Common and specific determinants of 9-year depression and anxiety course-trajectories: a machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA)

Wardenaar, K.J.; Riese, H.; Giltay, E.J.; Eikelenboom, M.; Hemert, A.J. van; Beekman, A.F.; ... ; Schoevers, R.A.

Journal of Affective Disorders 293 (2021) 295–304

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

Common and specific determinants of 9-year depression and anxiety course-trajectories: A machine-learning investigation in the Netherlands Study of Depression and Anxiety (NESDA).

Klaas J. Wardenaar [1,*], Harriëtte Riese [1], Erik J. Giltay [2], Merijn Eikelenboom [3], Albert J. van Hemert [2], Aartjan F. Beekman [3], Brenda W.J.H. Penninx [3], Robert A. Schoevers [1]

[1] University of Groningen, University Medical Center Groningen, Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion regulation (ICPE), Groningen, The Netherlands
[2] Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands
[3] Amsterdam UMC, Vrije Universiteit, Psychiatry, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Anxiety
Depression
Course
Machine Learning
Prediction
SuperLearner

ABSTRACT

Background: Given the strong relationship between depression and anxiety, there is an urge to investigate their shared and specific long-term course determinants. The current study aimed to identify and compare the main determinants of the 9-year trajectories of combined and pure depression and anxiety symptom severity.

Methods: Respondents with a 6-month depression and/or anxiety diagnosis (n=1,701) provided baseline data on 152 sociodemographic, clinical and biological variables. Depression and anxiety symptom severity assessed at baseline, 2-, 4-, 6- and 9-year follow-up, were used to identify data-driven course-trajectory subgroups for general psychological distress, pure depression, and pure anxiety severity scores. For each outcome (class-probability), a Superlearner (SL) algorithm identified an optimally weighted (minimum mean squared error) combination of machine-learning prediction algorithms. For each outcome, the top determinants in the SL were identified by determining variable-importance and correlations between each SL-predicted and observed outcome (ρ_{pred}) were calculated.

Results: Low to high prediction correlations (ρ_{pred}: 0.41-0.91, median=0.73) were found. In the SL, important determinants of psychological distress were age, young age of onset, respiratory rate, participation disability, somatic disease, low income, minor depressive disorder and mastery score. For course of pure depression and anxiety symptom severity, similar determinants were found. Specific determinants of pure depression included several types of healthcare-use, and of pure-anxiety course included somatic arousal and psychological distress.

Limitations: Limited sample size for machine learning.

Conclusions: The determinants of depression- and anxiety-severity course are mostly shared. Domain-specific exceptions are healthcare use for depression and somatic arousal and distress for anxiety-severity course.

1. Introduction

Both depression and anxiety generally follow chronic-intermittent course-trajectories (Verduijn et al., 2017) but considerable course heterogeneity exists (Musliner et al., 2016). Prospective studies of long-term (≥6 years) depression course indicate that patients can have recurrent episodes (Brodaty et al., 2001; Kennedy et al., 2003), never reach full remission (e.g., Angst & Volrath, 1991; Angst, 1996; Keller et al., 1992; Piccinelli & Wilkinson, 1994; Chen et al., 2000) or have persistent residual symptoms (Judd et al., 1998; Kennedy et al., 2004; Rhebergen et al., 2011), all of which is associated with higher costs (McIntyre & O'Donovan, 2004), healthcare use (e.g., Pettit et al., 2009) and impairments (e.g., Judd et al., 2000a;(Fichter et al., 2008). Prospective studies of long-term anxiety course have also shown high persistence over time (Cowley et al., 1996; Bruce et al., 2005; Keller, 2002; 2003; Wittchen & Fehm, 2003; Katschnig & Amering, 1998), although anxiety-disorder diagnoses may change over time (Hovenkamp-Hermelink et al., 2016). Often, depression and anxiety

* Corresponding author: Klaas J. Wardenaar PhD, University Medical Center Groningen, Department of Psychiatry, P.O. Box 30.001 (CC-72), 9700RB Groningen, The Netherlands.
E-mail address: k.j.wardenaar@umcg.nl (K.J. Wardenaar).

https://doi.org/10.1016/j.jad.2021.06.029
Received 15 March 2021; Received in revised form 15 June 2021; Accepted 17 June 2021
Available online 24 June 2021

disorders co-occur (e.g., Clark & Watson, 1991), which is associated with a less favourable course (e.g., Bruce et al., 2005; Rhebergen et al., 2011).

Both for clinical and public-health purposes, identifying the baseline determinants of the long-term course of depression and anxiety is of strong interest. Prospective studies have shown that baseline clinical characteristics such as comorbidity (Coryell et al., 2012), severity, number of previous episodes (Steinert et al., 2014), residual symptoms (Judd et al., 2000b), psychosis (Coryell et al., 1996), suicidality (Moos & Cronkite, 1999), as well as psychological characteristics (Surtees & Wainwright, 1996, Struijs et al., 2018; Hovenkamp-Hermelink et al., 2019) are associated with increased risk of long-term chronicity and/or recurrences. In addition, physical and health-related variables, such as increased triglyceride, decreased HDL cholesterol (Virtanen et al., 2017), lower birth weight, older age at first standing/walking (Colman et al., 2007), and pain intensity, duration and severity (Gerrits et al., 2015) were found to predict long-term depression persistence.

For anxiety, studies have found baseline clinical characteristics, including comorbidity (Hovenkamp-Hermelink et al., 2021; Keller, 2003; Bruce et al., 2005), young age of onset (Angst & Vollrath, 1991; Rubio et al., 2007), anxiety duration (Spinhoven et al., 2016), severity, parental history (Beesdo-Baum et al., 2012), personality traits (Angst & Vollrath, 1991; Schopman et al., in press), mental functioning and negative life events (Schopman et al., in press) to predict long-term persistence. So far, there has been little evidence for biological predictors (Hovenkamp-Hermelink, 2021).

Although informative, previous work has had limitations. First, most studies defined course based on presence or absence of discrete diagnoses over time, whereas mental health is a continuous phenomenon (e.g., Kendell & Jablensky, 2003; Clark & Watson, 1991). The use of course outcomes based on trajectories on continuous severity measurements (e.g., Rhebergen et al., 2011; Olino et al., 2010) could better capture naturally-occurring course variations, but studies using such outcomes have been scarce. Second, most studies focused either on the determinants of depression or anxiety course, whereas these domains are known to be strongly related (Goodwin, 2015; Kotov et al., 2020). Indeed, the few studies that looked at the combined course of depression and anxiety found considerable overlap between their determinants (e.g., comorbidity, neuroticism, childhood adversity; Rhebergen et al., 2011; Fichter et al., 2010; Merikangas et al., 2003). Importantly, some domain-specific course-determinants have also been found. For instance, a parental history of anxiety was found to predict a chronic anxiety course from an early age (<14 years); Olino et al., 2010). This aligns with the theory that both shared and domain-specific mechanisms underlie depression and anxiety (Clark & Watson, 1991). Third, studies have so far each investigated the predictive association of different and relatively small sets of baseline determinants, hampering identification of all potentially relevant (combinations of) course determinants.

Given the above, investigating the shared and domain-specific determinants of depression and anxiety course trajectories using a large number of baseline variables from different domains would be an ideal approach. For such purposes, machine-learning (ML) is ideal as it can handle larger numbers of variables and more complex associations than traditional regression (e.g., Smith, 2018). However, there are many different ML-algorithms that can each yield different results depending on their suitability for a given problem, with the optimal choice seldom being clear a priori. To navigate this problem, a Superlearner (SL) approach (van der Laan et al., 2007; van der Laan & Rose, 2011; Rose, 2013) is recommended, in which multiple ML-algorithms are estimated and 'stacked' in a SL that weights these algorithms as to obtain optimal predictions. This approach was previously shown useful in psychiatric studies (Kessler et al., 2014; Rosellini et al., 2018a; 2018b; 2020; Webb et al., 2020).

The current study aimed to use a SL to address the above-described knowledge gaps with regard to the shared and specific determinants of long-term course of dimensional anxiety and depression, using data from a large 9-year cohort study. This dataset contains many baseline predictor variables (n=152), which in combination with the SL allowed for estimation of prediction models incorporating most of the determinants that had previously not been studied or only in separate studies. Using this approach, as much information as possible was used to obtain optimal predictions for the 9-year course trajectories of (1) general psychological distress, (2) pure depression, and (3) pure anxiety symptom severity. The main baseline course determinants were compared between (a) pure depression and anxiety trajectories to evaluate overlap and (b) between the pure depression/anxiety and psychological distress trajectories to evaluate the degree of overlap between domain-specific and general psychological distress determinants.

2. Methods

2.1. Participants and procedures

Data came from the Netherlands Study of Depression and Anxiety (NESDA) cohort study (Penninx et al., 2008). Participants were recruited from the general population, primary care, and secondary care centres. The baseline sample (n=2,981) consists of adults aged 18-65 years (mean age: 41.6 years; 1,979 (66%) women). Of the participants, 652 (22.9%) had no lifetime diagnosis, and 2,329 (78.1%) had a lifetime diagnosis of a depressive and/or anxiety disorder at baseline. All participants had a detailed assessment at baseline and were followed up after 2, 4, 6 and 9 years. Exclusion criteria at baseline were: not being fluent in Dutch, a primary diagnosis of a psychotic, obsessive-compulsive disorder, bipolar disorder or severe addiction disorder. The detailed objectives and rationales of NESDA are given in Penninx et al. (2008). The Ethical Review Boards of all participating universities approved the study protocol. All participants signed informed consent. This study used data from participants with a diagnosis of dysthymia, major depressive disorder (MDD) or an anxiety disorder within 6 months before baseline (n=1,701; see **Figure S1**).

