WMH-sample, especially when these disorders started before the age of 19. The same was true for the VATSPSUD sample: early anxiety (generalized anxiety disorder, panic or phobia before the age of 18 years) predicted faster recurrence of a depressive episode. Our findings on anxiety in these two general population samples agree with studies on anxious depression in clinical samples [30, 31].

Second, a pre-adult-onset of depression was related to unfavorable course of illness in both samples. A first episode before the age of 16 years increased the risk of rapid recurrence of depression in the VATSPSUD sample, and an episode before the age of 19 was associated with more persistence and severity in the WMH-sample. Interaction effects in the WMH-sample showed that a juvenile age of onset was mainly unfavorable if subjects also suffered from anxiety and suicidal thoughts during their depressive episodes. Our results are conform findings in other general population and clinical outpatient samples. In these studies, young age of onset increased risk of recurrence, more lifetime episodes and suicide attempts [32-34].

Third, both the WMH and the VATSPSUD sample showed that a family history of depression or anxiety increased the risk for a more severe course of illness. The WMH-study indicated the risk of a history of depression in one or both parents, and the VATSPSUD-study showed the risk of a history of generalized anxiety disorder in the co-twin. Several other studies have also found that a family history of depression increased the risk of a more persistent course of depression [35, 36], although effects were not always significant [33, 37].

Course of depression is multifactorial and prediction is complex

Next to these three specific predictors, we found indications that course of depression is a multifactorial phenomenon. In the VATSPSUD sample we studied a rich number of baseline predictors, and multiple of these were related to a faster recurrence of MD, including depressive and anxiety symptoms during the index episode, the level of internalizing symptoms at the time of interview, psychiatric and family history, personality, early and recent adverse life events, marital status and problems with friends and finances (chapter 5). This multifactorial finding is in accordance with studies concerning the onset of MD [38, 39], although no previous studies have provided such a multifactorial account of the course of MD [37]. If the course of depression is a multifactorial phenomenon indeed, this would imply that prediction algorithms, and data-driven MDD-subtypes, most likely will improve by employing broad ranges of information to identify them (as opposed to using only symptoms or traumas).

Finally, we found indications that prediction of the course of MDD is a challenging enterprise, even when using a very rich set of baseline predictors and advanced statistical techniques to identify accurate prediction models. The best models predicting course of illness after a depressive episode were less accurate than the best model predicting mortality after myocardial infarction. Estimated AUCs of the models predicting course of depression varied from 0.61 for recurrence to 0.71 for hospitalization, as opposed to an AUC of 0.77 for the model
predicting mortality after myocardial infarction (based on independent test samples from VATSPSUD, NCS, and MINDMAPS, respectively; chapters 5, 6 and [17]). Part of this difference in estimated accuracy could be due to data and methodological limitations: prediction models for depression might improve if we study rich sets of predictors, including biological variables, in bigger training samples, or use more flexible statistical methods [11]. Also different theoretical starting-points, such as analyzing within-subject patterns instead of between-subject patterns, might be worthwhile strategies to improve prediction of course [8, 40].

Then, part of the difference in estimated accuracy could be due to a larger amount of irreducible error – unmeasured variables or unmeasurable variation [14] – connected to the course of MDD. Course of illness after a depressive episode might be a more complex phenomenon than course of illness after myocardial infarction, due to its multifactorial nature involving many time-dependent, and highly individual risk factors (chapter 5). That is, important predictors involved in depression are not fixed (such as sex), but might change over time (such as financial situation, life events, problems with relatives) (chapter 5) [41]. Some risk factors might be difficult to assess because they are highly individual, such as traumas and life events [42]. These factors might challenge the identification of accurate prediction models concerning long-term course of depression, but to what extent remains to be seen.

Starting-points for future studies
In sum, this thesis provides theoretical, methodological and empirical starting-points for future studies in search of data-driven subtypes of MDD. First, as the primary aim of psychiatric disease classification is to improve understanding and treatment of psychiatric disorders, we suggest to connect subtypes directly to clinically relevant outcomes, such as course of illness or treatment reaction. Although subtypes can potentially be chosen in many ways, these clinical outcomes provide objective criteria that can guide the search for and assessment of subtypes that are relevant in terms of patient outcomes. Second, supervised learning methods can be employed to discover classes with important similarities, because models are directly predicting a relevant outcome, such as course of illness. Third, models that inform subtype distinctions should be tested in independent datasets in order to obtain reliable estimates of model performance. Fourth, future studies should involve rich sets of patient characteristics, including interactions, instead of symptoms only to identify subtypes of depression predicting course of MDD. Using data-driven methods, we found indications that course of MDD is a multifactorial phenomenon, involving not only symptoms, but also comorbidity, psychiatric and family history, early and recent life events, and social and economic situation. Interactions of risk factors can contribute to more accurate prediction models and highlight subgroup differences in heterogeneous populations of patients having similar diagnoses. Finally, given the multifactorial nature of MDD, multidisciplinary research will probably have more potential for scientific breakthroughs in discovering data-driven subtypes of MDD, by connecting clinical insights with psychological, environmental and biological factors involved in the origins and course of MDD [43].

