Learning to swim the endless waves: examining self-management and risk factors for (chronic) depression

Solis, E.C.

CHAPTER 6

The 9-year clinical course of depressive and anxiety disorders: new NESDA findings.

Ericka C. Solis [a],* Albert M. van Hemert [a], Ingrid V.E. Carlier [a], Klaas J. Wardenaar [b], Robert A. Schoevers [b], Aartjan T.F. Beekman [c,d], Brenda W.J.H. Penninx [a,b,c], Erik J. Giltay [a]

[a] Leiden University Medical Center, Department of Psychiatry, Leiden, The Netherlands

[b] Department of Psychiatry, University Medical Center Groningen, Interdisciplinary Center Psychopathology and Emotion regulation, University of Groningen, Groningen, The Netherlands

[c] Department of Psychiatry and Amsterdam Public Health research institute, Amsterdam UMC, Vrije Universiteit, Amsterdam, The Netherlands

[d] GGZ inGeest Specialized Mental Health Care, Amsterdam, The Netherlands

Journal of Affective Disorders. 2021; 295: 1269-1279

Abstract

Background: In longitudinal research, switching between diagnoses should be considered when examining patients with depression and anxiety. We investigated course trajectories of affective disorders over a nine-year period, comparing a categorical approach using diagnoses to a dimensional approach using symptom severity. **Method:** Patients with a current depressive and/or anxiety disorder at baseline (*N* = 1701) were selected from the Netherlands Study of Depression and Anxiety (NESDA). Using psychiatric diagnoses, we described 'consistently recovered,' 'intermittently recovered, 'intermittently recurrent', and 'consistently chronic' at two-, four-, six-, and nine-year follow-up. Additionally, latent class growth analysis (LCGA) using depressive, anxiety, fear, and worry symptom severity scores was used to identify distinct classes. **Results:** Considering the categorical approach, 8.5% were chronic, 32.9% were intermittently recurrent, 37.6% were intermittently recovered, and 21.0% remained consistently recovered from any affective disorder at nine-year follow-up. In the dimensional approach, 66.6% were chronic, 25.9% showed partial recovery, and 7.6% had recovered. **Limitations:** 30.6% of patients were lost to follow-up. Diagnoses were rated by the interviewer and questionnaires were completed by the participant. **Conclusions:** Using diagnoses alone as discrete categories to describe clinical course fails to fully capture the persistence of affective symptoms that were observed when using a dimensional approach. The enduring, fluctuating presence of subthreshold affective symptoms likely predisposes patients to frequent relapse. The commonness of subthreshold symptoms and their adverse impact on long-term prognoses deserve continuous clinical attention in mental health care as well further research.

Keywords *(3-6 words):* Depression; Anxiety; Nine-year Course; Prognosis; Chronicity; Diagnostic Switching

Key findings

- When using a categorical approach (i.e., diagnoses), the majority of patients recovered from affective disorders at nine-year follow-up.
- When using a dimensional approach (i.e., symptom severity), the majority of patients had a more chronic course.
- The majority of patients who did not meet criteria for an affective disorder still experienced subthreshold symptomatology.

1. Introduction

Longitudinal research is essential to validate diagnostic classifications and tailor treatment plans for optimal effectiveness and efficiency for patients with affective disorders (Gillis et al., 1995; Kendell and Jablensky, 2003; Kraepelin, 1921; Lorenzo-Luaces et al., 2017; McGorry et al., 2016; Penninx et al., 2011). In longitudinal research, depressive and anxiety disorders should be investigated synchronously considering that major depressive disorder (MDD) and anxiety disorders often co-occur (Angst et al., 2009; Giandinoto and Edward, 2015; Hayden and Klein, 2001; Howland et al., 2009; Kessler et al., 2005; Kleiboer et al., 2016; Lamers et al., 2011; Moffitt et al., 2007; Rush et al., 2006), and as time passes, one disorder will often switch over to the other ("diagnostic switching") (Gregory et al., 2007; Hovenkamp-Hermelink et al., 2016; Scholten et al., 2016). For example, those recovered from MDD may then meet criteria for generalized anxiety disorder (GAD) and still experience functional impairment and a decreased quality of life (Angst et al., 2009; Maj et al., 2002; Wells et al., 1992). Patients may also switch over to dysthymic disorder, the more chronic but milder form of depression (Klein et al., 2008; Rhebergen et al., 2012).

The selected observation period may also reveal crucial insights into the course of depressive and anxiety disorders. A shorter window of less than two years has led to relatively positive findings with regard to prognosis (Gilchrist and Gunn, 2007; Hendriks et al., 2013; Richards, 2011; Steinert et al., 2014). For example, previous research has shown that more than half (50 – 70%) of patients recovered within one year from MDD (Richards, 2011; Steinert et al., 2014), and approximately half (43 – 73%) of patients recovered within two years from anxiety disorders (Hendriks et al., 2013). However, when extending the observation window, lower levels of recovery were found, which may be due to diagnostic switching (Caspi et al., 2020; Gregory et al., 2007; Hovenkamp-Hermelink et al., 2016; Plana-Ripoll et al., 2019; Scholten et al., 2016), relapse (Hardeveld et al., 2010; Mueller et al., 1999; Scholten et al., 2013; Verduijn et al., 2017), or continuous subthreshold or subclinical symptoms (Ormel et al., 1993; Wagner et al., 2000).

The course of depressive and anxiety disorders can be and have been described using two clinically-relevant approaches: with psychiatric diagnosis criteria according to the Diagnostic Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) (i.e. categorical approach) or scores on symptom severity scales (i.e. dimensional approach) (Wardenaar et al., 2014). Use of DSM diagnoses allows us to examine participants who (continue to) meet full DSM criteria over time, whether these are comorbid diagnoses or transitions to other diagnoses. Alternatively, symptom severity scores are regularly used throughout the treatment follow-up process to assess change of symptoms over time, such as in routine outcome monitoring (ROM) (e.g. de Beurs et al., 2011). Using symptom severity scores may also help reveal and describe the heterogeneity in (sub)clinical symptom profiles (Brunoni, 2017;

Fried and Nesse, 2015a). Considering that depression and anxiety DSM diagnoses can be viewed as discrete categorical syndromes imposed on a continuum of depressive or anxiety symptoms of varying severity and duration (Kendler and Gardner, 1998; Klein et al., 2006, 2008; Torpey and Klein, 2008), comparing both approaches is relevant. Latent class growth analysis (LCGA) with both depressive and anxiety symptoms would allow us to examine how symptoms change over time as a dimensional approach and to compare the data with DSM diagnoses as a categorical approach.

The present study aimed to describe and compare course trajectories of depressive and anxiety disorders over a nine-year period using a categorical approach for DSM-diagnosis trajectories and a dimensional approach for symptom pathways using LCGA. We considered depressive and anxiety comorbidity, switching to other diagnoses over time, and switching to other trajectories over time. We specified how participants intermittently recover and relapse over a nine-year period and how the DSM diagnoses concur with fluctuating symptom scores. For the categorical approach, we hypothesized that those with baseline comorbid depressive and anxiety disorders, compared to those with baseline depression or anxiety disorder, would have higher levels of chronicity at nine-year follow-up (Hendriks et al., 2013; Richards, 2011; Steinert et al., 2014). Likewise, for the dimensional approach, we hypothesized that those with higher baseline symptom levels would coincide with a more chronic course. Finally, we hypothesized that the trajectories according to the categorical approach would correspond highly with the dimensional symptom pathways.

2. Method

2.1 Study Sample and Procedure

Data were used from the Netherlands Study of Depression and Anxiety (NESDA) (Penninx et al., 2021), an ongoing longitudinal cohort study examining the course of depressive and anxiety disorders (for the extensive protocol, see Penninx et al., 2008). At baseline (2004-2007), a total of 2981 participants aged between 18 and 65 years were recruited from community (19.0%), primary care (54.0%), and specialized mental healthcare (27.0%). After baseline, face-to-face follow-up assessments were conducted at two years (87.1%, *n* = 2596), four years (80.6%, *n* = 2402), six years (75.5%, *n* = 2256), and nine years (69.4%, *n* = 2069). The nine-year follow-up was completed in 2016.

For the present study, we selected a total of 1701 participants with MDD, dysthymic disorder, and/or anxiety in the six months prior to baseline (i.e. a current six-month diagnosis) (see below).

2.2 Measures

2.2.1 Categorical course trajectories

To determine the categorical trajectories, we examined the presence of depressive (MDD, dysthymic disorder) and anxiety disorders (panic disorder, social phobia, generalized anxiety disorder (GAD), and agoraphobia) according to the DSM-IV criteria at each assessment. The disorders were assessed in person using the Composite Interview Diagnostic Instrument (CIDI) version 2.1 by trained research personnel (Wittchen, 1994; Wittchen and Nelson, 1996). To extrapolate whether the CIDI diagnoses were consistently present during follow-up periods, we examined the continuous presence of diagnosis-related symptoms reported in the Life Chart Interview (LCI; Lyketsos et al., 1994) for

that same follow-up period. More information regarding the instruments can be found in the **Supplementary Material**.