2.2. Measurements

The baseline assessment consisted of a face-to-face interview including a structured psychiatric interview by a trained research assistant, self-report questionnaires, biological measurements and a blood-draw. In total, 152 baseline variables (**Table S1**) were used as determinants. At 2-, 4-,6- and 9-year follow-up assessments were repeated. Longitudinal anxiety- and depression-severity self-report assessments were used to construct course outcomes.

2.2.1. Determinants

2.2.1.1. Demographics. Age, gender, relationship status, employment status, income, living situation, having children (yes/no), and having siblings (yes/no) were assessed in the interview.

2.2.1.2. Psychiatric and psychological variables. The Composite International Diagnostic Interview (CIDI; World Health Organization [WHO] v2.1) was used to establish DSM-IV diagnoses of depressive (minor depression, dysthymia and MDD) and anxiety disorders (Generalized Anxiety Disorder [GAD], Social Phobia, Agoraphobia, Panic Disorder [PD] with and without Agoraphobia). For all disorders, the lifetime, 6-month and 1-month presence were established. Also, comorbidity, number of comorbid disorders and first age of onset of any disorder were assessed.

Health-related disability and its subdomains ('cognition', 'mobility', 'self-care', 'getting along', 'life activities' and 'participation') were assessed with the WHO Disability Assessment Schedule (WHODAS; Üstün et al., 2010). The symptoms dimensions of General Distress, Anhedonic Depression and Anxious Arousal were measured with the

adapted Mood and Anxiety Symptoms Questionnaire (MASQ-D30; Wardenaar et al., 2010). Symptoms of mania or hypomania were assessed with the Mood Disorder Questionnaire (MDQ, Hirschfeld et al., 2000). Suicidal thoughts and occurrence of a lifetime suicide attempt were assessed with the Beck Scale for Suicide Ideation (BSSI; Beck et al., 1979). Insomnia severity and presence of sleep problems (dichotomized at IRS>9) were assessed with the Insomnia Rating Scale (IRS; Levine et al., 2003). Psychological distress and somatization were assessed with the Four-Dimensional Symptoms Questionnaire (4DSQ, Terluin et al., 2006). Loneliness was assessed with the de Jong-Gierveld Loneliness Scale, using the total, social and emotional loneliness subscales (de Jong Gierveld & Kamphuis, 1985). Big five personality traits were assessed with the Neuroticism-Extraversion-Openness-Five Factor Inventory (NEO-FFI, Costa & McCrae, 1992). Locus of control was assessed with the mastery scale (Pearlin & Schooler, 1978). Daily stressful events were assessed with the daily pressure and rejection subscales of the Daily Hassles questionnaire (Kanner et al., 1981). Threatening life events were assessed using the List of Threatening Events-Questionnaire (LTE-Q; Brugha et al., 1985), using both the occurrence and number of life events. Childhood trauma before the age of 16, was assessed with the NEMESIS-childhood trauma questionnaire (de Graaf et al., 2002) and used to calculate a childhood life event index (0-2) and trauma index (0-4; Hovens et al., 2010).

2.2.1.3. Medication use. Participants brought medication packaging/containers to the interview to be recorded. Medication was classified according to the WHO Anatomical Therapeutic Chemical system. Dichotomous medication variables were used for anxiolytics, antidepressants (selective serotonin reuptake inhibitors, tricyclic antidepressants, other [e.g., serotonin–norepinephrine reuptake inhibitors]), antihypertensive drugs and diabetes medication (see Licht et al., 2008 for details).

2.2.1.4. Lifestyle and somatic health. Current smoking, drugs consumption (cannabis, ecstasy, heroin, cocaine, LSD and speed), and the presence of ≥1 somatic diseases (e.g., diabetes, asthma) were assessed during the interview. The presence of the metabolic syndrome (MS) according to the (Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults 2001) (ATP-III) criteria, was assessed based on biological and medication-use measurements. Abdominal circumference was measured (in cm) with a tape measure and systolic and diastolic blood pressure were measured in supine position using an electronic Omron sphygmomanometer. The MS was considered present if ≥3 of the following criteria were met: (1) abdominal circumference >102 cm (men) or >88 cm (women), (2) triglyceride level ≥1.7 mmol/L, (3) high-density lipoprotein (HDL) cholesterol <1.03 mmol/L (men) or <1.30 mmol/L (women), (4) blood pressure ≥130/85 mm Hg or use of antihypertensive drugs, and (5) fasting plasma glucose ≥6.1 mmol/L or use of diabetes medication. Variables for the MS (yes/no) and for each separate component were used. Alcohol-use problems were assessed using the Alcohol Use Disorders Test (AUDIT; Saunders et al., 1993). Physical activity was assessed using the International Physical Activity Questionnaires (IPAQ; Craig et al., 2003). Pain was assessed using the Chronic Graded Pain Scale, using the intensity and disability subscales (von Korff et al., 1992).

2.2.1.5. Need and use of healthcare. Healthcare use and needs were assessed using the Perceived Need of Care Questionnaire (PNCQ; Meadows et al., 2000). Variables were created for the use of different healthcare types in the 6 months before baseline: homecare, alternative care, self-help group, hospital, physician, medical specialist, occupational physician, psychologist, mental healthcare [MH] institute, substance-use care, independent psychiatrist/psychotherapist, physiotherapist.

2.2.1.6. Cardiovascular measurements. Activity of the cardiac autonomic nervous system during the interview was assessed in sitting position with an ambulatory monitoring system (VU-AMS; de Geus et al., 1995). The VU-AMS continuously registers time series of inter-beat intervals (IBI) and impedance cardiograms (ICG) during rest and test conditions during the interview. Raw data were processed as described elsewhere (de Geus et al., 1995; Riese et al., 2003). IBIs assessed during rest and test conditions were used to derive mean heart rate and heartrate variability (HRV) indices (the Root-mean square of successive differences (RMSSD, in ms) and standard deviation of NN-intervals (SDNN, in ms), Respiration Rate (RR, in breaths per/min) and Respiratory Sinus Arrhythmia (RSA, in ms) were used. ICG signal assessed during test-conditions were used to derive mean Pre-Ejection Period (PEP, in ms).

2.2.2. Blood markers

Before the interview (between 8:00 and 9:00 AM), venous blood was drawn after an overnight fast. Routine assays of the blood samples were run in local labs. Within an hour after the draw the remaining samples were spun down to serum and plasma and stored at -80°C. The routine assays included markers of metabolic function (HDL [mmol/L], LDL [mmol/L], triglycerides [mmol/L], glucose [mmol/L], free thyroxine [pmol/L]), liver function (aspartate aminotransferase [ASAT-GOT; U/L], γ-glutamyl transferase [Gamma-GT, U/L], alanine aminotransferase [ALAT-GPT; U/L]), kidney function (creatinine; mg/dL) and haematological markers (haematocrit [L/L], haemoglobin [mmol/L] and erythrocyte count [$*10^{12}$/L]). Plasma Interleukin-6 (IL-6; mg/L) levels were measured in duplicate using a high sensitivity enzyme linked immunosorbent assay (ELISA) (PeliKine Compact TM ELISA, Sanquin, Amsterdam, The Netherlands). Tumor necrosis factor-alpha (TNF-α; pg/mL) plasma levels were assayed in duplicate with a high-sensitivity solid phase ELISA (Quantikine® HS Human TNF-α Immunoassay, R&D systems Inc, Minneapolis, MN, United States). C-reactive protein (CRP; mg/L) was measured in duplicate by an in-house ELISA, based on purified protein and polyclonal anti-hsCRP antibodies (Dako, Glostrup, Denmark). Tryptophan (μmol/L) and kynurenine (μmol/L) concentrations were assayed by an automated online solid phase extraction-liquid chromatographic-tandem mass spectrometric (XLC-MS/MS) method. Brain-derived neurotrophic factor (BDNF; ng/L) levels were measured in serum using the Emax Immuno Assay system from Promega (Madison, WI, USA).

2.2.3. Course outcome measures

Course outcomes were based on the longitudinal data collected with the Inventory of Depressive Symptomatology-Self Report (IDS-SR; Rush et al., 1996), the Beck Anxiety Inventory (BAI, Beck et al., 1988), the Fear Questionnaire (Marks & Matthews, 1979) and the Penn State Worry Questionnaire (Meyer et al., 1990). General psychological distress severity was calculated as the average of the standardized IDS-SR, BAI, FQ and PSWQ scores (**Table S2** gives scale correlations).

2.3. Analyses

2.3.1. Outcome trajectory classes

Outcomes for the current study were all based on LCGAs of the longitudinal course measures. In a previous study (Solis et al., submitted), LCGA had been run on the general psychological distress measure. For the current study, LCGAs were also run with (1) only the IDS-SR and (2) the BAI scores to identify pure-depression and pure-anxiety course trajectories, respectively.