This thesis has inspired us to plan future studies. A first extension would be to also analyze other types of predictors that possibly contribute to prediction, such as biological variables, which are available in the Netherlands Study of Anxiety and Depression [44, 45]. A second extension would be to include information on patterns in data over time instead of using one time point only to predict subsequent course of illness. Dynamical information, such as the data provided with momentary assessment methods, might improve the prediction of subsequent outcomes as these dynamical patterns seem to be related to course of illness [23, 24]. Advanced analytical
techniques have been developed to include longitudinal data patterns, also on group level [24]. The value of resulting subtype distinctions could then be compared to clinical staging models [7], of which one has recently been tested in the Netherlands Study of Anxiety and Depression [46]. Finally, we could apply the methods described above to clinical samples, in order to find a subtyping scheme that actually benefits decision-making in clinical practice. For this, we plan to use longitudinal data from psychiatric patients in the North of the Netherlands, who are intensively measured at the moment of intake in various mental health care institutions and will be followed up during the course of treatment. As these data will contain information on both treatment and course of illness in regular treatment settings, we could use both outcomes to inform subtype distinctions, instead of course of illness only.

Strengths and limitations

Limitations
This thesis attempted to formulate starting-points for the identification of data-driven subtypes of MDD, by answering three questions: what structures are we looking for, what methods can we use to identify these, and what patterns do the data show? We have answered part of these questions, but some issues are unresolved, which are the main limitations of this thesis.

Validating criteria
We argue that we should assess the performance of data-driven subtypes of MDD, and compare this to other possible subtype distinctions, but what are the best criteria to do this? Our results are limited as we do not provide an answer to this question. In this thesis, we focused on finding subtypes that predicted a certain course of illness, in order to assist clinicians in decisions involving long-term treatment and monitoring. In addition, we discussed two other potentially useful criteria: reaction to treatment and similar etiology (chapter 7). However, these criteria are only a few of a long list of so-called ‘validators’ – criteria that are supposed to lead to valid classifications (Box 1).

Ideally, a subtype distinction would meet many criteria, and thus define a group of patients with similar symptoms, characteristics and dysfunctions, similar etiology, similar course of illness, similar reaction to treatment, resulting in distress and disability, etc. This, however, will not happen any time soon: it is very unlikely that disorders will measure up to all of these criteria [13]. A subtype based on similar etiology does not necessarily predict a similar reaction to treatment (e.g. perinatal depression) [2], and a subtype based on similar symptoms does not necessarily predict a similar course of illness (e.g., melancholic depression) [16]. To make it even more complex, the current classification system is used for many different purposes (e.g., research, clinical practice, insurance purposes), and different disciplines might have different criteria for validity. For instance, psychiatrists are in general more concerned with external and predictive validity (does the disorder predict course of illness, treatment reaction or family history?), whereas psychologists are interested in psychometric validity (e.g., is the disorder unidimensional?) [47]. Accordingly, a more extensive study of the relative importance of all validators for successful classifications in psychiatry could support the identification of subtypes that outperform others. What are the validators that contribute to better classifications for which purposes?
Box 1. Potential validators for psychiatric disorders [13, 48-50]

- Shared genetic risk factors
- Shared environmental risk factors
- Shared neural substrates
- Shared temperamental antecedents
- Shared abnormalities of cognitive or emotional processing
- Symptom similarity
- Shared clinical features: race, sex, age at onset
- Shared course of illness
- Runs in families
- Results in distress
- Results in disability
- Manifestation of a dysfunction
- Clearly distinguished from normality
- Clearly different from other disorders
- Predicts treatment response

Data limitations
Our empirical findings are limited by sample size (VATSPSUD), the retrospective nature of the data (WMH), and the types of predictors and outcomes we investigated (WMH and VATSPSUD). First, the female samples of the VATSPSUD were relatively small (training data n=194, test data n=133), which might have limited the accuracy of the prediction model (chapter 5). The WMH-sample, on the other hand, was large (n=8,261) but retrospective so that we cannot exclude recall bias (chapters 3-4). Then, although our total set of baseline predictors was very rich in the VATSPSUD study, baseline predictors did not include (neuro)biological findings (structural brain, biomarkers, inflammation parameters) [51], which may also be relevant predictors. Also information on current treatments was lacking, which could have limited the results of our studies. Other missing predictors were information on prior treatment response [52] and work-load and stress [53]. Inclusion of these parameters possibly improves the prediction accuracy of our models.