Using the CIDI, participants were first divided into subgroups at baseline: 1) depression only (single or any combination of MDD and/or dysthymic disorder); 2) anxiety only (single or any combination of panic disorder, social phobia, GAD, and agoraphobia); and 3) comorbid depression-anxiety (any combination of both depressive and anxiety disorders). The following "primary" categorical diagnosis trajectories were then described at two, four-, six-, and nine-year follow-up for the total group and for each of the subgroups by examining whether participants had any CIDI diagnosis at each following assessment after baseline: *'consistently recovered' (fixed diagnosis-free period at each assessment), 'intermittently recovered' (variable diagnosis-free period preceded by a diagnosis-present period with or without discontinuous LCI symptoms), 'intermittently recurrent' (variable diagnosis-present period with discontinuous LCI symptoms) and 'consistently chronic' (fixed diagnosis-present period at each assessment with continuous LCI symptoms).* The trajectories of 'intermittently recovered' and 'intermittently recurrent' allowed us to explore the 'ebb and flow' of diagnoses across the nine-year period. At nine-year follow-up, we defined three "comparison" trajectories, where each percentage was aggregated at each subsequent assessment point: *'recovered' (diagnosis-free period), 'recurrent' (diagnosis-present period with discontinuous LCI symptoms), and 'chronic' (diagnosis-present period with continuous LCI symptoms).* This allowed us to compare our results to Verduijn et al. (2017), who examined the six-year follow-up NESDA data. An extended description of the trajectories can also be found in the **Supplementary Material** (see "Description of categorical trajectories"). The specific differences

with Verduijn et al. (2017) are also outlined in the **Supplementary Material** (see "Comparison with previous NESDA study").

2.2.2 Dimensional course trajectories

To determine dimensional trajectories, we used depressive and anxiety symptom severity measures. For depressive symptom severity, we used the Inventory of Depressive Symptomatology – Self-Report (IDS-SR; Rush et al., 1996; Trivedi et al., 2004). For anxiety symptom severity, we used the Beck Anxiety Inventory - Self-Report (BAI; Beck et al., 1988; De Ayala et al., 2005; Muntingh et al., 2011; Steer et al., 1993); the Fear Questionnaire (FQ; Marks and Mathews, 1979; Oei et al., 1991); and the Penn State Worry Scale (PSWQ; Meyer et al., 1990; Salarifar and Pouretemad, 2012; Verkuil and Brosschot, 2012). Finally, we examined functional (dis)ability, which was measured using the World Health Organization Disability Assessment Schedule (WHO-DAS-II; Chwastiak and Von Korff, 2003). More information regarding these instruments can be found in the **Supplementary Material**.

2.3 Statistical Analyses

Sociodemographic characteristics. We reported baseline characteristics as percentages or as means with standard deviations (*SD*) (see **Table 1**). We inspected baseline symptom severity scores for clinically-relevant threshold levels (IDS > 13; BAI > 7; approximations FQ > 30; PSWQ > 48) (Gillis et al., 1995; Rush et al., 1996).

Categorical approach using diagnoses. For the categorical approach, we described the DSM-diagnoses trajectories using variable and aggregated percentages. We also examined the trajectories without any missing CIDI data

at all follow-up assessments. Moreover, we conducted a contingency table analysis with adjusted residuals (Beasley and Schumacker, 1995; García-Pérez and Núñez-Antón, 2003), to test whether there was an association between baseline diagnostic subgroup (i.e., depression only, anxiety only, and depression-anxiety) and the four course trajectories at nine-year follow-up.

Dimensional approach using symptom severity. For the dimensional approach, we used latent class growth analysis (LCGA), a type of mixture modeling with latent class analysis, to identify subgroups of participants with distinct patterns of change in their longitudinal symptom severity assessments relative to baseline (using the 'hlme' function of the 'lcmm' package in R version 3.6.1) (Proust-Lima et al., 2017). In LCGA, participants are clustered into classes representing trajectories with similar mean levels and mean-level changes of symptom severity by modeling a latent categorical variable. Unlike conventional growth mixture modeling (GMM), the analysis was conducted with the variance of the latent slope and intercept fixed to zero within class, thus limiting the heterogeneity within the latent classes, allowing us to identify course-trajectory classes that were optimally differentiated from each other (Armour et al., 2012; Asparouhov, 2006; Berlin et al., 2014; Cicchetti and Rogosch, 1999; Curran and Hussong, 2003; deRoon-Cassini et al., 2010; Jung and Wickrama, 2008; Muthén and Muthén, 2000; Nylund et al., 2007). For more information, see "Description and interpretation criteria of the latent class growth analysis" and the checklist for the Guidelines for Reporting on Latent Trajectory Studies (GRoLTS; van de Schoot, 2015) in the **Supplementary Material.**

Comparison of categorical and dimensional approaches. We compared the outcome of both the categorical and dimensional approaches at nine-year

follow-up. Hereafter, we examined the overlap across the identified dimensional symptom severity pathways and the primary and comparison DSM-diagnosis trajectories at nine-year follow-up. For the comparison, or aggregated, trajectories, where a 3 x 3 contingency table was possible, we tested the level of agreement between the categorical and dimensional approaches with the Kappa statistic as a complete-case analysis, where .01-.20 = none to slight/very poor, .21-.39 = minimal, .40-.59 = weak, .60-.79 = moderate and .80-.90 = strong (Cohen, 1960; de Raadt et al., 2019; McHugh, 2012). We then examined the average functional (dis)ability scores across the identified dimensional symptom severity pathways.

To further compare the symptom severity pathways with the DSM-diagnoses trajectories, we examined the trends of symptom severity per questionnaire across the assessment points based on baseline diagnosis subgroups and functional (dis)ability across the assessment points. We split the IDS-SR, BAI, FQ, PSWQ, and WHO-DAS-II scores into quartiles based on the baseline total score of that particular questionnaire and used mixed model regression analysis to yield marginal mean values at each follow-up assessment per questionnaire. Given that depression and anxiety frequently emerge in younger adulthood (Caspi et al., 2020; Ernst and Angst, 1992; Moffitt et al., 2007), we tested whether there was a potential effect of age and gender on depressive, anxiety, fear and worry severity symptoms. Age was split into quartiles, or four age groups. We then examined these trajectories over the nine-year follow-up.

Statistics programs for the analyses. The descriptive analyses and categorical approach analysis were conducted using IBM SPSS version 25.0. For the dimensional approach, we used R statistical software (R version 3.6.0; R

Foundation for Statistical Computing, Vienna, Austria, 2016. URL: https://www.R-project.org/). The mixed-model analyses were also conducted using R software using the 'lme4' package. Alpha was set at 0.05, except for the chi-square post hoc contrasts, where alpha was set to 0.0042 after a Bonferroni correction.

3. Results

3.1 Baseline sociodemographic information

Table 1 shows the baseline characteristics of the study sample (*N* = 1701). Participants were on average 41.3 years old (*SD* = 12.4); 67.0% were female (*n* = 1140), and 65.3% had a spouse or partner (*n* = 1110). The mean total IDS and BAI scores were 31.5 (*SD* = 10.9) and 18.3 (*SD* = 10.6), respectively, reflecting a moderate level of depressive and anxiety symptom severity at baseline. Nearly half of our participant sample (44.8%) had comorbid depression and anxiety at baseline.

3.2 Categorical approach using depression and anxiety diagnoses

Figure 1 presents four stacked bar charts showing the proportions of participants at each follow-up assessment who were 'consistently recovered,' 'intermittently recovered,' 'intermittently recurrent,' and 'consistently chronic,' divided by baseline diagnostic status. Over the course of nine years, the prognosis of each group was relatively positive: the percentage of participants who recovered increased, and the percentage of those with consistently chronic diagnoses decreased. Specifically, at nine-year follow-up, across all groups, between 51.9 and 68.6% had recovered from any disorder (when combining both consistent and intermittent recovery) and between 2.9 and 11.8%

maintained consistently chronic disorders. Approximately one-third of participants experienced intermittent recurrence across all groups at four-, six- and nine-year follow-up. Looking specifically at nine-year follow-up, baseline diagnostic status was significantly associated with course, χ^2 (6, N = 966) = 61.5, p < .0001, with a lower proportion of participants with baseline depression-anxiety comorbidity experiencing consistent recovery than participants with baseline depression (p < .0001), and a higher proportion of participants with depression-anxiety comorbidity experiencing consistent chronicity than participants with baseline depression (p = .0012). Additionally, when examining recovery more closely at nine-year follow-up, only 10.1% of participants with baseline comorbid depression-anxiety were consistently recovered, 26.0% of the anxiety-only group were consistently recovered, and 32.2% of the depression-only group were consistently recovered. Lastly, for the total group with any disorder, 21.0% were consistently recovered. When using a sample without missing CIDI diagnoses at any follow-up assessment (N = 956), we found the same pattern in the categorical trajectories. For comparison with trajectories of Verduijn et al. (2017), see Figure A and "Comparison with previous NESDA study" in the **Supplementary Material**. The main finding of the comparison trajectories is that 78.5% of participants with any disorder experienced recurrence over a nine-year period.

Table 1.
Baseline Socio-Demographic and Clinical Characteristics (N = 1701)

Age, years, mean (*SD*)	41.3 (12.4)
Sex, female, % (*n*)	67.0 (1140)
Education, years, mean (*SD*)	11.8 (3.3)
Married or life partner, yes, % (*n*)	65.3 (1110)
Depression only, % (*n*)	23.3 (396)
Single depressive disorder, % (*n*)	19.9 (339)
Comorbid depressive disorders, % (*n*)	3.4 (57)
Severity of depressive symptoms (IDS-SR), mean (*SD*)	27.6 (11.6)
Severity of anxiety symptoms (BAI), mean (*SD*)	12.1 (8.9)
Severity of fear symptoms (FQ), mean (*SD*)	21.7 (16.8)
Severity of worry symptoms (PSWQ), mean (*SD*)	29.3 (15.7)
Severity of functional disability (WHO-DAS-II), mean (*SD*)	31.6 (16.0)
Anxiety only, % (*n*)	31.9 (543)
Single anxiety disorder, % (*n*)	22.6 (384)
Comorbid anxiety disorders (2 to 3), % (*n*)	9.3 (159)
Severity of depressive symptoms (IDS-SR), mean (*SD*)	21.9 (9.8)
Severity of anxiety symptoms (BAI), mean (*SD*)	14.7 (9.5)
Severity of fear symptoms (FQ), mean (*SD*)	31.1 (18.5)
Severity of worry symptoms (PSWQ), mean (*SD*)	29.7 (14.4)
Severity of functional disability (WHO-DAS-II), mean (*SD*)	25.1 (15.4)
Comorbid depression and anxiety, % (*n*)	44.8 (762)
Severity of depressive symptoms (IDS-SR), mean (*SD*)	34.2 (13.1)
Severity of anxiety symptoms (BAI), mean (*SD*)	20.8 (11.5)
Severity of fear symptoms (FQ), mean (*SD*)	38.8 (21.9)
Severity of worry symptoms (PSWQ), mean (*SD*)	32.4 (18.2)
Severity of functional disability (WHO-DAS-II), mean (*SD*)	41.1 (17.2)

Note. Depression only = depressive disorders such as major depressive disorder or dysthymia. Anxiety only = anxiety disorders such as social phobia and generalized anxiety disorder. Comorbid depression and anxiety = MDD, dysthymic disorder and anxiety disorders. *IDS-SR* = Inventory of Depressive Symptomatology – Self-Report; *BAI* = Beck Anxiety Inventory; *FQ* = Fear Questionnaire; PSWQ = Penn State Worry Scale; WHO-DAS-II = World Health Organization Disability Assessment Schedule (standardized).