Prior to analysis, all repeated outcome scores were standardized to zero mean and unit variance and baseline values were subtracted from the follow-up values to focus the LCGAs on change rather than baseline differences. LCGA was used rather than growth mixture modelling as the latter allows for more within-class heterogeneity, making the classes less suitable as differentiated outcome categories. Inspection of the pre-

processed data revealed that most change occurred between baseline (all values are 0) and the first follow-up, after which changes were more stable. To model this pattern of change, a log-linear function was chosen for the LCGA (i.e. outcomes were log-transformed).

LCGA models with increasing numbers of classes were fitted to the data. The best model was selected using the Bayesian and Akaike Information Criteria (BIC and AIC), with lower values indicating better fit. Also all model classes were required to have a size of n>100 to ensure that they would be usable in the subsequent prediction analyses. All models were estimated using Maximum Likelihood estimation, using each participant's non-missing data. One-hundred random starts were used to prevent solutions at local maxima. LCGAs were conducted with package *'lcmm'* (Proust-Lima et al., 2017) in R (R core team, 2020). In subsequent analyses, the continuous posterior class-probabilities rather than discrete class membership (based on highest posterior class-probability) were used if class-allocation involved too much uncertainty as indicated by model entropy<0.8 (Clark & Muthén, 2009).

2.3.2. Prediction modelling

2.3.2.7. Data pre-processing. All categorical variables were recoded to dichotomous variables, coded as 0/1. Continuous variables were Z-transformed. Overall, only 1.7% of the data were missing: 66/152 determinants had ≥1 missing value (range: 5.9%-15.8%; **Table S1**). These were imputed once using predictive mean matching with the '*mice*' R-package (van Buuren and Groothuis-Oudshoorn, 2011). See supplement for some complete-case analyses.

2.3.2.8. Super Learner. SL analyses were performed with the '*sl3*' R-package (Coyle et al., 2020) to identify the optimal prediction model for each outcome-trajectory. The mean squared error (MSE) of predicted versus observed outcomes was used to determine prediction accuracy. In the SL method, a number of individual ML-algorithms (*'base learners'*; leDell et al., 2016) are run. Base-learners were selected that can work with continuous outcomes and large numbers of correlated determinants: Elasticnet (Zou & Hastie, 2005), Random Forests (Breiman 2001), Gradient Boosting (Friedman, 2001), and Support Vector Machines (details: **Table S3**). Each base-learner was run with 10-fold cross validation (CV) to limit the risk of overfitting. We chose 10-fold CV in favour of a single hold-out sample as the former is more efficient (all subjects are included in training and validation) and yields MSEs with less random variation (Hastie et al., 2009; James et al., 2013). Next, a stacked SL-model formula, weighting each of the individual learners' predictions as to minimize the SL MSE (MSE_{SL}) was estimated in the whole sample using a Non-Negative Least Squares (NNLS, van der Laan et al., 2007) estimator, which was set to yield a convex solution with the learners' weights summing to one. This yields SL predictions that improve upon any of the stack's base-learners (van der Laan et al., 2007). The resulting SL formula was then used to generate predictions based on the base-learners that were fit to the complete dataset, and the MSE_{SL} was evaluated. See Rose (2013) for details. Although the MSE_{SL} is a good measure of precision, is hard to interpret on its own. Therefore, the Spearman correlation between SL-predicted and observed scores (ρ_{pred}) and its squared value (ρ^2_{pred}: coefficient of determination) were also calculated and evaluated using the following performance cut-offs: $\rho_{pred}<0.2$ (negligible: less than practically significant), $0.2\geq\rho_{pred}<0.5$ (small), $0.5\geq\rho_{pred}<0.8$ (moderate) and $\rho_{pred}\geq0.8$ (strong; Ferguson, 2009). Finally, SL performance was compared to performance of Ordinary Least Squares (OLS) regression with the identified top 15 important determinants (see below).

2.3.3. Determinant importance

For each outcome, the importance of each determinant in the SL was investigated by evaluating the change in the MSE_{SL} (i.e. MSE difference) when the given determinant's data were randomly scrambled (Coyle et al., 2020). For each outcome's SL, this resulted in a list of determinants' MSE-differences, with higher values indicating higher importance. For each outcome, the top 15 important determinants were listed and Spearman correlations between each determinant and the model-predicted score were calculated to gain insight into the determinants' roles in the SL.

3. Results

3.1. Sample characteristics

The baseline sample characteristics are given in **Table 1**. The mean age was 41.3 years and the majority was female (67.2%). Most participants had an MDD diagnosis (65.4%) and a majority of these also had comorbid anxiety (68.0%). Of participants, 23.3% and 32.0% had only a depressive or anxiety disorder, respectively.

3.2. LCGA

3.2.1. General psychological distress

LCGA identified three classes with different severity trajectories: a 'chronic' (n=1,131), 'partial recovery' (n=435) and 'full recovery' trajectory (n=127) (Solis et al., submitted).

3.2.2. Pure depression and anxiety

The preprocessed IDS-SR and BAI data showed a moderate relationship (Spearman ρ=0.64; p<0.001). In the LCGAs (**Table 2**), for both domains, the BIC decreased little when adding a 4th class and for both outcomes, the 4-class model contained classes that were too small (n<100). For depression, the selected model had a 'chronic' (n=1,078), 'partial recovery' (n=502) and 'full recovery' (n=112) class. For anxiety, the model had a 'full recovery' (n=236), 'partial recovery' (n=1,306), and 'increasing severity' (n=151) class (**Figure S2** gives the trajectories). Because entropy was <0.8 for both models, the posterior class-probabilities were used as outcomes (See **Table S4** for distributions).

Table 1
Study sample characteristics (n=1,693)

Variable	*Sample*
Demographics	
Age (in years), mean (sd)	41.3 (12.4)
Women, n (%)	1,137 (67.2%)
Years of education	11.8 (3.3)
Unemployed, n (%)	597 (35.3%)
Partner, n (%)	1,107 (65.4%)
6-month CIDI diagnoses, n (%)	
MDD	1,108 (65.4%)
Dysthymia	303 (17.9%)
GAD	460 (27.2%)
PD with Agoraphobia	421 (24.9%)
PD without Agoraphobia	245 (14.5%)
Social phobia	665 (39.3%)
Agoraphobia	187 (11.0%)
Pure depressive disorder(s), n (%)	394 (23.3%)
Only anxiety disorder(s), n (%)	542 (32.0%)
Comorbid depression and anxiety, n (%)	757 (44.7%)
Ever a suicide attempt, n (%)	283 (16.7%)
Severity scales, median (IQR)	
Depression severity (IDS-SR)	29 (20-38)
Anxiety Severity (BAI)	15 (9-24)
Worrying (PSWQ)	37 (30-44)
Fear/Phobia (FQ)	30 (17-46)

Note: CIDI= Composite International Diagnostic Interview; MDD=Major Depressive Disorder, GAD=Generalized Anxiety Disorder, PD=Panic Disorder, IDS-SR=Inventory of Depressive Symptomatology - Self Report; BAI=Beck Anxiety Inventory; PSWQ=Penn State Worry Questionnaire; FQ=Fear Questionnaire; IQR=interquartile range.

Table 2
latent class growth analysis (LCGA) results to explain variation in the 9-year course of (A) depression severity and (B) anxiety severity over 9 years.

Outcome	Classes	Degrees of freedom	AIC	BIC	Entropy	Class sizes
Depression (n=1,692)[a]	1	3	15241.3	15257.6	1.00	1692
	2	6	13821.9	13854.5	0.69	1357, 335
	3	**9**	**13378.5**	**13427.4**	**0.65**	**502, 112, 1,078**
	4	12	13214.2	13279.4	0.67	70, 102, 1169, 351
Anxiety (n=1,693)[b]	1	3	15402.0	15418.3	1.00	1693
	2	6	13971.3	14003.9	0.68	340, 1353
	3	**9**	**13507.2**	**13556.1**	**0.70**	**151, 1306, 236**
	4	12	13161.8	13227.0	0.70	80, 361, 96, 1156

Note: IDS-SR=Inventory of Depressive Symptomatology - Self Report; BAI=Beck Anxiety Inventory; PSWQ=Penn State Worry Questionnaire; FQ=Fear Questionnaire; AIC=Akaike Information Criterion; BIC=Bayesian Information Criterion. Bold print indicates the model that was selected in this study.
[a] Based on the IDS-SR.
[b] Based on the BAI.

3.3. Prediction modelling

3.3.1. Prediction of general psychological distress course

Prediction performance indices are given in Table 3 (all ρ_{pred}: $p<0.001$). The 'chronic' class probability was predicted with MSE=0.150-0.159 across base learners and MSE_{SL}=0.148, with a strong ρ_{pred}=0.91 (ρ^2_{pred}=0.83). 'Partial-recovery' was predicted with MSE=0.118-0.131 across base learners and MSE_{SL}=0.118 with a small ρ_{pred} of 0.41 (ρ^2_{pred}=0.17). 'Full-recovery' was predicted with MSE=0.054-0.056 across base learners and MSE_{SL}=0.054 with a moderate ρ_{pred} of 0.70 (ρ^2_{pred}=0.50). SL-based ρ_{pred} were all higher than small OLS-based ρ_{pred} (0.18-0.27).