Limited number of statistical learning methods
We used a limited number of supervised statistical learning methods, mainly penalized linear regression models in combination with classification and regression trees to search for similar subgroups of MDD-patients (chapters 3-6). Many other statistical learning methods are available and should be explored, such as random forests, support vector machines or regression splines, and compared to the methods used here [11, 14]. It could be that more complex and flexible models outperform restrictive methods in predicting course of depression, which is a complex phenomenon itself. The learning method should relate systematically to the specifics of the question and the data, but we do not know beforehand which method will work best [14].

The strength of the restrictive penalized regression methods used in this thesis is that prediction models are interpretable: the model is relatively simple (linear, with a limited
number of predictors), and one can get a grasp of how predictors are related to the outcome. More complex and flexible models might outperform restrictive methods in terms of prediction accuracy (especially when the phenomenon of interest is complex), but it would be very difficult to interpret such a model [14, 54].

Strengths
Notwithstanding these limitations, this thesis adds novel insights to the current literature on data-driven subtypes of depression. Combining insights from different disciplines – psychiatry, statistics and philosophy of science – has been helpful to develop a broader perspective on data-driven subtypes of depression, clarifying what to search for, how to search for it, and resulted in starting-points for future research. First, we have provided a new account of psychiatric comorbidity and psychiatric disease classifications. This account clarifies what types of structures we are searching for, how data can be used to identify classifications, and the value of empirical findings based on these classifications. This position motivates investigation of new disease classifications, without \textit{a priori} discarding current classifications. Second, we have been the first to use relevant outcome variables such as depression recurrence, chronicity, hospitalization and disability, as guiding principles in search for MDD-subtypes. Although previous studies have used clinically relevant variables as validators for data-driven subtypes of depression, subtype distinctions were primarily based on similar symptom patterns instead of on clinically relevant outcomes (chapter 2) [45]. Third, we showed that supervised learning methods can be applied to identify clinically relevant MDD-subtypes, in order to directly increase the predictive power of clusters of patients. Supervised methods might improve the efficiency of identifying clusters of patients predictive for clinically relevant outcomes. Fourth, the definition of subtypes in terms of broad sets of variables predicting clinically relevant outcomes could lead to a move from a pure symptom-based classification to a more pluralistic classification of psychiatric disorders, which is in accordance with the pluralistic nature of disease classifications in medicine in general. In the last two sections of this discussion we will discuss the implications of our findings for clinical practice, society (section 4) and end with some concluding remarks (section 5).

Implications for clinical practice and policy decisions

Clinical practice: prediction of course of MDD
The data-driven search for subtypes of depression predicting recurrence, persistence, disability and hospitalization resulted in observations relevant for estimating course of depression in clinical practice. First, certain risk factors – anxiety symptoms and comorbidity, pre-adult-onset of MDD and anxiety disorders, and a family history of depression and anxiety – pointed towards a more persistent and severe course of illness. Besides these specific risk factors, we found indications for many other risk factors predicting a rapid recurrence of MDD: the (residual) level of symptoms, a psychiatric history with frequent and long-lasting depressive episodes, past and recent adverse life events, and social and financial problems. This finding supports the clinical assessment of all these different risk factors, and base risk estimation on a broad set of risk factors, instead of relying on only a few patient characteristics.

Second, given that prediction of course of MDD over the course of several years is hard and dependent on many time-dependent variables, this would be an argument to update risk estimates over time, as opposed to base risk assessment on one assessment in time. There could
be several ways to monitor changes in time-dependent risk factors. Recovered patients and their relatives could be instructed to contact their general practitioner or psychiatrist in case of adverse life events, other types of serious stress (e.g. financial, social, marital), or early symptoms, in order to update risk estimates for recurrence of MDD. Alternatively, risk estimates could be updated at certain time-intervals, by asking patients to respond to questionnaires, for which software applications might be of assistance. Ultimately, accurate assessment of risk will enhance possibilities for early intervention and prevention of episodes, and at the same time prevents overtreatment of patients with a mild course of illness.

**Politics and society**

Our theoretical findings – what structures are psychiatric disorders? – connect to a broader debate on psychiatric disorders, which has influenced public opinion and policy decisions. Classifications of psychiatric disorders are sometimes criticized because they are in part based on consensus among experts (the ‘bogusat’ method mentioned in the introduction [55]), and may be influenced by pharmaceutical industry, politics and bureaucracy [56, 57]. Therefore, some authors have concluded that psychiatric disorders are no real disorders but man-made inventions, and that psychiatry should not belong to medicine [58, 59]. Others have protested against changes in classification criteria that may have important and unintended consequences in everyday practice [60]. These views have influenced public opinion and policy decisions, and may also have added to the stigma that is still associated with psychiatric disorders [61]. The Dutch minister of health care questioned whether mental disorders are ‘real disorders’ [62]. This took place in the same year that psychiatric patients, as the only patient group in secondary healthcare, were forced to pay an extra charge to receive specialist treatment in order to reduce care consumption in this field [62]. Eventually, framing mental disorders as artificial constructs may have serious societal and political consequences for psychiatric patients regarding employment, treatment reimbursements, and accessibility of health care [63].