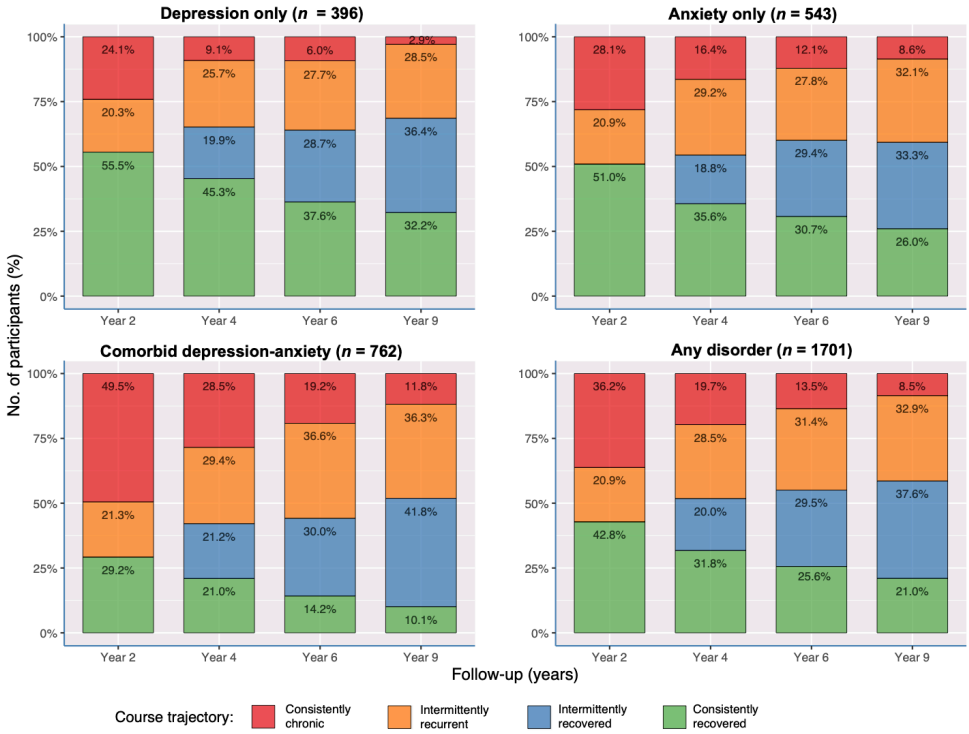

Figure 1. Categorical trajectories of diagnoses at two-, four-, six-, and nine-year assessments, based on any CIDI affective disorder and LCI data. Depression only = depressive disorders such as major depressive disorder or dysthymia. Anxiety only = anxiety disorders such as social phobia and generalized anxiety disorder.

3.3 Dimensional approach using symptom severity

Table 2 shows the fit indices resulting from the log-linear LCGA with one- to six-classes.

Table 2.
Fit Indices of One- to Six-Class Latent Class Growth Analysis over a Nine-Year Follow-Up (*N* = 1693)

Classes	Log-Likelihood	AIC	BIC	SA-BIC	Entropy	Percentage of Individuals in Class					
						1	2	3	4	5	6
1	-5735.8	11477.6	11493.9	11484.3	1.00	100	-	-	-	-	-
2	-4827.9	9667.8	9700.5	9681.4	0.71	22.1	77.8	-	-	-	-
3	-4598.8	9215.6	9264.5	9235.9	0.69	7.5	66.6	25.9	-	-	-
4	-4505.9	9035.9	9101.1	9063.0	0.62	4.7	14.1	18.9	62.4	-	-
5	--4452.1	8934.2	9015.7	8968.1	0.65	2.4	7.6	62.3	19.5	8.2	-
6	-4427.9	8891.8	8989.7	8932.5	0.66	1.2	2.4	61.4	7.5	19.8	7.8

Note. AIC = Akaike Information Criterion (Akaike, 1987); BIC = Bayesian Information Criterion (Schwartz, 1978); SA-BIC = sample-size-adjusted BIC (Sclove, 1987); Entropy (Ramaswamy et al., 1993).

Taking all fit indices into account, the three-class model was considered the optimal solution. The AIC and SA-BIC values showed a marked drop between one-, two- and three-class models (see elbow plots in **Figure B** in **Supplemental Material**), and hereafter, the differences in AIC and SA-BIC values were smaller, suggesting that the addition of more classes did not further improve the model. Also, solutions with four to six classes had very small numbers of participants (<5%) in some classes. Examining the entropy value of the three-class model (0.69) indicated that a moderate to high proportion of participants were correctly classified. The grid search also confirmed a three-class model.

Figure 2A delineates the three latent classes that were identified using pooled standardized mean severity scores of depressive, anxiety, fear, and worry symptom questionnaires: class 1: *chronic* (66.8%); class 2: *partially recovered* (25.7%); and class 3: *(sustained) recovered* (7.5%). Class 1 (red) remained stable and severe; whereas symptom levels decreased moderately in class 2 (blue) and markedly in class 3 (green). This was consistent with the latent classes of pooled symptom severity for depression, anxiety, fear, and worry symptom levels, viewing each questionnaire separately (see **Figure 2B**): the group with the highest symptom levels maintained high symptom levels throughout the nine-year period; the moderate level group showed slow decline and intermittent increases in anxiety and worry; and the lowest symptom level group showed a larger and steady decline.

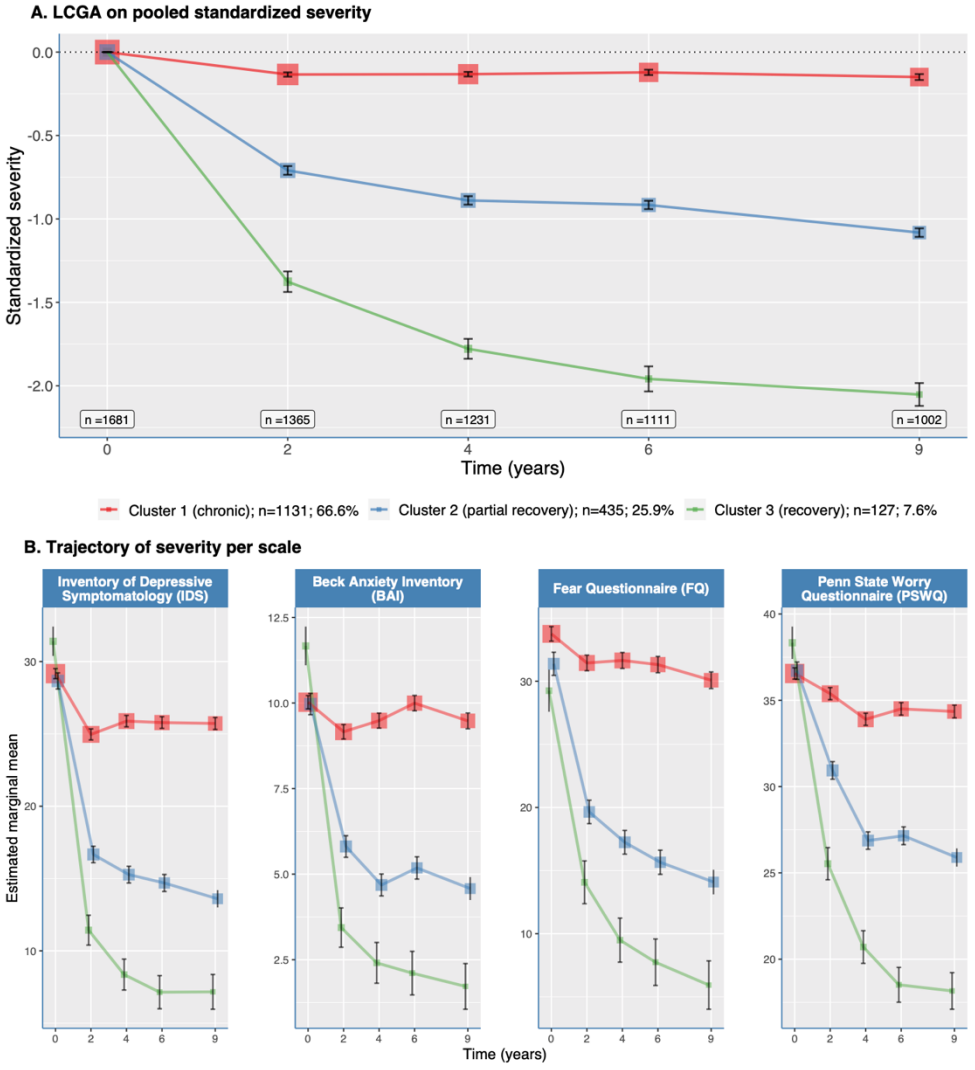

Figure 2. Three latent class solution of pooled standardized mean severity scores of depressive, anxiety, fear, and worry symptoms (LCGA: latent class growth analysis). Panel A shows the average change in standardized severity over time as compared to baseline severity, and Panel B shows the standardized severity scores per scale over time. Error bars represent standard errors, and the box size is proportional to the number of participants.