Figure 1 shows the top determinants. Age, age of onset, RR, WHODAS participation, somatic disease, low income, 1-M minor depression and mastery score were among the most important overall determinants. Prediction correlations showed that age and somatic disease were positively correlated with predicted 'chronic' course and negatively with 'partial' and/or 'full recovery', whereas age of onset, RR, WHODAS participation, mastery and minor depression were positively correlated with 'partial' and/or 'full recovery'.

3.3.2. Prediction of pure depression course

'Chronic' course probability was predicted with MSEs=0.133-0.137 across base learners (Table 3) and MSE_{SL}=0.131, with a strong ρ_{pred}=0.81 (ρ^2_{pred}=0.66). 'Partial-recovery' was predicted with MSE=0.114-0.121 across base learners and MSE_{SL}=0.114 with a strong

Table 3
10-fold cross-validated prediction performance of individual learners and of the total sample Super Learner in predicting the probability of different course trajectories of joint depression and anxiety.

	Learners	Probability of 'Chronic course'			Probability of 'Partial recovery'			Probability of 'Full recovery'		
		MSE[a]	SE	SL Weight	MSE[a]	SE	SL Weight	MSE[a]	SE	SL Weight
General psycho-pathology severity course[c]	Elasticnet	0.152	0.004	0.18	0.118	0.003	0.68	0.054	0.004	0.19
	Random forest (100)[b]	0.151	0.003	0.11	0.120	0.003	0.00	0.056	0.004	0.00
	Random forest (250)[b]	0.150	0.003	0.01	0.120	0.003	0.00	0.055	0.004	0.08
	Random forest (500)[b]	0.150	0.003	0.30	0.119	0.003	0.00	0.055	0.004	0.00
	Gradient Boosting	0.151	0.003	0.09	0.119	0.003	0.32	0.055	0.004	0.00
	Support Vector Machine	0.159	0.004	0.31	0.131	0.005	0.00	0.055	0.005	0.73
	SuperLearner	0.148	0.003	-	0.118	0.003	-	0.054	0.005	-
		Probability of 'Chronic course'			Probability of 'Partial recovery'			Probability of 'Full recovery'		
		MSE[a]	SE	SL Weight	MSE[a]	SE	SL Weight	MSE[a]	SE	SL Weight
Pure depressive symptom severity course[d]	Elasticnet	0.133	0.003	0.37	0.114	0.003	0.56	0.049	0.004	0.21
	Random forest (100)[b]	0.136	0.003	0.10	0.117	0.003	0.00	0.050	0.004	0.08
	Random forest (250)[b]	0.137	0.003	0.00	0.116	0.003	0.03	0.050	0.004	0.00
	Random forest (500)[b]	0.137	0.003	0.00	0.116	0.003	0.10	0.050	0.004	0.00
	Gradient Boosting	0.134	0.003	0.16	0.116	0.003	0.07	0.050	0.004	0.00
	Support Vector Machine	0.137	0.004	0.37	0.121	0.004	0.24	0.050	0.005	0.71
	SuperLearner	0.131	0.003	-	0.114	0.003	-	0.049	0.004	-
		Probability of 'Increasing severity'			Probability of 'Partial recovery'			Probability of 'Full recovery'		
		MSE[a]	SE	SL Weight	MSE[a]	SE	SL Weight	MSE[a]	SE	SL Weight
Anxiety severity course[e]	Elasticnet	0.058	0.004	0.37	0.117	0.004	0.63	0.096	0.005	0.11
	Random forest (100)[b]	0.059	0.004	0.00	0.118	0.004	0.16	0.097	0.004	0.00
	Random forest (250)[b]	0.058	0.004	0.04	0.118	0.004	0.00	0.096	0.004	0.00
	Random forest (500)[b]	0.058	0.004	0.00	0.117	0.004	0.20	0.096	0.004	0.29
	Gradient Boosting	0.058	0.004	0.32	0.119	0.004	0.00	0.095	0.005	0.43
	Support Vector Machine	0.062	0.005	0.26	0.130	0.005	0.00	0.102	0.006	0.17
	SuperLearner	0.057	0.004	-	0.116	0.004	-	0.095	0.005	-

Note: IDS-SR=Inventory of Depressive Symptomatology - Self Report; BAI=Beck Anxiety Inventory; PSWQ=Penn State Worry Questionnaire; MSE=mean squared errors; SL weight= NNLS estimated weight (range: 0-1) of the algorithm in the Super Learner.
[a] The average of the MSEs obtained in each test fold;
[b] Numbers between parentheses indicate number of trees in the *ranger* package.
[c] Based on the IDS-SR, BAI, PSWQ and FQ
[d] Based on the IDS-SR
[e] Based on the BAI

Figure 1. Variable importance for the course of general psychopathology severity, expressed as the percentage of change in the mean squared error if the given determinant is removed from the model ('MSE difference (%)' on x-axis). For added interpretability, the Spearman correlations of the determinants with the predicted probabilities of the Super Learner (SL) are printed behind each determinant's bar. A green bar indicates a positive correlation and a red bar indicates a negative correlation with the SL-predicted score.

ρ_{pred}=0.82 (ρ^2_{pred}=0.67). 'Full-recovery' was predicted with MSE=0.049-0.050 across learners and MSE_{SL}=0.049 with a moderate ρ_{pred}=0.60 (ρ^2_{pred}=0.36). SL-based ρ_{pred} were larger than the OLS-based ρ_{pred} (0.22-0.31).

Figure 2 shows the top-15 determinants, which were age, youngest age of onset, mastery, visiting an occupational physician, physiotherapist or MH institute, somatic disease, extraversion and WHODAS participation score. Age was positively correlated with the probability of a 'chronic' course and negatively with 'partial' and 'full recovery', whereas age of onset, mastery and visiting a MH institute showed the reversed pattern.

3.3.3. *Prediction of pure anxiety course*

An 'increasing-severity' probability was predicted with MSE=0.058-0.062 across learners (Table 3) and MSE_{SL}=0.057 with a moderate ρ_{pred}=0.53 (ρ^2_{pred}=0.28). 'Partial-recovery' was predicted with MSE=0.117-0.130 across learners and MSE_{SL}=0.116, with a high ρ_{pred}=0.88; ρ^2_{pred}=0.77). 'Full-recovery' was predicted with MSE=0.095-0.102 across learners and MSE_{SL}=0.095 with a moderate ρ_{pred}=0.73; ρ^2_{pred}=0.53). The SL-based ρ_{pred} were much higher than negligible OLS-based ρ_{pred} (0.12-0.20).

Figure 2 shows that MASQ-D30 Anxious Arousal, 4DSQ Somatic, 4DSQ Distress, WHODAS participation, 1-M minor depression, extraversion, age and RR were among the top determinants. The MASQ-D30 Anxious Arousal and 4DSQ Somatic scales were negatively correlated with 'increasing severity' and positively with 'partial recovery'. Minor depression, extraversion and RR were positively correlated and age was negatively correlated with 'full-recovery'. 4DSQ Distress and WHODAS Participation were positively correlated with 'partial-recovery'.

4. Discussion

This study aimed to train models that optimally predict the 9-year course of general psychological distress, pure depression and pure anxiety, and to identify and compare the main determinants of these outcomes. The used 9-year course outcomes for general psychological distress were previously-estimated LCGA-based classes with 'chronic', 'partial recovery' and 'full recovery' course trajectories. For pure depression, LCGAs showed similar classes. For pure anxiety, LCGA showed 'partial recovery', 'full recovery' and 'increasing severity' classes, aligning with previous observations that depression and anxiety-trajectories do not always run parallel (Olino et al., 2010; Wardenaar et al., 2015). For each outcome, optimized SLs were estimated with base-learner weights differing across outcomes, indicating that different learners were optimal for different outcomes and using a single learner would have been suboptimal. Prediction-correlations ranged from small (ρ_{pred}=0.41) to high (ρ_{pred}=0.91) with a moderate median of 0.73 and consistently better performance than OLS regression. Interestingly, the only outcome class, for which prediction performance was weak for the SL was the general psychological distress 'partial recovery' class. This could be due to this outcome's high within-class heterogeneity (many different symptoms can partially recover), making it a comparatively noisy outcome.