We disagree with this argumentation, and thus with societal and political decisions based on such views. We do not deny political or pharmaceutical influences on the DSM: they do unquestionably exist and may have harmful effects on decisions regarding psychiatric classification. However, we deny that these influences imply that psychiatric disorders as defined in the DSM are ‘artificial’ or ‘unreal’. As we argued in the foregoing (chapter 8, 9), it does not make sense to say that classifications are true or false, because they are no empirical facts. Instead, psychiatric categories are classifications, and some classifications are more successful than others. New scientific developments may change the way we classify today based on a more profound insight into the etiology or other aspects of disorders. In this sense, classifications of mental disorders are not different from classifications in medicine in general: some classifications are more useful than others and very few survive unaltered. For example, a committee from the World Health Organization recently tested the utility of the diagnosis metabolic syndrome, and concluded that metabolic syndrome is useful as educational concept, but not as a diagnostic or management tool [64].

Pure constructivist views on psychiatric disorders are therefore not justified and might impair research, policy decisions, and public opinion. The alternative perspective provided in this thesis aims to be more constructive (no pun intended) for purposes of research and science communication. It motivates empirical studies into the merits of each category and into possible improvements of these: what are the strengths and weaknesses of a specific classification? What is known about the course of the disorder, the reaction to treatment, and the consequences
of treatment delay? How large is the proportion of subjects within this category recovering spontaneously? Which patients will benefit more from primary care than from secondary care? How can we recognize different types of patients? Policy decisions, like treatment decisions, should be based on these types of considerations, instead of on broad generalizations.

Concluding remarks

These questions bring us back to the starting point of this thesis: the heterogeneity of depressive patients concerning etiology, course of illness, and treatment reaction. In our eyes, one central question should guide future research: which classifications maximize the possibility to understand and interfere in processes underlying MDD? This thesis showed the potential of data-driven subtypes of MDD as bases for such classifications: data-analysis can detect subgroups of patients directly predicting course of illness profiles or other clinically relevant outcomes. The search for data-driven subtypes of MDD does not finish here. On the contrary, we hope that some of our findings will contribute to a continuing search for better classifications for MDD, and for psychiatric disorders in general.

Presumably, psychiatric disease classifications will not have a fixed form any time soon as they will develop with scientific progress [65, 66]. Empirical findings can lead to refinement or self-correction of initial affirmed systems, as developments in other sciences have illustrated (e.g., the invention of temperature) [66]. Change might be prompted by new data, different cognitive needs, development of new experimentation devices, questionnaires, imaging tools, statistical methods, or encounters with new frameworks [67]. Thus, scientific change is most likely to be the normal state of affairs for psychiatric disease classifications, and for disease classifications in general [65, 68]. Which direction of research will be right or optimal for scientific progress is a priori uncertain, and there might even be different ways to enhance the same epistemic virtue [66]. With its clear focus to enhance particular epistemic virtues – understanding and treatment of MDD – the search for data-driven subtypes of MDD is one of the classification directions worthwhile to further explore, with the final aim to improve the lives of patients with depression.
References


Nederlandstalige samenvatting

Curriculum Vitae

List of publications

Acknowledgements
Nederlandstalige samenvatting

Data-gedreven subtypen van depressie

Depressie is een ziekte die wereldwijd veel lijden veroorzaakt, maar huidige behandelingen slaan niet bij elke patiënt aan en onderzoek naar de oorzaken van de stoornis stagneert. Deels komt dit doordat patiënten met een depressieve stoornis aanzienlijk van elkaar verschillen. Zij variëren in klinisch relevante opzichten, zoals in de aard van de klachten, oorzakelijke factoren en biologische mechanismen. Ook het beloop van de stoornis loopt uiteen: ongeveer 35% van de patiënten heeft één depressieve episode tijdens het leven, terwijl 50% van de patiënten een terugkerende ziekte heeft en 15% lijdt aan een ernstige en langdurige chronische vorm. Dit compliceert behandelbeslissingen (bijvoorbeeld wat betreft het continueren van antidepressiva) en kan leiden tot over- of onderbehandeling.

Deze heterogeniteit heeft geleid tot twijfel aan de waarde van de huidige classificatie van depressie: wellicht is de groep patiënten met deze classificatie te verschillend om algemene onderliggende ziektemechanismen of effectieve behandelingen te vinden. Om meer homogene subgroepen te creëren is een aantal klinische subtypen van depressie gedefinieerd – i.e., subgroepen gebaseerd op waarneming en patroonherkenning in de klinische praktijk – bijvoorbeeld op basis van symptomen (melancholische, atypische, psychotische depressie), vermeende oorzaken (postnatale depressie, aanpassingsstoornis) en moment van ontstaan (seizoensgebonden, late leeftijd). Helaas vormen deze klinische subtypen onvoldoende homogene subgroepen en hebben deze weinig voorspellende waarde voor het beloop van de stoornis of het effect van de behandeling.