Moreover, inspection of the classes, or symptom pathways, showed that there was a relatively greater change in symptom severity between baseline and the first follow-up, after which the course showed less change between follow-up assessments and began to stabilize, suggesting that the log-linear function best described these symptom pathways. For a comparison with alternative models, see "Alternative models using latent class growth analysis" in the **Supplementary Material**. Recomputing the model without any missing data and comparing these to the abovementioned classes resulted in highly corresponding models: 94.9% similarity for class 1 (chronicity), 89.9% similarity for class 2 (partial recovery), and 100% similarity for class 3 (recovery).

3.4 Comparison of categorical and dimensional approaches

We compared the diagnosis trajectories of the total sample using the categorical approach at nine-year follow-up with the three identified symptom severity pathways of the dimensional approach. The 'intermittently recurrent' trajectory and the "partially recovered" pathway corresponded relatively well (32.9% versus 25.7%). However, the percentages of the '(consistently) chronic' and '(consistently) recovered' trajectories did not align. In the categorical approach compared to the dimensional approach, chronicity was lower (8.5% versus 66.8%), and (combined) recovery was higher (58.6% versus 7.5%). The 'recovered' dimensional pathway appeared to correspond more to the 'consistently recovered' diagnosis trajectory (10.1%) at nine-year follow-up.

We then compared the primary and comparison DSM-diagnosis trajectories and three identified symptom severity pathways using crosstabulation at nine-year follow-up (see **Table 3**).

Table 3.
Comparison of the Three Identified Dimensional Symptom Severity Pathways of the Latent Class Growth Analysis (LCGA) with the Three and Four Categorical Diagnosis Trajectories based on Overlap and Disability at Nine-Year Follow-Up

Categorical diagnoses trajectories[a]	Dimensional symptom severity pathways or LCGA classes		
	Recovered	Partially Recovered	Chronic
Consistently Recovered	33	92	78
% within categorical trajectories	16.3%	45.3%	38.4%
% within LCGA classes	52.4%	29.7%	13.2%
Intermittently Recovered	25	135	203
% within categorical trajectories	6.9%	37.2%	55.9%
% within LCGA classes	39.7%	43.5%	34.2%
Intermittently Recurrent	5	75	238
% within categorical trajectories	1.6%	23.6%	74.8%
% within LCGA classes	7.9%	24.2%	40.1%
Consistently Chronic	0	8	74
% within categorical trajectories	0.0%	9.8%	90.2%
% within LCGA classes	0.0%	2.6%	12.5%
Categorical diagnoses trajectories[b]	Recovered	Partially Recovered	Chronic
Recovered	**33**	92	78
% within LCGA classes	40.7%	24.3%	9.0%
% within categorical trajectories	16.3%	45.3%	38.4%
Recurrent	48	**278**	715
% within LCGA classes	59.3%	73.5%	82.5%
% within categorical trajectories	4.6%	26.7%	68.7%
Chronic	0	8	**74**
% within LCGA classes	0.0%	2.1%	8.5%
% within categorical trajectories	0.0%	9.8%	90.2%
Disability (WHO-DAS-II), mean (*SD*)	5.3 (7.9)	13.9 (12.6)	28.8 (17.6)

[a]4 primary trajectories of the categorical approach, *N* = 966.
[b]3 comparison/aggregated trajectories of the categorical approach, *N* = 1326.
In **bold** = overlap between the aggregated categorical trajectories and dimensional pathways.
WHO-DAS-II = World Health Organization disability assessment schedule.

First, we looked at the four primary/variable DSM-diagnosis trajectories. Participants who were 'consistently recovered,' 'intermittently recovered,' and 'intermittently recurrent' from a DSM diagnosis were spread across all three symptom severity pathways. No clear pattern of correspondence emerged. Second, we looked at the three comparison/aggregated DSM-diagnosis trajectories with the symptom severity pathways, and the overlap of was low, with very poor agreement, $\kappa = .02$. Participants from the categorical 'recovered' trajectory, for example, were also allocated to the 'partially recovered' and 'chronic' dimensional pathways. Regarding functional disability, participants in the 'chronic' symptom severity pathway experienced the highest functional disability, followed by the 'partial recovery', then 'recovered' pathway.

Finally, we tested whether there was a potential effect of age in quartiles and gender on depressive, anxiety, fear and worry severity symptoms (see **Figure C** in **Supplemental Material**). There were age group and gender differences, with age having a slightly higher effect on severity symptoms. **Figure 3** presents the trends of depressive (A), anxiety (B), fear (C), and worry (D) severity symptoms for subsamples of participants with depression only, anxiety only, or comorbid depression-anxiety, based on baseline severity scores split into quartiles. These were adjusted for the effects of age and gender. However, the adjusted mixed model regression analyses resulted in similar trajectories as the non-adjusted analyses.

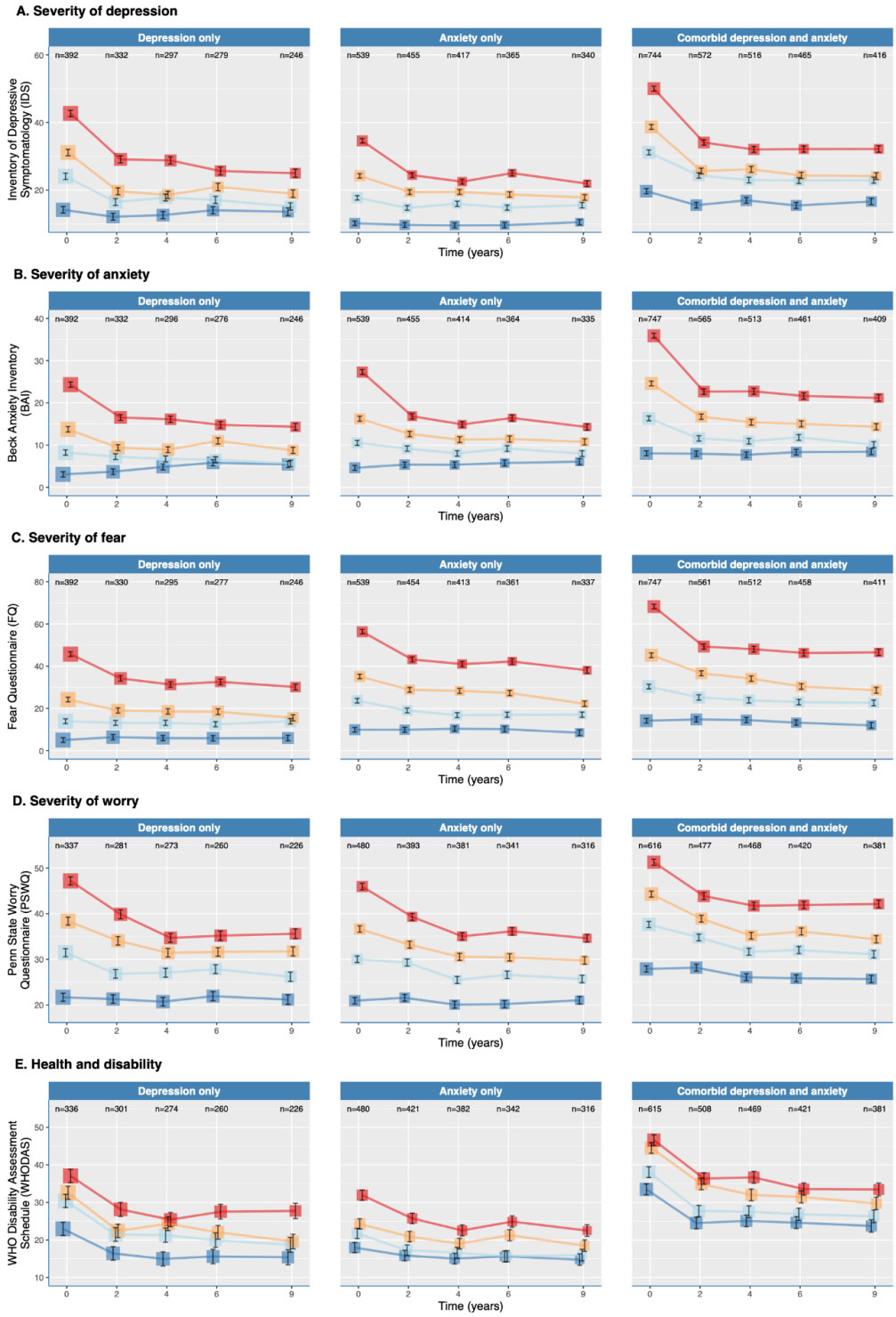

Figure 3. Trajectories of depressive, anxiety, fear, and worry symptoms, controlled for the effect of age and gender. Error bars represent standard errors, and the box size is proportional to the number of participants.

Throughout the nine-year period, those with the highest baseline severity levels (red line) on average maintained the highest severity scores, remaining above clinically-relevant threshold levels. In general, all baseline quartiles had trends that ran rather parallel over time and remained stable after an initial drop in severity at two-year follow-up. Furthermore, comparing the three subsamples, those with comorbid depression-anxiety at baseline appeared to have the highest scores across all symptom measures compared to the depression only and anxiety only groups.

Finally, to assess the trends of functional disability, we reviewed the WHO-DAS-II scores at each follow-up assessment over the nine-year period for subgroups of depression only, anxiety only, and comorbid depression-anxiety based on baseline total scores split into quartiles (see **Figure 2E**). Those with comorbid depression-anxiety appeared to have higher disability than those with depression only or anxiety only.