Inspection of variable importance in the SL showed that for the general psychological distress course, a mix of different types of important determinants emerged, age, low income, youngest age of onset, WHODAS participation score, 1-M minor depression, mastery, RR and somatic disease variables. For pure depression course, mostly the same determinants were found, but also unique determinants (visit occupational physician, physiotherapist, or MH institute). For pure anxiety, some top determinants overlapped with pure depression (age, WHODAS participation, minor depression and extraversion), and unique determinants included self-reported somatic symptoms and psychological distress. The identification of different types of determinants aligns with the view that psychopathology is influenced by factors, functioning at different levels (e.g., McNamara et al., 2021; Eronen, 2019). The observed overlap in determinants between outcome domains and lower number of domain-specific determinants aligns with the view of depression and anxiety as related domains with shared and

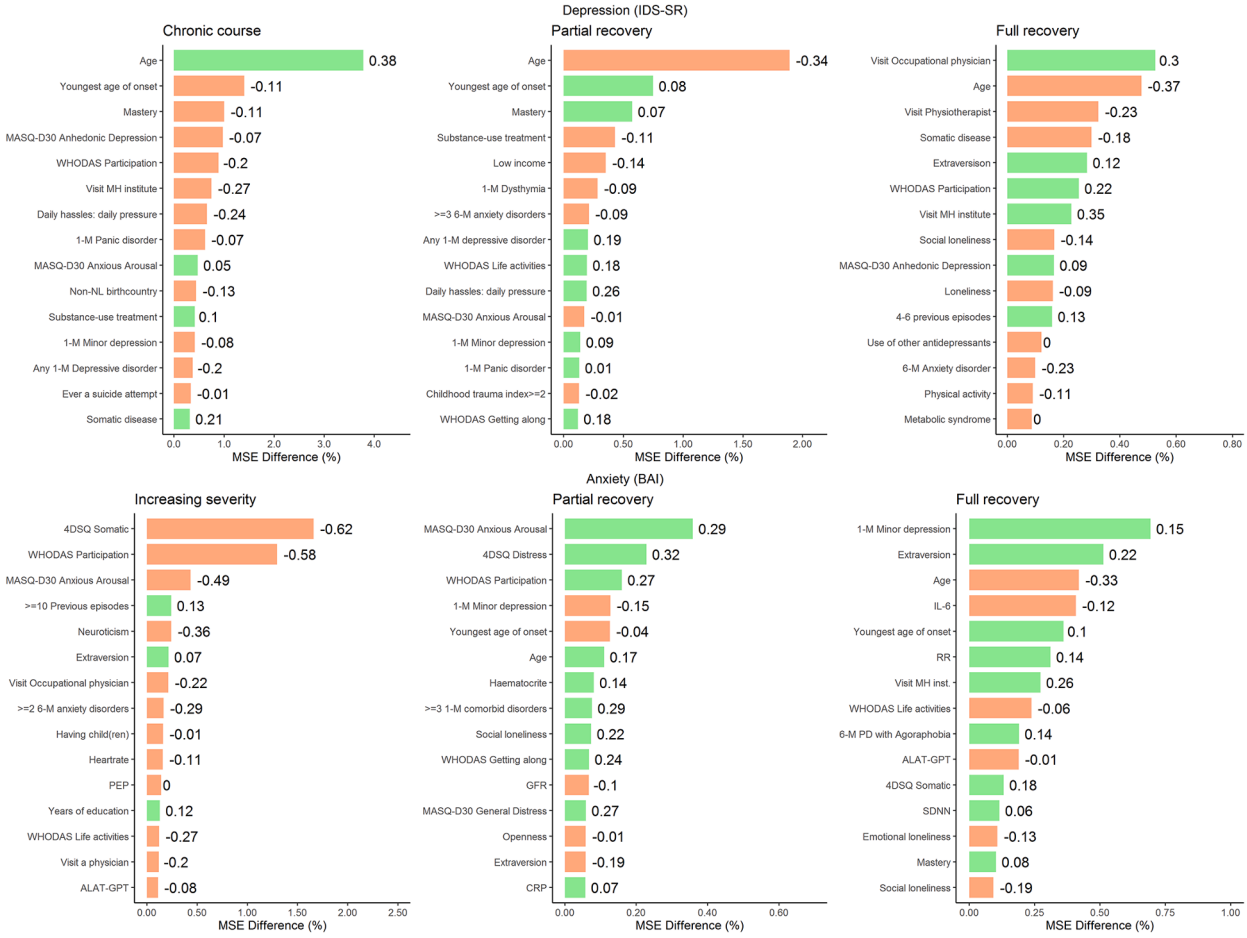

Figure 2. Variable importance for the course of individual depression (IDS-SR) and anxiety (BAI) severity, expressed as the percentage of change in the mean squared error if the given determinant is removed from the model ('MSE difference (%)' on x-axis). The Spearman correlations of the determinants with the predicted probabilities of the Super Learner (SL) are printed for each determinant. A green bar indicates a positive correlation and red indicates a negative correlation with the SL-predicted score.

domain-specific mechanisms and with the shared predictors overlapping with a general, higher-order psychological distress dimension (Clark & Watson, 1991; Goodwin, 2015; Kotov et al., 2020).

The roles of determinants in the algorithm were investigated by inspecting determinants' correlations with SL predicted probabilities. These should be interpreted carefully because they only capture linear relationships whereas variable importance also incorporates variables' roles in interactions and non-linear associations (Archer & Kimes, 2008; Hastie et al., 2009). Still, the correlations revealed some noteworthy patterns. Across outcomes, age was positively correlated with the chronicity and negatively with partial/full recovery. Indeed, previous work has found higher chronicity/recurrence with increasing age (e.g., Schaakxs et al., 2018). This may reflect that depressed patients that were older at baseline were more likely to already have had a history of persistent depression, making future persistence likely. Age of first onset was negatively correlated with an adverse course of psychological distress and pure depression, but not anxiety, indicating that it may be a depression-specific determinant. Others also showed that a young onset of depression is an important component of depression chronicity (Pettit et al., 2009) and related to higher illness burden (Zisook et al., 2007). Measures of extraversion and mastery were found be positively correlated with recovery and negatively with adverse course trajectories across outcomes, aligning with earlier work showing these domains to be positively associated with lower chronicity/recurrence (Steunenberg et al., 2007; Colman et al., 2007; Wardenaar et al., 2014; Wiersma et al., 2011; Hovenkamp-Hermelink et al., 2021), pointing toward a role as cross-diagnostic vulnerability markers (Ormel et al., 2013; Struijs et al., 2018) that likely predict over longer time periods given their temporal stability (Hovenkamp_Hermelink et al., 2019; (Mund et al., 2020). Somatic illness correlated positively with an unfavourable course, aligning with work showing poor somatic health to be associated with less favourable depression and anxiety course (Ferro et al., 2015; Ambresin et al., 2014). RR was the only biological determinant for both general psychological distress and pure anxiety, indicating a more anxiety-specific role. Indeed, respiratory deviations, such as a higher RR, are known to be related to anxiety symptom severity and worry (Papp et al., 1997; Blechert et al., 2007). However, here RR was positively correlated with more favourable course trajectories. It is possible that this reflects the role of RR in a more complex interactive/non-linear model, or it could reflect the fact that anxiety severity and related somatic arousal (including higher RR) were on average higher (data not shown) in those with a 'chronic' or 'increasing severity' course. The finding of a single biological determinant indicates that in algorithms for clinical use it could be preferable to predominantly include other measures that are easier and less invasive to measure.

Interestingly, previous ML prediction studies of the 2-year course of depression (Dinga et al., 2018) and anxiety (Bokma et al., in press) using the same data found baseline symptom severity to be the most important predictor. Another ML-study of 9-year diagnosis-based course prediction (van Eeden et al., 2021) used an automated algorithm-selection approach for prediction (auto-sklearn; Feurer et al., 2015), but found only moderate prediction accuracy. Differences between these and the

current findings could be explained by different methodological choices. First, the first two studies used only a 2-year follow-up interval. Second, the current study looked at dimensional rather than discrete diagnostic course outcomes. This meant a conceptually different approach, but also that baseline severity was used in outcome definition and could not be used as determinants. Interestingly though, other severity scales (e.g., MASQ-D30 General Distress) did not emerge as important either. Dinga et al., (2018) also used LCGA-based outcomes, but these were for 2-year course only. Third, two studies used a single learner (Elasticnet [Dinga et al., 2018]; Random Forests [(Bokma et al., in press)]), so although many baseline variables were included (up to 567 in (Bokma et al., in press)), the chosen learner may not have been suitable to identify more predictors. Vice versa, (van Eeden et al., 2021) did use a more flexible algorithm, but a relatively small set of 35 baseline predictors.

There are several study limitations. First, the findings were based on data-driven dimensional course outcomes, which do justice to the way course-variations naturally occur, but are harder to translate to DSM-oriented clinical practice. Second, the SL analyses were computationally demanding and to save computation time, single imputation was used to handle missing data, whereas multiple imputation would have been optimal. Third, the optimized SL-algorithm was based on cross-validated base-learners to prevent overfitting, but generalizability of the findings should still be independently evaluated. Also, the performance of the SL itself was not evaluated through cross-validation, adding to the necessity of external evaluation. Fourth, the use of class-probabilities rather than memberships as outcomes make results harder to interpret, but the low entropy indicated that participants could not be allocated to classes with acceptable certainty (Celeux & Soromenho, 1996). Fifth, LCGA model selection was strongly driven by the requirement of sufficiently large classes for the SL, whereas a range of fit indices is often primarily used to identify the optimal class-solution (e.g., Nylund et al., 2007). Sixth, to allow inclusion of the same variables in all analyses, only baseline variables measured irrespective of diagnosis were included; more diagnosis-specific course determinants may have been missed. Finally, the final SL algorithm can be seen as a 'black box' that is hard to interpret. For instance, the SL may include complex interactions, but provides no direct insight into their nature or importance. Importantly, model transparency is sacrificed here in favour of optimal predictions.