In deze dissertatie onderzoeken wij data-gedreven subtypen van depressie als een nieuwe manier om subtypen van depressie te ontdekken. Data-gedreven subtypen zijn groepen van depressieve patiënten die op elkaar lijken en die zijn gebaseerd op patroonherkenning door statistische data-analyses. De afgelopen jaren zijn verschillende statistische methoden ontwikkeld die grote hoeveelheden data van depressieve patiënten kunnen analyseren om subgroepen met dezelfde kenmerken op te sporen. Deze subgroepen kunnen de basis vormen voor empirische subtypen van depressie. Het doel van deze dissertatie is het exploeren van deze nieuwe mogelijkheden en het formuleren van uitgangspunten voor data-gedreven subtypen. Daarbij hebben we ons toegelegd op methodologische, empirische en theoretische aspecten, die relevant zijn voor de zoektocht naar data-gedreven subtypen van depressie. Grofweg beantwoordt deze dissertatie drie vragen: wat zijn geschikte methoden om data-gedreven subtypen van depressie te ontdekken? Wat laten data ons zien aan patronen? En als we zoeken naar subtypen van depressie, naar wat voor structuren zijn we dan op zoek?

Deel 1: Methodologische en empirische bevindingen
Het eerste deel van dit proefschrift richt zich op de eerste twee vragen: wat zijn geschikte methoden om subtypen te vinden en wat zijn de voorlopige resultaten als we deze methoden toepassen? Allereerst hebben we gekeken naar reeds bestaande studies naar data-gedreven subtypen van depressie (hoofdstuk 2). In de literatuur worden regelmatig subtypen van depressie beschreven met andersoortige symptomen. Bijvoorbeeld, men spreekt van een melancholische depressie als patiënten nergens meer plezier in hebben en lijden aan slaaptekort en gewichtsverlies, terwijl patiënten die kunnen genieten van plezierige activiteiten, en juist meer slapen en eten dan normaal worden geclassificeerd als atypisch.
Eerst hebben we in een systematische beschouwing van de literatuur alle reeds bestaande data-analyses op een rij gezet, die het doel hadden subtypen met verschillende typen depressieve symptomen te onderscheiden. In totaal vonden we 20 studies die 34 data-analyses hadden verricht om subtypen van depressie met andersoortige depressieve symptomen te vinden. Al deze studies gebruikten hiervoor latente variabele technieken zoals factoranalyses of latente-klassenanalyses. De resultaten van al deze studies waren echter zo verschillend dat ze geen eenduidig beeld opleverden over het bestaan van subtypen met andersoortige depressieve symptomen. De meest consistente bevinding was een clustering van patiënten op basis van ernst: de data-analyses onderscheiden voornamelijk subgroepen patiënten die op alle depressieve symptomen hoog scoren, of juist laag, maar geen subtypen met verschillende typen depressieve symptomen. Het zou kunnen betekenen dat zulke subtypen niet bestaan, of dat de gebruikte onderzoeksmethoden niet toereikend waren om ze te herkennen: de meeste studies onderzochten kleine groepen patiënten, van wie beperkte gegevens beschikbaar waren, en de gebruikte statistische methoden mogelijk niet geschikt waren.

De uitkomsten van dit review motiveerden ons om nieuwe statistische methoden toe te passen in een grote wereldwijde bevolkingsstudie van psychiatrische ziekten, de World Mental Health Survey (hoofdstukken 3 en 4). We onderzochten in een dataset van 8.261 depressieve patiënten of we data-gedreven subtypen konden herkennen die een gelijk beloop van de ziekte zouden voorspellen. Onze onderzoeksvraag was: kunnen we subtypen herkennen op basis van initiële symptomen en psychiatrische comorbiditeit (i.e., bijkomende psychiatrische stoornissen), zodanig dat deze subtypen voorspellend zijn voor een mild of juist ernstig beloop? Kunnen we subtypen vinden die voorspellen hoe vaak en hoe lang de ziekte terugkomt, of er toekomstige ziekenhuisopnames en beperkingen in functioneren zullen optreden? Daarvoor gebruikten we een combinatie van statistische methoden (classification and regression trees, Lasso regressie en k-means clusteranalyse), die niet eerder waren toegepast in deze context.