4. Discussion

Our findings show that using DSM diagnoses alone fails to fully capture the persistence of depressive and anxiety symptoms over time. The categorical trajectories using diagnoses were not congruent with results of the identified dimensional pathways using symptom severity, further emphasizing the need to better capture clinical course (Batelaan et al., 2014). In other words, diagnostic recovery was not equivalent to symptomatic recovery. When using the categorical approach with any depressive or anxiety diagnosis, approximately 8% were consistently chronic, 33% had intermittently relapsed, 38% had intermittently recovered, and 21% remained consistently recovered at nine-year follow-up, showing a higher estimation of recovery compared to the dimensional

approach. However, when we examined participants with baseline comorbid depression and anxiety, only 10.1% remained consistently recovered at nine-year follow-up, which paralleled the 7.5% who were recovered using the dimensional approach. Additionally, compared to the categorical trajectories of comorbid depression and anxiety, the chronicity of the dimensional symptom severity pathway was much higher (11.8% versus 66.8%). When using aggregated trajectories, we found that 78.5% of participants with any disorder experienced recurrence over a nine-year period. This was much higher than the 25.7% who were partially recovered when using the dimensional approach.

The discrepancy between the trajectories could be explained by the following reasons.

First, the discrepancy may be due to the inherent differences of the methods used to assess the dimensional versus the categorical outcomes. Specifically, the dimensional approach used self-report questionnaires to determine the change in symptom severity over time while the categorical approach used a structured interview to establish DSM diagnoses and examine them over time. Although the self-report questionnaires encompassed the same symptoms as the depression and anxiety syndromes in the structured interview, the questionnaires assessed a shorter duration period of one to two weeks. For the diagnoses, we allowed a longer recency period of six months. The diagnosis, therefore, could no longer be present and might have been present up to five and a half months ago. Moreover, in the structured interview, the trained interviewer examined the number of symptoms, duration of the symptoms, and the patient's level of functioning. Thus, if a patient had the required number of symptoms yet a higher level of functioning, a DSM diagnosis was not recorded. At the outset, due to the specific diagnostic criteria, using a categorical

approach with DSM diagnoses may be less inclusive when compared to the dimensional approach, which may have resulted in higher recovery rates and lower chronicity. Because of the inclusion of the functional disability criterion, we consider DSM diagnoses to be a better indicator of psychological suffering than self-report questionnaires. Second, the discrepancy in results could be due to enduring, fluctuating subthreshold affective symptoms and the higher prevalence of individuals experiencing recurrence of diagnoses. Those with remitted affective diagnoses, for instance, may still suffer from subthreshold disease (e.g., minor depression), or they could have been assessed during a short period in which affective symptoms had slightly receded (Batelaan et al., 2014; Torpey and Klein, 2008). Previous research has shown that, compared to those without symptoms, those with subthreshold symptoms relapsed sooner (Cuijpers and Smit, 2004; Fawcett, 1994; Judd et al., 1998; Karsten et al., 2011) and were found to suffer more chronic episodes and fewer symptom-free weeks (Arnow and Constantino, 2003; Judd and Akiskal, 2000). Other studies have also found a recurrent course with (non)chronic episodes in at least 50% of participants (Bobes et al., 2018; Bruce et al., 2005; Scholten et al., 2016; Scholten et al., 2013; Verduijn et al., 2017; Yonkers et al., 2003) and up to 80% (Judd, 1997).

Furthermore, when using the categorical approach, as expected, those with comorbid depressive and anxiety disorders had a relatively poorer long-term course when compared to those baseline depressive disorders only. This finding is in line with previous research examining anxiety and (comorbid) depressive disorders (Batelaan et al., 2014; Bruce et al., 2005; Rhebergen et al., 2011; ter Meulen et al., 2021). According to Batelaan et al. (2014),

however, baseline severity, duration of symptoms, and disability appear to be better indicators of a poor prognosis than DSM-categories.

Clinical significance of the results

Kraepelinian nosology (Kraepelin, 1921) emphasizes that course, functional outcome, and etiology support a dimensional approach of depression and anxiety. In light of our findings, the question arises whether we should view individuals with comorbidity as having two or more distinct DSM disorders or as having a single dimensional disorder, or symptoms on a continuum, in which myriad etiological factors may result in diverse syndromes that are modified with time and environmental exposures (Angst and Wicki, 1991; Cardno et al., 2002; Hyman, 2010; Kendler et al., 1992; Krueger and Markon, 2006). Previous studies (Fried and Nesse, 2015b; Fried et al., 2014; Lux and Kendler, 2010; van Eeden et al., 2019) have identified risk factors, such as neuroticism and baseline chronicity, that affected individual depressive symptoms to varying degrees, suggesting that depression is not one unified latent construct. On the other hand, abandoning the current DSM criteria would reduce reliability and leave psychiatry without a common language (Kendell and Jablensky, 2003; Patten, 2015). Therefore, until a better dimensional model is proposed, we suggest the integrative use of diagnostic classification, symptom severity, symptom duration, and levels of impairment (Batelaan et al., 2014; Hyman, 2010; Patten, 2015). An example is the Activity, Cognition, and Emotion (ACE) model that groups symptoms commonly present in mood disorders like depression and bipolar disorder according to functional domains (Malhi et al., 2018). Another option is the symptom-based framework which looks at individual symptoms and how they influence each other (Fried, 2015). However, more research is

needed to confirm whether a dimensional symptom-oriented tool will lead to more accurate symptom-specific tailored treatment plans and better outcomes. Moreover, clinicians may consider the adoption of longer-term treatment strategies (Vos et al., 2004) that may include relapse prevention strategies and psychiatric rehabilitation (i.e., functional recovery) for those with high levels of subthreshold symptoms and a high risk of recurrence of disorders over nine years. Well-designed interventions studies looking into the types and effectiveness of relapse prevention strategies are warranted.

Strengths and Limitations

This study has some noteworthy strengths. First, NESDA was designed to gain more insight into the long-term clinical course of depression and anxiety in a large cohort study with access to the full range of depression and anxiety disorders from primary and secondary care. Patients were rigorously diagnosed using diagnostic interviews and followed up over a nine-year time period with multiple assessment points. Moreover, rather than investigating the course of single MDD or anxiety disorders, we examined variable course categories that allowed comorbidity and diagnostic switching, which may be more ecologically and clinically relevant. To strengthen our conclusions, we used the data-driven LCGA method, to distinguish groups with similar symptom severity trajectories, which handles missing data rather well. The findings were further confirmed by inspecting symptom severity scores, diagnoses, and functional (dis)ability. Additionally, we compared our models to those without missing data, and results were not significantly altered.

With regard to our study limitations and research recommendations, the following may be mentioned. First, the current study used data from NESDA, which concerns DSM-IV criteria rather than the newest DSM-5 criteria. In the DSM-5, there is a duration criterion for social phobia and agoraphobia, which could be useful in excluding patients with only transient fears (Batelaan et al., 2014). Second, NESDA was set up as an observational, naturalistic study, and therefore, the effects of treatment on outcomes could not be examined (e.g., due to confounding by indication effects). While general data were collected regarding medication use, treatment duration, and setting (e.g., primary or secondary care, or psychotropic use), data on the specifics of type, duration, and intensity of psychotherapeutic interventions were not collected. Third, to maximize our sample, we selected participants with a diagnosis in the six months prior to baseline, a period during which some participants may have remitted. Fourth, NESDA did not include participants with obsessive compulsive disorder (OCD), bipolar disorders, or other common mental disorders, which participants may also be experiencing, thereby potentially increasing prevalence rates of chronicity. Fifth, our sample included participants recruited from the community, primary care and specialized mental health care, resulting in a heterogenous cohort (Patten, 2015; Regier et al., 1998). Thus, our results may not be generalizable to each population. Sixth, at each follow-up there was a selective loss of participants, which were those with the highest baseline severity levels. Thus, chronicity may have been underestimated. Seventh, the Life Chart Interview (LCI), which relies on the recall of past memories, was used to establish the continuity of symptoms between assessment points in the categorical approach. Recall error of autobiographical information may have had an impact on the accuracy of the sustained recovered or consistently chronic

categorical trajectories (Drasch and Matthes, 2013; Reimer and Matthes, 2007). Although the LCI uses an event history calendar with landmark events, which has been shown to be an effective tool for the reliable retrieval of older memories (Belli, 1998; Belli et al., 2001; Drasch and Matthes, 2013; Vaart and Glasner, 2011), memory is not always reliable, and any results using retrospective recall should be carefully considered (Glasner and Vaart, 2009; Patten et al., 2010). Eighth, NESDA used the CIDI to diagnose DSM-IV disorders. However, there is an alternative semi-structured clinical interview, namely the Structured Psychopathological Interview and Rating of the Social Consequences for Epidemiology (SPIKE) (Angst and Dobler-Mikola, 1985; Angst et al., 2009; Angst and Wicki, 1991; Ernst and Angst, 1992), which has a lower threshold than the CIDI to meet diagnostic criteria. The strict criteria of the CIDI may make it less inclusive, potentially leading to an underestimation of chronicity levels in the current study. Ninth, LCGA classes were calculated primarily with log-linear functions and with too few assessments. Therefore, our dimensional approach may not have been able to fully describe the "waxing and waning" of symptom severity. However, examining alternative models with other shape trajectories corresponded highly with the log-linear primary model. Tenth, while we did examine the functional disability in the dimensional approach using mixed model regression analyses, functional disability was not included in the main LCGA due to bias. Future research using a longitudinal design and a dimensional approach that includes both rater-based versus self-rated symptom scales would be of value and may help improve the prediction of the clinical course.

5. Conclusion

This study showed that, when taking comorbidity, switching between diagnoses, and symptom duration into consideration to examine the long-term course of depression and anxiety, using categorical diagnoses alone to describe clinical course led to a high estimation of recovery, low estimation of chronicity, and appeared to inadequately capture the persistence of affective symptoms. Discrepancies in the clinical course using DSM-categories or symptom severity could be explained by the enduring, fluctuating presence of subthreshold affective symptoms that do not meet diagnosis criteria but still may lead to recurrence of disorders. The commonness of subthreshold symptoms and their adverse impact on long-term prognoses deserve continuous clinical attention in mental health care as well further research.