When implemented in clinical practice, ML-based algorithms could primarily be used to predict patients' general prognosis irrespective of treatment. A next step is to develop algorithms that predict course conditional on treatment type, which could be used to select a patient's most suitable treatment. To develop such algorithms, larger datasets with detailed treatment information are needed.

In conclusion, when optimal SL-based prediction models are applied, the most important determinants in the algorithms were mostly shared for the pure-depression and pure-anxiety outcomes and also showed sizable overlap with determinants of general psychological distress course. Domain-specific exceptions were healthcare use for pure depression and somatic arousal and distress for pure anxiety-severity course. This indicates that the both depression and anxiety course can be contemporaneously predicted based on mostly shared and some domain-specific of predictors.

Role of the funding source

The funders of this study played no role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication.

Author contributions

KJW, HR, EJG and BP came up with the study concept. KJW conducted the data analyses with input from EJG. KJW wrote the initial draft with feedback from HR. EJG, ME, AvH, BP, AJB, and RAS contributed further feedback on the manuscript. All authors approved the final version of the article.

Data availability statement

According to European law (GDPR) data containing potentially identifying or sensitive patient information are restricted; our data involving clinical participants are not freely available in a public repository. However, data are – under some specifications - available upon request via the NESDA Data Access Committee (nesda@ggzingeest.nl). See also our website: www.nesda.nl.

Declaration of Competing Interest

None

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 (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). All authors had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2021.06.029.

References

Ambresin, G., Chondros, P., Dowrick, C., Herrman, H., Gunn, J.M., 2014. Self-rated health and long-term prognosis of depression. Ann. Fam. Med. 12 (1), 57–65.
Angst, J., 1996. Comorbidity of mood disorders: a longitudinal prospective study. Br. J. Psychiatry. 168, 58–67.
Angst, J., Vollrath, M., 1991. The natural history of anxiety disorders. Acta. Psychiatr. Scand. 84, 446–452.
Archer, K.J., Kimes, R.V., 2008. Empirical characterization of random forest variable importance measures. Comput. Stat. Data Anal. 52, 2249–2260.
Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. J. Consult. Clin. Psychol. 56, 893–897.
Beck, A.T., Kovacs, M., Weissman, A., 1979. Assessment of suicidal intention: the Scale for Suicide Ideation. J. Consult. Clin. Psych. 47, 343–352.
Beesdo-Baum, K., Knappe, S., Fehm, L., Höfler, M., Lieb, R., Hofmann, S.G., Wittchen, H. U., 2012. The natural course of social anxiety disorder among adolescents and young adults. Acta. Psychiatr. Scand. 126 (6), 411–425.
Blechert, J., Michael, T., Grossman, P., Lajtman, M., Wilhelm, F.H., 2007. Autonomic and respiratory characteristics of posttraumatic stress disorder and panic disorder. Psychosom. Med. 69, 935–943.
Bokma, W.A., Zhutovsky, P., Giltay, E.J., Schoevers, R.A., Penninx, B.W.J.H., van Balkom, A.L.J.M., Batelaan, N.M., van Wingen, G.A., in press. Predicting the naturalistic course in anxiety disorders using clinical and biological markers: a machine learning approach. Psychol. Med.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brodaty, H., Luscombe, G., Peisah, C., Anstey, K., Andrews, G., 2001. A 25-year longitudinal, comparison study of the outcome of depression. Psychol. Med. 31 (8), 1347–1359.
Bruce, S.E., Yonkers, K.A., Otto, M.W., Eisen, J.L., Weisberg, R.B., Pagano, M., Shea, M. T., Keller, M.B., 2005. Influence of psychiatric comorbidity on recovery and recurrence in generalized anxiety disorder, social phobia, and panic disorder: a 12-year prospective study. Am. J. Psychiatry 162 (6), 1179–1187.
Brugha, T., Bebbington, P., Tennant, C., Hurry, J., 1985. The list of threatening experiences: a subset of 12 life event categories with considerable long-term contextual threat. Psychol. Med. 15, 189–194.
Celeux, G., Soromenho, G., 1996. An entropy criterion for assessing the number of clusters in a mixture model. J. Classif. 13, 195–212.
Chen, L.S., Eaton, W.W., Gallo, J.J., Nestadt, G., Crum, R., 2000. Empirical examination of current depression categories in a population-based study: symptoms, course, and risk factors. Am. J. Psych. 157, 573–580.