Deze studies waren vernieuwend omdat we subtypen direct probeerden te verbinden aan klinisch relevante uitkomsten – ernst en chroniciteit van de stoornis – zodat deze subtypen behulpzaam kunnen zijn voor het maken van behandelbeslissingen in de klinische praktijk. Immers, als depressieve patiënten een subtype hebben met een ernstig en chronisch beloop zullen deze een intensievere behandeling en monitoring nodig hebben dan een subtype patiënten dat slechts eenmalig een depressieve episode ondergaat. De resultaten van deze studies waren bemoedigend: we vonden drie voorlopige subtypen die gekarakteriseerd werden door een ernstig, gematigd en mild beloop van de ziekte, die met name konden worden herkend aan een combinatie van comorbid angstsymptomen en angststoornissen, suicidaliteit en een jonge leeftijd op het moment van het ontstaan van de ziekte.

In twee vervolgstudies hebben we gekeken of we de voorsprielmodellen – de basis van onze data-gedreven subtypen – konden verbeteren door rekening te houden met extra voorspellers en statistische interacties. In hoofdstuk 5 bekeken we of het zinvol is rekening te houden met extra kenmerken naast de eerder onderzochte psychiatrische symptomen en comorbiditeit om accuraat te kunnen voorspellen of patiënten last zullen hebben van een snel terugkerende depressie of niet. We onderzochten hiervoor de gegevens van 194 vrouwelijke patiënten die het afgelopen jaar een depressie hadden doorgemaakt (Virginia Adult Twin Study of Psychiatric and Substance Use Disorders), en van wie we veel informatie hadden over hun voorgeschiedenis, huidige sociale situatie en de meest recente depressieve episode. We vonden dat het beste voorspelmmodel bestond uit een divers en groot aantal risicofactoren betreffende specifieke depressie- en angstsymptomen, de psychiatrische voorgeschiedenis, psychiatrische ziekten in
de familie en de voorgeschiedenis, persoonlijkheidskenmerken, vroege en late trauma's en de problemen in de huidige financiële en sociale situatie. Deze studie suggereert dat data-gedreven subtypen mogelijk preciezer worden in het voorspellen van beloop van de stoornis als rekening wordt gehouden met al deze risicofactoren.

Vervolgens onderzochten we in hoofdstuk 6 of statistische interacties – het rekening houden met speciale effecten van combinaties van risicofactoren – bestaande voorspelmodellen kunnen verbeteren. Daarom onderzochten we een andere populatie van patiënten, namelijk patiënten met een hartinfarct, voor wie al langer algoritmes bestaan die de kans op overlijden in de jaren na een hartinfarct voorspellen. We onderzochten de samengevoegde data van 16 studies (individual patient data meta-analysis, MINDMAPS), met in totaal 10.512 patiënten die in het ziekenhuis opgenomen waren na een hartinfarct. Op een systematische wijze onderzochten we meer dan 500 interacties van risicofactoren en vonden dat een aantal interacties betekenis hadden voor het voorspellen van overlijden na een myocardinfarct: een aantal factoren bleek een ander effect te hebben op het overlijdensrisico voor mannen dan voor vrouwen. Zo verhoogde depressie het risico op overlijden voor mannen, maar niet voor vrouwen. Statistische interacties kunnen rekening houden met dit soort subgroep-verschillen, en kunnen dus ook relevant zijn voor voorspelmodellen voor de heterogene groep patiënten met depressie.

Deel 2: Theoretische bevindingen
Psychiatrische patiënten hebben vaak twee of meer psychiatrische stoornissen tegelijkertijd: er is sprake van veel comorbiditeit. Ook depressie komt vaak voor samen met andere psychiatrische ziekten, bijvoorbeeld met angststoornissen en alcoholverslaving. Deze hoge mate van comorbiditeit is regelmatig onderwerp van discussie in de psychiatrie. Waarom is er zo veel psychiatrische comorbiditeit? Wat zegt deze hoge mate van comorbiditeit over de classificatie van psychiatrische stoornissen in de Diagnostic and Statistical Manual of Mental Disorders (DSM)? Het tweede deel van deze dissertatie onderzoekt comorbiditeit om zodoende een beter beeld te krijgen op wat voor structuren psychiatrische ziekte classificaties zijn. Daarvoor maken we gebruik van data-analyses en inzichten uit de wetenschapsfilosofie.

Zoals gezegd is er een levendig debat gaande in de psychiatrie over comorbiditeit. Een deel van dit debat gaat over de vraag of comorbiditeit een probleem is voor psychiatrische ziekteclassificaties of deze juist valideert. In hoofdstuk 7 onderzochten we welke argumenten gebruikt worden in dit debat. We vonden dat beide posities dezelfde aannamen delen, namelijk dat: i) de hoge mate van comorbiditeit zou aantonen dat huidige psychiatrische ziekten niet samenvallen met onderliggende oorzaken, en dat ii) een causale ziekteclassificatie beter zou zijn dan een classificatie die niet gebaseerd is op oorzaken. We nuanceerden deze aannamen door voorbeelden van ziekten uit de algemene geneeskunde. Deze voorbeelden laten zien dat i) een hoge mate van comorbiditeit ook voorkomt als causale ziekteclassificaties worden gebruikt en ii) dat causaliteit niet een noodzakelijke voorwaarde is voor een bruikbaar ziekteconcept.