References in Special Issue

Akaike, H., 1987. Factor analysis and aic, Selected papers of hirotugu akaike. Springer, pp. 371-386.
American Psychiatric Association, 2013. Diagnostic and statistical manual of mental disorders, dsm-5, 5th ed. ed. Washington, DC [etc.] : American Psychiatric Publishing.
Andrews, G., Peters, L., 1998. The psychometric properties of the composite international diagnostic interview. Social psychiatry and psychiatric epidemiology 33, 80-88.
Andruff, H., Carraro, N., Thompson, A., Gaudreau, P., Louvet, B., 2009. Latent class growth modelling: A tutorial. Tutorials in quantitative methods for psychology 5, 11-24.
Angst, J., Dobler-Mikola, A., 1985. The zurich study--a prospective epidemiological study of depressive, neurotic and psychosomatic syndromes. Iv. Recurrent and nonrecurrent brief depression. European Archives of Psychiatry and Clinical Neuroscience 234, 408-416.
Angst, J., Gamma, A., Rössler, W., Ajdacic, V., Klein, D.N., 2009. Long-term depression versus episodic major depression: Results from the prospective zurich study of a community sample. Journal of Affective Disorders 115, 112-121.
Angst, J., Wicki, W., 1991. The zurich study. Xi. Is dysthymia a separate form of depression? Results of the zurich cohort study. European Archives of Psychiatry and Clinical Neuroscience 240, 349-354.
Armour, C., Shevlin, M., Elklit, A., Mroczek, D., 2012. A latent growth mixture modeling approach to ptsd symptoms in rape victims. Traumatology (Tallahass Fla) 18, 20-28.
Arnow, B.A., Constantino, M.J., 2003. Effectiveness of psychotherapy and combination treatment for chronic depression. Journal of clinical psychology 59, 893-905.
Asparouhov, T., 2006. Growth mixture analysis: Models with non-gaussian random effects. G.
Batelaan, N.M., Rhebergen, D., Spinhoven, P., van Balkom, A.J., Penninx, B.W.J.H., 2014. Two-year course trajectories of anxiety disorders: Do dsm classifications matter? The Journal of clinical psychiatry 75, 985.
Beasley, T.M., Schumacker, R.E., 1995. Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. The Journal of experimental education 64, 79-93.
Beck, A., Epstein, N., Brown, G., Steer, R., 1988. An inventory for measuring clinical anxiety: Psychometric properties. Journal of Consulting and Clinical Psychology 56, 893-897.
Belli, R.F., 1998. The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. Memory 6, 383-406.
Belli, R.F., Shay, W.L., Stafford, F.P., 2001. Event history calendars and question list surveys: A direct comparison of interviewing methods. Public Opinion Quarterly 65, 45-74.
Berlin, K.S., Parra, G.R., Williams, N.A., 2014. An introduction to latent variable mixture modeling (part 2): Longitudinal latent class growth analysis and growth mixture models. Journal of Pediatric Psychology 39, 188-203.
Biernacki, C., Celeux, G., Govaert, G., 2003. Choosing starting values for the em algorithm for getting the highest likelihood in multivariate gaussian mixture models. Computational statistics & data analysis 41, 561-575.
Bobes, J., Saiz-Ruiz, J., Pérez, V., 2018. Barriers to complete recovery of major depression: Cross-sectional, multi-centre study on clinical practice. Record study. Journal of Psychiatry and Mental Health (Revista de Psiquiatría y Salud Mental, English Edition) 12, 141-150.
Bokma, W.A., Batelaan, N.M., van Balkom, A.J.L.M., Penninx, B.W.J.H., 2017. Impact of anxiety and/or depressive disorders and chronic somatic diseases on disability and work impairment. Journal of Psychosomatic Research 94, 10-16.
Brown, T.A., Antony, M.M., Barlow, D.H., 1992. Psychometric properties of the penn state worry questionnaire in a clinical anxiety disorders sample. Behaviour Research and Therapy 30, 33-37.
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. American Journal of Psychiatry 162, 1179-1187.
Brunoni, A.R., 2017. Beyond the dsm: Trends in psychiatry diagnoses. Revista de psiquiatria clínica 44, 154-158.
Cardno, A.G., Rijsdijk, F.V., Sham, P.C., Murray, R.M., McGuffin, P., 2002. A twin study of genetic relationships between psychotic symptoms. American Journal of Psychiatry 159, 539-545.
Caspi, A., Houts, R.M., Ambler, A., al., e., 2020. Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. JAMA Network Open 3, e203221.
Chwastiak, L.A., Von Korff, M., 2003. Disability in depression and back pain: Evaluation of the world health organization disability assessment schedule (who das ii) in a primary care setting. Journal of Clinical Epidemiology 56, 507-514.
Cicchetti, D., Rogosch, F.A., 1999. Conceptual and methodological issues in developmental psychopathology research, In: Kendall, P.C., Butcher, J.N., Holmbeck, G.N. (Eds.), Handbook of research methods in clinical psychology. Wiley, New York, NY, pp. 433-465.
Cohen, J., 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement 20, 37-46.
Cuijpers, P., Smit, F., 2004. Subthreshold depression as a risk indicator for major depressive disorder: A systematic review of prospective studies. Acta Psychiatrica Scandinavica 109, 325-331.
Curran, P.J., Hussong, A.M., 2003. The use of latent trajectory models in psychopathology research. Journal of Abnormal Psychology 112, 526-544.
Davey, G.C.L., 1993. A comparison of three worry questionnaires. Behaviour Research and Therapy 31, 51-56.
De Ayala, R.J., Vonderharr-Carlson, D.J., Kim, D., 2005. Assessing the reliability of the beck anxiety inventory scores. Educational and Psychological Measurement 65, 742-756.
de Beurs, E., Den Hollander-Gijsman, M., van Rood, Y., Giltay, E., van Noorden, M., van Der Lem, R., van Fenema, E., Zitman, F., 2011. Routine outcome monitoring in the netherlands: Practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. Clinical Psychology & Psychotherapy 18, 1-12.
de Raadt, A., Warrens, M.J., Bosker, R.J., Kiers, H.A.L., 2019. Kappa coefficients for missing data. Educational and Psychological Measurement 79, 558-576.
deRoon-Cassini, T.A., Mancini, A.D., Rusch, M.D., Bonanno, G.A., 2010. Psychopathology and resilience following traumatic injury: A latent growth mixture model analysis. Rehabilitation Psychology 55, 1-11.
DiStefano, C., Kamphaus, R.W., 2006. Investigating subtypes of child development: A comparison of cluster analysis and latent class cluster analysis in typology creation. Educational and Psychological Measurement 66, 778-794.
Drasch, K., Matthes, B., 2013. Improving retrospective life course data by combining modularized self-reports and event history calendars: Experiences from a large scale survey. Quality & quantity 47, 817-838.
Ernst, C., Angst, J., 1992. The zurich study. Sex-differences in depression - evidence from longitudinal epidemiologic data. European archives of psychiatry and clinical neuroscience 241, 222-230.
Fawcett, J., 1994. Antidepressants: Partial response in chronic depression. The British Journal of Psychiatry 165, 37-41.
Fried, E.I., 2015. Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. Frontiers in Psychology 6, 309-309.
Fried, E.I., Nesse, R.M., 2015a. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the stard study. Journal of Affective Disorders 172, 96-102.
Fried, E.I., Nesse, R.M., 2015b. Depression sum-scores don't add up: Why analyzing specific depression symptoms is essential. BMC Medicine 13, 72.
Fried, E.I., Nesse, R.M., Zivin, K., Guille, C., Sen, S., 2014. Depression is more than the sum score of its parts: Individual dsm symptoms have different risk factors. Psychological Medicine 44, 2067-2076.

Fydrich, T., Dowdall, D., Chambless, D.L., 1992. Reliability and validity of the beck anxiety inventory. Journal of anxiety disorders 6, 55-61.
García-Pérez, M.A., Núñez-Antón, V., 2003. Cellwise residual analysis in two-way contingency tables. Educational and psychological measurement 63, 825-839.
Giandinoto, J.-A., Edward, K.-l., 2015. The phenomenon of co-morbid physical and mental illness in acute medical care: The lived experience of australian health professionals. BMC Research Notes 8, 295.
Gilchrist, G., Gunn, J., 2007. Observational studies of depression in primary care: What do we know? BMC Family Practice 8, 28-28.
Gillis, M.M., Haaga, D.A., Ford, G.T., 1995. Normative values for the beck anxiety inventory, fear questionnaire, penn state worry questionnaire, and social phobia and anxiety inventory. Psychological Assessment 7, 450.
Glasner, T.J., Vaart, v.d.W., 2009. Applications of calendar instruments in social surveys: A review. Quality & quantity 43, 333-349.
Gregory, A.M., Caspi, A., Moffitt, T.E., Koenen, K., Eley, T.C., Poulton, R., 2007. Juvenile mental health histories of adults with anxiety disorders. American Journal of Psychiatry 164, 301-308.
Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W.A., Beekman, A.T.F., 2010. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta psychiatrica scandinavica 122, 184-191.
Haro, J.M., Arbabzadeh-Bouchez, S., Brugha, T.S., De Girolamo, G., Guyer, M.E., Jin, R., Lepine, J.P., Mazzi, F., Reneses, B., Vilagut, G., Sampson, N.A., Kessler, R.C., 2006. Concordance of the composite international diagnostic interview version 3.0 (cidi 3.0) with standardized clinical assessments in the who world mental health surveys. International Journal of Methods in Psychiatric Research 15, 167-180.
Hayden, E.P., Klein, D.N., 2001. Outcome of dysthymic disorder at 5-year follow-up: The effect of familial psychopathology, early adversity, personality, comorbidity, and chronic stress. American Journal of Psychiatry 158, 1864-1870.
Hendriks, S.M., Spijker, J., Licht, C.M.M., Beekman, A.T.F., Penninx, B.W.J.H., 2013. Two-year course of anxiety disorders: Different across disorders or dimensions? Acta psychiatrica scandinavica 128, 212-221.
Hovenkamp-Hermelink, J.H., Riese, H., 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 nesda study. Journal of affective disorders 190, 310-315.
Howland, R.H., John Rush, A., Wisniewski, S.R., Trivedi, M.H., Warden, D., Fava, M., Davis, L.L., Balasubramani, G.K., McGrath, P.J., Berman, S.R., 2009. Concurrent anxiety and substance use disorders among outpatients with major depression: Clinical features and effect on treatment outcome. Drug and Alcohol Dependence 99, 248-260.
Hyman, S.E., 2010. The diagnosis of mental disorders: The problem of reification. Annu Rev Clin Psychol 6, 155-179.
Judd, L.L., 1997. The clinical course of unipolar major depressive disorders. Archives of general psychiatry 54, 989-991.
Judd, L.L., Akiskal, H.S., 2000. Delineating the longitudinal structure of depressive illness: Beyond clinical subtypes and duration thresholds. Pharmacopsychiatry 33, 3-7.
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. Major depressive disorder: A prospective study of residual subthreshold depressive symptoms as predictor of rapid relapse. Journal of Affective Disorders 50, 97-108.
Jung, T., Wickrama, K.A.S., 2008. An introduction to latent class growth analysis and growth mixture modeling. Social and personality psychology compass 2, 302-317.
Karsten, J., Hartman, C.A., Smit, J.H., Zitman, F.G., Beekman, A.T.F., Cuijpers, P., van Der Does, A.J.W., Ormel, J., Nolen, W.A., Penninx, B.W.J.H., 2011. Psychiatric history and subthreshold symptoms as predictors of the occurrence of depressive or anxiety disorder within 2 years. The British journal of psychiatry : the journal of mental science 198, 206.
Kendell, R., Jablensky, A., 2003. Distinguishing between the validity and utility of psychiatric diagnoses. American Journal of Psychiatry 160, 4-12.