Clark, L.A., Watson, D., 1991. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. J. Abnorm. Psychol. 100 (3), 316–336.
Clark, S.L., Muthén, B.O., 2009. Relating latent class analysis results to variables not included in the analysis. https://www.statmodel.com/download/relatinglca.pdf (Accessed 11 June 2021).
Colman, I, Ploubidis, GB, Wadsworth, ME, Jones, PB, Croudace, TJ, 2007. Biol Psychiatry. 62, 1265–1271. https://doi.org/10.1016/j.biopsych.2007.05.012.
Coryell, W., Fiedorowicz, J.G., Solomon, D., Leon, A.C., Rice, J.P., Keller, M.B., 2012. Effects of anxiety on the long-term course of depressive disorders. Br. J. Psychiatry. 200 (3), 210–215.
Coryell, W., Leon, A., Winokur, G., Endicott, J., Keller, M., Akiskal, H., Solomon, D., 1996. Importance of psychotic features to long-term course in major depressive disorder. Am. J. Psychiatry 153 (4), 483–489.
Costa jr, P.T., McCrae, R.R., 1992. Revised NEO Personality Inventory and NEO Five-Factor Inventory professional manual. Psychological Assessment Resources, Odessa, Florida.
Cowley, D.S., Flick, S.N., Roy-Byrne, P.P., 1996. Long-term course and outcome in panic disorder: a naturalistic follow-up study. Anxiety 2 (1), 13–21.
Coyle, J.R., Hejazi, N.S., Malenica, I., Sofrygin, O., 2020. sl3: Modern Pipelines for Machine Learning and Super Learning. package version 1.3.7. (https://github.com/tlverse/sl3).
Craig, C.L., Marshall, A.L., Sjostrom, M., Bauman, A., Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F., Oja, P., 2003. International physical activity questionnaire: 12-country reliability and validity. Med. Sci. Sports. Exerc. 35, 1381–1395.
de Geus, E.J., Willemsen, G.H., Klaver, C.H., van Doornen, L.J., 1995. Ambulatory measurement of respiratory sinus arrhythmia and respiration rate. Biol. Psychol. 41, 205–227.
de Graaf, R., Bijl, R.V., Smit, F., Vollebergh, W.A., Spijker, J., 2002. Risk factors for 12-month comorbidity of mood, anxiety, and substance use disorders: findings from the Netherlands Mental Health Survey and Incidence Study. Am. J. Psychiatry. 159, 620–629.
de Jong Gierveld, J., Kamphuis, F., 1985. The development of a Rasch-type loneliness scale. Appl. Psychol. Meas. 9, 289–299.
Dinga, R., Marquand, A.F., Veltman, D.J., Beekman, A.T.F., Schoevers, R.A., van Hemert, A.M., Penninx, B.W.J.H., Schmaal, L., 2018. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. Transl. Psychiatry. 8 (1), 241.
Eronen, M.I., 2019. The levels problem in psychopathology. Psychol. Med. 24, 1–7.
Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults, 2001. Executive Summary of the Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). JAMA. 285, 2486–2497.
Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., Hutter, F., 2015. Efficient and robust automated machine learning. Adv. Neural Inf. Process. Syst. 28, 2962–2970.
Ferguson, C.J., 2009. An effect size primer: A guide for clinicians and researchers. Prof. Psychol. Res. Pr. 40, 532–538.
Ferro, M.A., Gorter, J.W., Boyle, M.H., 2015. Trajectories of depressive symptoms during the transition to young adulthood: the role of chronic illness. J. Affect. Disord. 174, 594–601.
Fichter, M.M., Kohlboeck, G., Quadflieg, N., 2008. The Upper Bavarian longitudinal community study 1975-2004. 2. Long-term course and outcome of depression. A controlled study. Eur. Arch. Psychiatry. Clin. Neurosci. 258 (8), 476–488.
Fichter, M.M., Quadflieg, N., Fischer, U.C., Kohlboeck, G., 2010. Twenty-five-year course and outcome in anxiety and depression in the Upper Bavarian Longitudinal Community Study. Acta. Psychiatr. Scand. 122, 75–85.
Friedman., J, 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist. 29, 1189–1232.
Gerrits, M.M., van Marwijk, H.W., van Oppen, P., van der Horst, H., Penninx, B.W., 2015. Longitudinal association between pain, and depression and anxiety over four years. J. Psychosom. Res. 78 (1), 64–70.
Goodwin, G.M., 2015. The overlap between anxiety, depression, and obsessive-compulsive disorder. Dialogues Clin. Neurosci. 17, 249–260.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition. Springer-Verlag.
Hirschfeld, R.M., Williams, J.B., Spitzer, R.L., Calabrese, J.R., Flynn, L., Keck Jr, P.E., Lewis, L., McElroy, S.L., Post, R.M., Rapport, D.J., Russell, J.M., Sachs, G.S., Zajecka, J., 2000. Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. Am. J. Psychiatry. 157, 1873–1875.
Hovenkamp-Hermelink, J.H., Riese, H., van der Veen, D.C., Batelaan, N.M., Penninx, B. W., Schoevers, R.A., 2016. Low stability of diagnostic classifications of anxiety disorders over time: A six-year follow-up of the N.ESDA study. J. Affect. Disord. 190, 310–315.
Hovenkamp-Hermelink, J.H.M., Jeronimus, B.F., van der Veen, D.C., Spinhoven, P., Penninx, B.W.J.H., Schoevers, R.A., Riese, H., 2019. Differential associations of locus of control with anxiety, depression and life-events: A five-wave, nine-year study to test stability and change. J. Affect. Disord. 253, 26–34.
Hovenkamp-Hermelink, J.H.M., Jeronimus, B.F., Myroniuk, S., Riese, H., Schoevers, R. A., 2021. What predicts persistence of anxiety disorders across the lifespan? A systematic review. Lancet Psychiatry 8 (5), 428–443.
Hovens, J.G., Wiersma, J.E., Giltay, E.J., van Oppen, P., Spinhoven., P., Penninx, B.W., Zitman, F.G., 2010. Childhood life events and childhood trauma in adult patients with depressive, anxiety and comorbid disorders vs. controls. Acta. Psychiatr. Scand. 122 (1), 66–74 s.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: With Applications in R. Springer, New York.
Judd, L.L., Akiskal, H.S., Maser, J.D., Zeller, P.J., Endicott, J., Coryell, W., Paulus, M.P., Kunovac, J.L., Leon, A.C., Mueller, T.I., Rice, J.A., Keller, M.B., 1998. A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. Arch. Gen. Psychiatry. 55 (8), 694–700.
Judd, L.L., Akiskal, H.S., Zeller, P.J., Paulus, M., Leon, A.C., Maser, J.D., Endicott, J., Coryell, W., Kunovac, J.L., Mueller, T.I., Rice, J.P., Keller, M.B., 2000a. Psychosocial disability during the long-term course of unipolar major depressive disorder. Arch. Gen. Psychiatry. 57 (4), 375–380.
Judd, L.L., Paulus, M.J., Schettler, P.J., Akiskal, H.S., Endicott, J., Leon, A.C., Maser, J. D., Mueller, T., Solomon, D.A., Keller, M.B., 2000b. Does incomplete recovery from first lifetime major depressive episode herald a chronic course of illness? Am. J. Psychiatry. 157 (9), 1501–1504.
Kanner, A.D., Coyne, J.C., Schaefer, C., Lazarus, R.S., 1981. Comparison of two modes of stress measurement: daily hassles and uplifts versus major life events. J. Behav. Med. 4, 1–39.
Katschnig, H., Amering, M., 1998. The long-term course of panic disorder and its predictors. J. Clin. Psychopharmacol. 18 (6 Suppl 2), 6S–11S.
Keller, M.B., Lavori, P.W., Mueller, T.I., Endicott, J., Coryell, W., Hirschfeld, R.M., Shea, T., 1992. Time to recovery, chronicity, and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects. Arch. Gen. Psychiatry. 49 (10), 809–816.
Keller, M.B., 2003. The lifelong course of social anxiety disorder: a clinical perspective. Acta. Psychiatr. Scand. (Suppl. 417), 85–94.
Keller, M.B., 2002. The long-term clinical course of generalized anxiety disorder. J. Clin. Psychiatry. 63 (Suppl 8), 11–16.
Kendell, R., Jablensky, A., 2003. Distinguishing between the validity and utility of psychiatric diagnoses. Am. J. Psychiatry. 160 (1), 4–12.
Kennedy, N., Abbott, R., Paykel, E.S., 2004. Longitudinal syndromal and sub-syndromal symptoms after severe depression: 10-year follow-up study. Br. J. Psychiatry. 184, 330–336.
Kennedy, N., Abbott, R., Paykel, E.S., 2003. Remission and recurrence of depression in the maintenance era: long-term outcome in a Cambridge cohort. Psychol. Med. 33 (5), 827–838.
Kessler, R.C., Rose, S., Koenen, K.C., Karam, E.G., Stang, P.E., Stein, D.J., Heeringa, S.G., Hill, E.D., Liberzon, I., McLaughlin, K.A., McLean, S.A., Pennell, B.E., Petukhova, M., Rosellini, A.J., Ruscio, A.M., Shahly, V., Shalev, A.Y., Silove, D., Zaslavsky, A.M., Angermeyer, M.C., Bromet, E.J., de Almeida, J.M., de Girolamo, G., de Jonge, P., Demyttenaere, K., Florescu, S.E., Gureje, O., Haro, J.M., Hinkov, H., Kawakami, N., Kovess-Masfety, V., Lee, S., Medina-Mora, M.E., Murphy, S.D., Navarro-Mateu, F., Piazza, M., Posada-Villa, J., Scott, K., Torres, Y., Viana, Carmen, M., 2014. How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? An exploratory study in the WHO World Mental Health Surveys. World Psychiatry 13, 265–274.
Kotov, R., Jonas, K.G., Carpenter, W.T., Dretsch, M.N., Eaton, N.R., Forbes, M.K., Forbush, K.T., Hobbs, K., Reininghaus, U., Slade, T., South, S.C., Sunderland, M., Waszczuk, M.A., Widiger, T.A., Wright, A.G.C., Zald, D.H., Krueger, R.F., Watson, D., HiTOP Utility Workgroup, 2020. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): I. Psychosis superspectrum. World Psychiatry 19, 151–172.
LeDell, E., Van der Laan, M.J., Petersen, M., 2016. AUC-maximizing ensembles through metalearning. Int. J. Biostat. 12, 203–218.
Levine, D.W., Kripke, D.F., Kaplan, R.M., Lewis, M.A., 2003. Reliability and validity of the women's health initiative insomnia scale. Psychol. Assess. 15, 123–136.
Licht, C.M.M., de Geus, E.J.C., Zitman, F.G., Hoogendijk, W.J.G., van Dyck, R., Penninx, B.W.J.H., 2008. Association Between Major Depressive Disorder and Heart Rate Variability in the Netherlands Study of Depression and Anxiety (NESDA). Arch. Gen. Psychiatry. 65, 1358–1367.
Nylund, K.L., Asparouhov, T., Muthén, B.O., 2007. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. SEM 14 (4), 535–569.
Marks, I.M., Mathews, A.M., 1979. Brief standard self-rating for phobic patients. Behav. Res. Ther. 17, 263–267.
McIntyre, R.S., O'Donovan, C., 2004. The human cost of not achieving full remission in depression. Can. J. Psychiatry. 49 (3 Suppl 1), 10S–16S.
McNamara, M.E., Shumake, J., Stewart, R.A., Labrada, J., Alario, A., Allen, J.J.B., Palmer, R., Schnyer, D.M., McGeary, J.E., Beevers, C.G., 2021. Multifactorial prediction of depression diagnosis and symptom dimensions. Psychiatry Res 298, 113805.
Meadows, G., Harvey, C., Fossey, E., Burgess, P., 2000. Assessing perceived need for mental health care in a community survey: development of the Perceived Need for Care Questionnaire (PNCQ). Soc. Psychiatry. Psychiatr. Epidemiol. 35, 427–435.
Merikangas, K.R., Zhang, H., Avenevoli, S., Acharyya, S., Neuenschwander, M., Angst, J., 2003. Zurich Cohort Study. Longitudinal trajectories of depression and anxiety in a prospective community study: the Zurich Cohort Study. Arch. Gen. Psychiatry. 60 (10), 993–1000.
Meyer, T.J., Miller, M.L., Metzger, R.L., Borkovec, T.D., 1990. Development and validation of the Penn State Worry Questionnaire. Behav. Res. Ther. 28, 487–495.
Moos, R.H., Cronkite, R.C., 1999. Symptom-based predictors of a 10-year chronic course of treated depression. J. Nerv. Ment. Dis. 187 (6), 360–368.
Mund, M., Freuding, M.M., Möbius, K., Horn, N., Neyer, F.J., 2020. The Stability and Change of Loneliness Across the Life Span: A Meta-Analysis of Longitudinal Studies. Personality and social psychology review 24 (1), 24–52.