Een andere vraag in het debat gaat over de oorsprong van comorbiditeit: is de hoge mate van comorbiditeit het gevolg van classificatiekeuzen in de DSM of is het een reëel aspect van psychiatrische stoornissen en de onderliggende oorzakelijke factoren? In de DSM zijn er allerlei classificatiekeuzen gemaakt, zoals drempelwaarden voor psychische stoornissen. Zo spreekt men pas van een depressieve stoornis als er tenminste 5 van de 9 depressie symptomen aanwezig zijn. Sommige onderzoekers denken dat de mate van psychiatrische comorbiditeit sterk afhankelijk is van dit soort keuzen: een lagere drempelwaarde (bijv. tenminste 4 van de 9 symptomen) zou leiden tot meer patiënten met depressie en daardoor kunstmatig tot meer comorbiditeit. In een
data-analyse van meer dan 74.000 individuen uit de algemene bevolking in Noord-Nederland (LifeLines) onderzochten we of het klopt dat comorbiditeit tussen depressie en gegeneraliseerde angststoornis toeneemt bij lagere drempelwaarden voor deze stoornissen (hoofdstuk 8). Een lagere drempelwaarde voor depressie leidde inderdaad tot meer comorbiditeit, maar verrassenderwijs beïnvloedde een lagere drempelwaarde voor gegeneraliseerde angststoornis de mate van comorbiditeit vrijwel niet. Dit had te maken met het voorkomen van symptomen in de bevolking: alle angstsymptomen bleken vaak al aanwezig te zijn bij mensen die ook last hadden van somberheid of interesseverlies, dus het verlagen van de drempelwaarde voor gegeneraliseerde angststoornis leidde tot weinig 'extra' patiënten.

In een vervolgstudie exploreerden we vervolgens de betekenis van deze bevinding, namelijk dat comorbiditeit enerzijds afhangt van classificatiekeuzen en anderzijds afhangt van het voorkomen van psychiatrische symptomen in de bevolking (hoofdstuk 9). Comorbiditeit is dus geen kunstmatig fenomeen, noch een vaststaand feit onafhankelijk van gekozen indelingen, maar een coproductie van gekozen classificaties en het voorkomen van psychiatrische symptomen in de realiteit. Een soortgelijke conclusie is eerder getrokken in een wetenschapsfilosofische discussie over de objectiviteit van meetresultaten in de natuurkunde betreffende ruimte en tijd. Deze discussie leerde ons dat psychiatrische ziekteclassificaties te vergelijken zijn met zogenaamde coördinatieve definities. Coördinatieve definities zijn noodzakelijke voorwaarden om empirische feiten te kunnen meten en organiseren en zij vormen een basis voor empirische kennis. Meetresultaten in de psychiatrie zijn inderdaad afhankelijk van de gekozen definities, maar tegelijkertijd informatief over de empirische werkelijkheid. Dit soort definities kunnen meer of minder succesvol gekozen worden als basis voor empirische kennis. Het doel is dan ook om de psychiatrische classificaties zo te kiezen dat deze zo veel mogelijk kennis opleveren, bijvoorbeeld over de oorzaken, het beloop, en de behandeling van depressie.

De comorbiditeitsdiscussie verheldert de vraag die centraal zou moeten staan in toekomstig onderzoek naar subtypen van depressie: welke subtypen optimaliseren de mogelijkheden om ziekteprocessen van depressie te doorgronden en te behandelen? Deze dissertatie toont aan dat data-gedreven subtypen van depressie veelbelovend zijn om deze vraag te beantwoorden, en beschrijft een aantal theoretische, methodologische en empirische uitgangspunten voor de zoektocht naar klinisch relevante subtypen van depressie, met als doel de behandeling van depressie te verbeteren.
Curriculum Vitae

Hanna Maria van Loo (1984) was born in Rotterdam, the Netherlands. After completing secondary education (Gymnasium, SG Werenfridus, Hoorn), she moved to Groningen in 2002 to study Medicine and Philosophy. She graduated with degrees in Philosophy (2007, *cum laude*) and Medicine (2010, *cum laude*) from the University of Groningen. During her studies she was involved in teaching (mathematics for high school students, medical education committees), management (as a member of the Faculty Council of the University Medical Center Groningen, the Board of the Faculty of Philosophy, and the University Council) and several international volunteering projects.