Kendler, K.S., Gardner, C.O., 1998. Boundaries of major depression: An evaluation of dsm-iv criteria. American Journal of Psychiatry 155, 172-177.
Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., Eaves, L.J., 1992. Major depression and generalized anxiety disorder: Same genes, (partly) different environments? Archives of general psychiatry 49, 716-722.
Kessler, R.C., Chui, W.T., Demler, O., Walters, E.E., 2005. Prevalence, severity, and comorbidity of 12-month dsm-iv disorders in the national comorbidity survey replication. Arch Gen Psychiatry 62, 617-627.
Kleiboer, A., Smit, J., Bosmans, J., Ruwaard, J., Andersson, G., Topooco, N., Berger, T., Krieger, T., Botella, C., Baños, R., Chevreul, K., Araya, R., Cerga-Pashoja, A., Cieślak, R., Rogala, A., Vis, C., Draisma, S., van Schaik, A., Kemmeren, L., Ebert, D., Berking, M., Funk, B., Cuijpers, P., Riper, H., 2016. European comparative effectiveness research on blended depression treatment versus treatment-as-usual (e-compared): Study protocol for a randomized controlled, non-inferiority trial in eight european countries. Trials 17, 1-10.
Klein, D.N., Shankman, S.A., Rose, S., 2006. Ten-year prospective follow-up study of the naturalistic course of dysthymic disorder and double depression. American Journal of Psychiatry 163, 872-880.
Klein, D.N., Shankman, S.A., Rose, S., 2008. Dysthymic disorder and double depression: Prediction of 10-year course trajectories and outcomes. Journal of Psychiatric Research 42, 408-415.
Kraepelin, E., 1921. Manic depressive insanity and paranoia. The journal of nervous and mental disease 53, 350.
Krueger, R.F., Markon, K.E., 2006. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. Annual Review of Clinical Psychology 2, 111-133.
Lamers, F., van Oppen, P., Comijs, H.C., Smit, J.H., Spinhoven, P., van Balkom, A.J.L.M., Nolen, W.A., Zitman, F.G., Beekman, A.T.F., Penninx, B.W.J.H., 2011. Comorbidity patterns of anxiety and depressive disorders in a large cohort study: The netherlands study of depression and anxiety (nesda). The Journal of clinical psychiatry 72, 341-348.
Lee, H.-c.B., Oei, T.P.S., 1994. Factor structure, validity, and reliability of the fear questionnaire in a hong kong chinese population. Journal of psychopathology and behavioral assessment 16, 189-199.
Lorenzo-Luaces, L., DeRubeis, R.J., van Straten, A., Tiemens, B., 2017. A prognostic index (pi) as a moderator of outcomes in the treatment of depression: A proof of concept combining multiple variables to inform risk-stratified stepped care models. Journal of Affective Disorders 213, 78-85.
Lux, V., Kendler, K.S., 2010. Deconstructing major depression: A validation study of the dsm-iv symptomatic criteria. Psychological Medicine 40, 1679-1690.
Lyketsos, C.G., Nestadt, G., Cwi, J., Heithoff, K., 1994. The life chart interview: A standardized method to describe the course of psychopathology. International Journal of Methods in Psychiatric Research.
Maj, M., Pirozzi, R., Magliano, L., Bartoli, L., 2002. The prognostic significance of "switching" in patients with bipolar disorder: A 10-year prospective follow-up study. American Journal of Psychiatry 159, 1711-1717.
Malhi, G.S., Irwin, L., Hamilton, A., Morris, G., Boyce, P., Mulder, R., Porter, R.J., 2018. Modelling mood disorders: An ace solution? Bipolar Disorders 20, 4-16.
Marks, I.M., Mathews, A.M., 1979. Brief standard self-rating for phobic patients. Behaviour research and therapy 17, 263-267.
McGorry, P.D., Hickie, I.B., Yung, A.R., Pantelis, C., Jackson, H.J., 2016. Clinical staging of psychiatric disorders: A heuristic framework for choosing earlier, safer and more effective interventions. Australian and New Zealand Journal of Psychiatry 40, 616-622.
McHugh, M.L., 2012. Interrater reliability: The kappa statistic. Biochemia Medica (Zagreb) 22, 276-282.
Meyer, T.J., Miller, M.L., Metzger, R.L., Borkovec, T.D., 1990. Development and validation of the penn state worry questionnaire. Behaviour research and therapy 28, 487-495.

Moffitt, T.E., Caspi, A., Harrington, H., Milne, B.J., Melchior, M., Goldberg, D., Poulton, R., 2007. Generalized anxiety disorder and depression: Childhood risk factors in a birth cohort followed to age 32. Psychological Medicine 37, 441-452.
Mueller, T., Leon, A., Keller, M., Solomon, D., Endicott, J., Coryell, W., Warshaw, M., Maser, J., 1999. Recurrence after recovery from major depressive disorder during 15 years of observational follow-up. American Journal of Psychiatry 156, 1000.
Muntingh, A.D.T., van der Feltz-Cornelis, C.M., van Marwijk, H.W.J., Spinhoven, P., Penninx, B.W.J.H., van Balkom, A.J.L.M., 2011. Is the beck anxiety inventory a good tool to assess the severity of anxiety? A primary care study in the netherlands study of depression and anxiety (nesda). BMC Family Practice 12, 66-66.
Muthén, B., 2003. Statistical and substantive checking in growth mixture modeling: Comment on bauer and curran (2003). Psychological Methods 8, 369-377.
Muthén, B., Muthén, L.K., 2000. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. Alcoholism: Clinical and experimental research 24, 882-891.
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. Structural equation modeling 14, 535-569.
Oei, T.P., Moylan, A., Evans, L., 1991. Validity and clinical utility of the fear questionnaire for anxiety-disorder patients. Psychological Assessment: A Journal of Consulting and Clinical Psychology 3, 391.
Ormel, J., Oldehinkel, T., Brilman, E., van den Brink, W., 1993. Outcome of depression and anxiety in primary care. A three-wave 3 1/2-year study of psychopathology and disability. Archives of general psychiatry 50, 759-766.
Patten, S.B., 2015. Major depressive disorder: Reification and (maybe) rheostasis. Epidemiology and Psychiatric Sciences 24, 473-475.
Patten, S.B., Gordon-Brown, L., Meadows, G., 2010. Simulation studies of age-specific lifetime major depression prevalence. BMC Psychiatry 10, 85-85.
Penninx, B.W.J.H., Beekman, A.T.F., Smit, J.H., Zitman, F.G., Nolen, W.A., Spinhoven, P., Cuijpers, P., De Jong, P.J., Van Marwijk, H.W.J., Assendelft, W.J.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, objectives and methods. International Journal of Methods in Psychiatric Research 17, 121-140.
Penninx, B.W.J.H., Eikelenboom, M., Giltay, E.J., van Hemert, A.M., Riese, H., Schoevers, R.A., Beekman, A.T.F., 2021. Cohort profile of the longitudinal netherlands study of depression and anxiety (nesda) on etiology, course and consequences of depressive and anxiety disorders. Journal of Affective Disorders 287, 69-77.
Penninx, B.W.J.H., Nolen, W.A., Lamers, F., Zitman, F.G., Smit, J.H., Spinhoven, P., Cuijpers, P., de Jong, P.J., van Marwijk, H.W.J., Der Meer, K.v., Verhaak, P., Laurant, M.G.H., de Graaf, R., Hoogendijk, W.J., Der Wee, N.v., Ormel, J., van Dyck, R., Beekman, A.T.F., 2011. Two-year course of depressive and anxiety disorders: Results from the netherlands study of depression and anxiety (nesda). Journal of Affective Disorders 133, 76-85.
Plana-Ripoll, O., Pedersen, C.B., Holtz, Y., Benros, M.E., Dalsgaard, S., de Jonge, P., Fan, C.C., Degenhardt, L., Ganna, A., Greve, A.N., Gunn, J., Iburg, K.M., Kessing, L.V., Lee, B.K., Lim, C.C.W., Mors, O., Nordentoft, M., Prior, A., Roest, A.M., Saha, S., Schork, A., Scott, J.G., Scott, K.M., Stedman, T., Sørensen, H.J., Werge, T., Whiteford, H.A., Laursen, T.M., Agerbo, E., Kessler, R.C., Mortensen, P.B., McGrath, J.J., 2019. Exploring comorbidity within mental disorders among a danish national population. JAMA Psychiatry 76, 259-270.
Proust-Lima, C., Philipps, V., Liquet, B., 2017. Estimation of extended mixed models using latent classes and latent processes: The r package lcmm. Journal of statistical software 78, 1-56.
Ramaswamy, V., Desarbo, W.S., Reibstein, D.J., Robinson, W.T., 1993. An empirical pooling approach for estimating marketing mix elasticities with pims data. Marketing science (Providence, R.I.) 12, 103-124.