Musliner, K.L., Munk-Olsen, T., Laursen, T.M., Eaton, W.W., Zandi, P.P., Mortensen, P.B., 2016. Heterogeneity in 10-Year Course Trajectories of Moderate to Severe Major Depressive Disorder: A Danish National Register-Based Study. JAMA psychiatry 73, 346–353.
Olino, T.M., Klein, D.N., Lewinsohn, P.M., Rohde, P., Seeley, J.R., 2010. Latent trajectory classes of depressive and anxiety disorders from adolescence to adulthood: descriptions of classes and associations with risk factors. Compr. Psychiatry. 51 (3), 224–235.
Ormel, J., Jeronimus, B.F., Kotov, R., Riese, H., Bos, E.H., Hankin, B., Rosmalen, J.G.M., Oldehinkel, A.J., 2013. Neuroticism and common mental disorders: meaning and utility of a complex relationship. Clin. Psychol. Rev. 33 (5), 686–697.
Papp, L.A., Martinez, J.M., Klein, D.F., Coplan, J.D., Norman, R.G., Cole, R., de Jesus, M. J., Ross, D., Goetz, R., Gorman, J.M., 1997. Respiratory psychophysiology of panic disorder: three respiratory challenges in 98 subjects. Am. J. Psychiatry. 154 (11), 1557–1565.
Pearlin, L.I., Schooler, C., 1978. The structure of coping. J. Health. Soc. Behav. 19, 2–21.
Penninx, B.W., Beekman, A.T., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W., Assendelft, W.J., Van Der Meer, K., Verhaak, P., Wensing, M., De Graaf, R., Hoogendijk, W.J., Ormel, J., Van Dyck, R., 2008. The Netherlands Study of Depression and Anxiety (NESDA): rationale, nobjectives and methods. Int. J. Methods. Psychiatr. Res. 17, 121–140.
Pettit, J.W., Lewinsohn, P.M., Roberts, R.E., Seeley, J.R., Monteith, L., 2009. The long-term course of depression: development of an empirical index and identification of early adult outcomes. Psychol. Med. 39 (3), 403–412.
Piccinelli, M., Wilkinson, G., 1994. Outcome of depression in psychiatric settings. Br. J. Psychiatry. 164 (3), 297–304.
Proust-Lima, C., Philipps, V., Liquet, B., 2017. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. J. Stat. Softw. 78, 1–56.
R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.
Rhebergen, D., Batelaan, N.M., de Graaf, R., Nolen, W.A., Spijker, J., Beekman, A.T., Penninx, B.W., 2011. The 7-year course of depression and anxiety in the general population. Acta. Psychiatr. Scand. 123 (4), 297–306.
Riese, H., Groot, P.F., van den Berg, M., Kupper, N.H., Magnee, E.H., Rohaan, E.J., Vrijkotte, T.G., Willemsen, G., de Geus, E.J., 2003. Large-scale ensemble averaging of ambulatory impedance cardiograms. Behav. Res. Methods Instrum. Comput. 35, 467–477.
Rose, S., 2013. Mortality risk score prediction in an elderly population using machine learning. Am. J. Epidemiol. 177, 443–452.
Rosellini, A.J., Liu, S., Anderson, G.N., Sbi, S., Tung, A.S., Knyazhanskaya, E., 2020. Developing algorithms to predict adult onset internalizing disorders: An ensemble learning approach. J. Psychiatr. Res. 121, 189–196.
Rosellini, A.J., Dussaillant, F., Zubizarreta, J.R., Kessler, R.C., Rose, S., 2018a. Predicting posttraumatic stress disorder following a natural disaster. J. Psychiatr. Res. 96, 15–22.
Rosellini, A.J., Stein, M.B., Benedek, D.M., Bliese, P.D., Chiu, W.T., Hwang, I., Monahan, J., Nock, M.K., Sampson, N.A., Street, A.E., Zaslavsky, A.M., Ursano, R.J., Kessler, R.C., 2018b. Predeployment predictors of psychiatric disorder-symptoms and interpersonal violence during combat deployment. Depress. Anxiety. 35, 1073–1080.
Rubio, G., López-Ibor, J.J., 2007. Generalized anxiety disorder: a 40-year follow-up study. Acta Psychiatr Scand 115 (5), 372–379.
Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol. Med. 26, 477–486.
Saunders, J.B., Aasland, O.G., Babor, T.F., de la Fuente, J.R., Grant, M., 1993. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption II. Addiction 88, 791–804.
Schaakxs, R., Comijs, H.C., Lamers, F., Kok, R.M., Beekman, A.T.F., Penninx, B.W.J.H., 2018. Associations between age and the course of major depressive disorder: a 2-year longitudinal cohort study. Lancet Psychiatry 5, 581–590.
Schopman, S, Ten Have, M, van Balkom, AJ, de Graaf, R, Batelaan, N, n.d.. Course trajectories of anxiety disorders: Results from a 6-year follow-up in a general population study. Aust NZ J Psychiatry. In press.
Smith, G., 2018. Step away from stepwise. J. Big. Data. 5, 32.
Solis, E, van Hemert, AM, Carlier, IVE, Wardenaar, KJ, Schoevers, RA, Beekman, ATF, Penninx, BWJH, Giltay, EJ, n.d. The 9-year clinical course of depressive and anxiety disorders: new NESDA findings. J Affect Disord. Submitted for publication.
Spinhoven, P., Batelaan, N., Rhebergen, D., van Balkom, A., Schoevers, R., Penninx, B. W., 2016. Prediction of 6-yr symptom course trajectories of anxiety disorders by diagnostic, clinical and psychological variables. J Anxiety. Disord. 44, 92–101.
Steinert, C., Hofmann, M., Kruse, J., Leichsenring, F., 2014. The prospective long-term course of adult depression in general practice and the community. A systematic literature review. J. Affect. Disord. 152-154, 65–75.
Steunenberg, B., Beekman, A.T., Deeg, D.J., Bremmer, M.A., Kerkhof, A.J., 2007. Mastery and neuroticism predict recovery of depression in later life. Am. J. Geriatr. Psychiatry. 2007 15 (3), 234–242.
Struijs, S.Y., Lamers, F., Spinhoven, P., van der Does, W., Penninx, B.W.J.H., 2018. The predictive specificity of psychological vulnerability markers for the course of affective disorders. J. Psychiatr. Res. 103, 10–17.
Surtees, P.G., Wainwright, N.W., 1996. Fragile states of mind: neuroticism, vulnerability and the long-term outcome of depression. Br. J. Psychiatry. 169 (3), 338–347.
Terluin, B., van Marwijk, H.W., Adèr, H.J., de Vet, H.C., Penninx, B.W., Hermens, M.L., van Boeijen, C.A., van Balkom, A.J., van der Klink, J.J., Stalman, W.A., 2006. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. BMC Psychiatry 22, 6–34.
Ustun, T.B., Kostanjsek, N., Chatterji, S., Rehm, J., 2010. Measuring health and disability: manual for WHO Disability Assessment Schedule (WHODAS 2.0). World Health Organization, Geneva, Switzerland.
van Buuren, S., Groothuis-Oudshoorn, K., 2011. MICE: Multivariate Imputation by Chained Equations in R. J. Stat. Softw. 45 (3), 1–67.
van der Laan, M.J., Polley, E.C., Hubbard, A.E, 2007. Super learner. Stat. Appl. Genet. Mol. Biol. 6, 25.
van der Laan, M.J., Rose, S., 2011. Targeted Learning: Causal Inference for Observational and Experimental Data. Springer, New York.
van Eeden, W.A., Luo, C., van Hemert, A.M., Carlier, I.V.E., Penninx, B.W., Wardenaar, K. J., Hoos, H., Giltay, E.J., 2021. 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 Res 299, 113823.
Verduijn, J., Verhoeven, J.E., Milaneschi, Y., Schoevers, R.A., van Hemert, A.M., Beekman, A.T.F., Penninx, B.W.J.H., 2017. Reconsidering the prognosis of major depressive disorder across diagnostic boundaries: full recovery is the exception rather than the rule. BMC Med 15 (1), 215.
Virtanen, M., Ferrie, J.E., Akbaraly, T., Tabak, A., Jokela, M., Ebmeier, K.P., Singh-Manoux, A., Kivimäki, M., 2017. Metabolic Syndrome and Symptom Resolution in Depression: A 5-Year Follow-Up of Older Adults. J. Clin. Psychiatry. 78 (1), e1–e7.
Von Korff, M., Ormel, J., Keefe, F.J., Dworkin, SF., 1992. Grading the severity of chronic pain. Pain 50, 133–149.
Wardenaar, K.J., van Veen, T., Giltay, E.J., de Beurs, E., Penninx, B.W., Zitman, F.G., 2010. Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). Psychiatry Res 179, 101–106.
Wardenaar, K.J., Conradi, H.J., Bos, E.H., de Jonge, P., 2014. Personality modulates the efficacy of treatment in patients with major depressive disorder. J. Clin. Psychiatry. 75 (9), e916–e923.
Wardenaar, K.J., Monden, R., Conradi, H.J., de Jonge, P., 2015. Symptom-specific course trajectories and their determinants in primary care patients with Major Depressive Disorder: Evidence for two etiologically distinct prototypes. J. Affect. Disord. 179, 38–46.
Webb, C.A., Cohen, Z.D., Beard, C., Forgeard, M., Peckham, A.D., Björgvinsson, T., 2020. Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. J. Consult. Clin. Psych. 88, 25–38.
Wiersma, J.E., van Oppen, P., van Schaik, D.J., van der Does, A.J., Beekman, A.T., Penninx, B.W., 2011. Psychological characteristics of chronic depression: a longitudinal cohort study. J Clin Psychiatry 72 (3), 288–294.
Wittchen, H.U., Fehm, L., 2003. Epidemiology and natural course of social fears and social phobia. Acta. Psychiatr. Scand. Suppl. (417), 4–18.
Zisook, S., Lesser, I., Stewart, J.W., Wisniewski, S.R., Balasubramani, G.K., Fava, M., Gilmer, W.S., Dresselhaus, T.R., Thase, M.E., Nierenberg, A.A., Trivedi, M.H., Rush, AJ., 2007. Effect of age at onset on the course of major depressive disorder. Am. J. Psychiatry. 164, 1539–1546.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Series. B. 67, 301–320.