In 2010, she started work as a resident in Psychiatry at the University Medical Center Groningen. From 2011 through 2014, she performed her PhD research at the Interdisciplinary Center Psychopathology and Emotion regulation (ICPE, Department of Psychiatry, UMCG) under the supervision of Robert Schoevers, Peter de Jonge, and Jan-Willem Romeijn. During her PhD program, she collaborated with international renowned scientists Ronald Kessler (Department of Health Care Policy, Harvard Medical School) and Kenneth Kendler (Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University). After finishing her PhD she continued her specialization in Psychiatry, combining this with new research projects.

List of publications


Submitted for publication


Research Institute SHARE

This thesis is published within the Research Institute SHARE (Science in Healthy Ageing and healthcaRE) of the University Medical Center Groningen / University of Groningen. Further information regarding the institute and its research can be obtained from our internetsite: http://www.share.umcg.nl/.

More recent theses can be found in the list below. ((co-) supervisors are between brackets)

2015
Mutsaerts M
Lifestyle and reproduction
(prof JA Land, prof BW Mol, dr A Hoek, dr H Groen)

Sulkers E
Psychological adaptation to childhood cancer
(prof R Sanderman, prof PF Roodbol, prof ESJM de Bont, dr J Fleer, dr WJE Tissing)

Febrianna SA
Skin problems related to Indonesian leather & shoe production and the use of footwear in Indonesia
(prof PJ Coenraads, prof H Soebono, dr MLA Schuttelaar)

2014
Schneeberger C
Asymptomatic bacteriuria and urinary tract infections in women: focus on diabetes mellitus and pregnancy
(prof RP Stolk, prof JJHM Erwich, dr SE Geerlings)

Skorvanek, M
Fatigue, apathy and quality of life in patients with Parkinson's disease
(prof JW Groothoff, prof Z Gdovinova, dr JP van Dijk, dr J Rosenberger)

Kolvek G
Etiology and prognosis of chronic kidney disease in children: Roma ethnicity and other risk factors
(prof SAReiljneveld, prof L Podracka, dr JP van Dijk, dr J Rosenberger)

Mikula P
Health related quality of life in people with multiple sclerosis; the role of coping, social participation and self-esteem
(prof JW Groothoff, prof Z Gdovinova, dr JP van Dijk, dr I Nagyova)

Amalia R
Improving a school-based dental programme through a sociodental risk group approach
(prof RMH Schaub, prof JW Groothoff, prof N Widyanti)

Christoffers WA
Hand eczema; interventions and contact allergies
(prof PJ Coenraads, dr MLA Schuttelaar)

Troquete NAC
START-ing risk assessment and shared care planning in out-patient forensic psychiatry; results from a cluster randomized controlled trial
(prof D Wiersma, prof RA Schoevers, dr RHS van den Brink)
Golea E
Functioning of young individuals with upper limb reduction deficiencies
(prof CK van der Sluis, dr RM Bongers, dr HA Reinders-Messelink)

Nguyen HT
Medication safety in Vietnamese hospitals; a focus on medication errors and safety culture
(prof K Taxis, prof FM Haaijer-Ruskamp, prof JRBj Brouwers, dr TD Nguyen)

Lehmann V
Singlehood and partnerships in healthy people and childhood cancer survivors; a focus on satisfaction
(prof M Hagedoorn, prof R Sanderman, dr MA Tuinman)

Jaarsma EA
Sports participation and physical disabilities: taking the hurdle?!
(prof JHB Geertzen, prof PU Dijkstra, dr R Dekker)

Ockenburg SL van
Psychological states and physical fates; studying the role of psychosocial stress in the etiology of cardiovascular disease: a nomothetic versus an idiographic approach
(prof JGM Rosmalen, prof P de Jonge, prof ROB Gans)

Beijers CGHM
Unhealthy behaviors during pregnancy; who continues to smoke and consume alcohol, and is treatment of anxiety and depressive symptoms effective?
(prof J Ormel, prof CLH Bockting, dr H Burger)

Kerdijk W
Strategic choices in curriculum design to facilitate knowledge and competency development
(prof J Cohen-Schotanus, prof JW Snoek, dr R Tio)

Spaans F
Hemopexin activity and extracellular ATP in the pathogenesis of preeclampsia
(prof H van Goor, dr MM Faas, dr WW Bakker)

Brinksma A
Nutritional status in children with cancer
(prof PF Roodbol, prof R Sanderman, prof ESJM de Bont, dr WJE Tissing)

Prihodova L
Psychological and medical determinants of long-term patient outcomes; a specific focus on patients after kidney transplantation and with haemophilia
(prof JW Groothoff, dr JP van Dijk, dr I Rajnicova-Nagyova, dr J Rosenberger)

Snippe E
Understanding change in psychological treatments for depressive symptoms; the individual matters
(prof R Sanderman, prof PMG Emmelkamp, dr MJ Schroevers, dr J Fleer)

Groen B
Complications in diabetic pregnancy; role of immunology and Advanced Glycation End products
(prof TP Links, prof PP van den Berg, dr MM Faas)

For more 2014 and earlier theses visit our website.