Regier, D.A., Kaelber, C.T., Rae, D.S., Farmer, M.E., Knauper, B., Kessler, R.C., Norquist, G.S., 1998. Limitations of diagnostic criteria and assessment instruments for mental disorders: Implications for research and policy. Archives of General Psychiatry 55, 109-115.
Reimer, M., Matthes, B., 2007. Collecting event histories with truetales:Techniques to improve autobiographical recall problems in standardized interviews. Quality & quantity 41, 711-735.
Rhebergen, D., Batelaan, N.M., De Graaf, R., Nolen, W.A., Spijker, J., Beekman, A.T.F., Penninx, B.W.J.H., 2011. The 7-year course of depression and anxiety in the general population. Acta Psychiatrica Scandinavica 123, 297-306.
Rhebergen, D., Lamers, F., Spijker, J., de Graaf, R., Beekman, A.T.F., Penninx, B.W.J.H., 2012. Course trajectories of unipolar depressive disorders identified by latent class growth analysis. Psychological Medicine 42, 1383-1396.
Richards, D., 2011. Prevalence and clinical course of depression: A review. Clinical Psychology Review 31, 1117-1125.
Rizvi, S.J., Cyriac, A., Grima, E., Tan, M., Lin, P., Gallaugher, L.A., McIntyre, R.S., Kennedy, S.H., 2015. Depression and employment status in primary and tertiary care settings. The Canadian Journal of Psychiatry 60, 14-22.
Rush, A., Trivedi, M., Wisniewski, S., Nierenberg, A., Stewart, J., Warden, D., Niederehe, G., Thase, M., Lavori, P., Lebowitz, B., McGrath, P., Rosenbaum, J., Sackeim, H., Kupfer, D., Luther, J., Fava, M., 2006. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A star*d report. American Journal of Psychiatry 163, 1905-1917.
Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The inventory of depressive symptomatology (ids): Psychometric properties. Psychological Medicine 26, 477-486.
Salarifar, M., Pouretemad, H., 2012. The study of factorial stucture, validity, and reliability of the penn state worry questionnaire (pswq). European Psychiatry 27.
Scholten, W.D., Batelaan, N.M., Penninx, B.W., van Balkom, A.J., Smit, J.H., Schoevers, R.A., van Oppen, P., 2016. Diagnostic instability of recurrence and the impact on recurrence rates in depressive and anxiety disorders. Journal of Affective Disorders 195, 185-190.
Scholten, W.D., Batelaan, N.M., van Balkom, A.J., Wjh. Penninx, B., Smit, J.H., van Oppen, P., 2013. Recurrence of anxiety disorders and its predictors. Journal of Affective Disorders 147, 180-185.
Schwarz, G., 1978. Estimating the dimension of a model. The annals of statistics 6, 461-464.
Sclove, S.L., 1987. Application of model-selection criteria to some problems in multivariate analysis. Psychometrika 52, 333-343.
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. Journal of Anxiety Disorders 44, 92-101.
Steer, R.A., Ranieri, W.F., Beck, A.T., Clark, D.A., 1993. Further evidence for the validity of the beck anxiety inventory with psychiatric outpatients. Journal of anxiety disorders 7, 195-205.
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. Journal of Affective Disorders 152-154, 65-75.
ter Meulen, W.G., Draisma, S., van Hemert, A.M., Schoevers, R.A., Kupka, R.W., Beekman, A.T.F., Penninx, B.W.J.H., 2021. Depressive and anxiety disorders in concert–a synthesis of findings on comorbidity in the nesda study. Journal of Affective Disorders 284, 85-97.
Torpey, D., Klein, D., 2008. Chronic depression: Update on classification and treatment. Current Psychiatry Reports 10, 458-464.
Trivedi, M.H., Rush, A.J., Ibrahim, H.M., Carmody, T.J., Biggs, M.M., Suppes, T., Crismon, M.L., Shores-Wilson, K., Toprac, M.G., Dennehy, E.B., Witte, B., Kashner, T.M., 2004. The inventory of depressive symptomatology, clinician rating (ids-c) and self-report (ids-sr),

and the quick inventory of depressive symptomatology, clinician rating (qids-c) and self-report (qids-sr) in public sector patients with mood disorders: A psychometric evaluation. Psychological Medicine 34, 73-82.
Vaart, W.v.d., Glasner, T.J., 2011. Personal landmarks as recall aids in survey interviews. Field methods 23, 37-56.
van de Schoot, R., 2015. Latent trajectory studies: The basics, how to interpret the results, and what to report. European journal of psychotraumatology 6, 27514-27511.
van Eeden, W.A., van Hemert, A.M., Carlier, I.V.E., Penninx, B.W., Spinhoven, P., Giltay, E.J., 2019. Neuroticism and chronicity as predictors of 9-year course of individual depressive symptoms. Journal of Affective Disorders 252, 484-492.
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 medicine 15, 215-215.
Verkuil, B., Brosschot, J.F., 2012. The online version of the dutch penn state worry questionnaire: Factor structure, predictive validity and reliability. Journal of Anxiety Disorders 26, 844-848.
Vos, T., Haby, M.M., Barendregt, J.J., Kruijshaar, M., Corry, J., Andrews, G., 2004. The burden of major depression avoidable by longer-term treatment strategies. Archives of general psychiatry 61, 1097-1103.
Wagner, H.R., Burns, B.J., Broadhead, W.E., Yarnall, K.S.H., Sigmon, A., Gaynes, B.N., 2000. Minor depression in family practice: Functional morbidity, co-morbidity, service utilization and outcomes. Psychological Medicine 30, 1377-1390.
Wardenaar, K.J., Conradi, H.J., de Jonge, P., 2014. Data-driven course trajectories in primary care patients with major depressive disorder. Depression and anxiety 31, 778-786.
Warshaw, M.G., Keller, M.B., Stout, R.L., 1994. Reliability and validity of the longitudinal interval follow-up evaluation for assessing outcome of anxiety disorders. Journal of Psychiatric Research 28, 531-545.
Wells, K.B., Burnam, M.A., Rogers, W., Hays, R., Camp, P., 1992. The course of depression in adult outpatients: Results from the medical outcomes study. Archives of general psychiatry 49, 788-794.
Wittchen, H.-U., 1994. Reliability and validity studies of the who-composite international diagnostic interview (cidi): A critical review. Journal of Psychiatric Research 28, 57-84.
Wittchen, H.-U., Nelson, C.B., 1996. The composite international diagnostic interview: An instrument for measuring mental health outcome?, Mental health outcome measures. Springer, pp. 179-187.
Wittchen, H.U., Robins, L.N., Cottler, L.B., Sartorius, N., Burke, J.D., Regier, D., 1991. Cross-cultural feasibility, reliability and sources of variance of the composite international diagnostic interview (cidi). The British Journal of Psychiatry 159, 645-653.
Yonkers, K.A., Bruce, S.E., Dyck, I.R., Keller, M.B., 2003. Chronicity, relapse, and illness—course of panic disorder, social phobia, and generalized anxiety disorder: Findings in men and women from 8 years of follow-up. Depression and Anxiety 17, 173-179.
Zimmerman, F.J., Katon, W., 2005. Socioeconomic status, depression disparities, and financial strain: What lies behind the income-depression relationship? Health economics 14, 1197-1215.

SUPPLEMENTARY MATERIAL

I- Instruments

Composite Interview Diagnostic Instrument (CIDI). The CIDI is a standardized diagnostic interview that is used worldwide to classify DSM-IV psychiatric diagnoses. It has been shown to have high interrater reliability, high test-retest reliability, and high validity for depressive and anxiety disorders (Andrews and Peters, 1998; Haro et al., 2006; Wittchen, 1994; Wittchen et al., 1991).

Life Chart Interview (LCI). The LCI uses both age- and calendar-linked personal landmark events and an interactive life event chart to aid in memory recall. The methodology of the LCI has been shown to have high validity and reliability (Warshaw et al., 1994). Examining diagnosis-related symptoms was conditional upon meeting DSM-criteria for a psychiatric disorder as per the CIDI at an assessment point. Specifically, for each month between assessments, the presence of depressive and/or anxiety diagnostic symptoms was determined. If diagnostic symptoms were reported during ≥85% of the months between assessments (i.e., no more than two months without symptoms), we considered this continuous or chronic symptomatology and confirmation of the continuous presence of a psychiatric disorder.

Inventory of Depressive Symptomatology - Self-Report (IDS-SR). The 30-item IDS-SR measures depressive symptoms regarding the previous week on a four-point Likert scale (0-3) (Rush et al., 1996; Trivedi et al., 2004). The IDS-SR has shown adequate internal consistency (Cronbach's alpha ranged from 0.92 to 0.94), test-retest stability, as well as adequate convergent and discriminant validity (Rush et al., 1996).

Beck Anxiety Inventory self-report (BAI). When completing the 21-item BAI, participants rated how much each anxiety symptom troubled them in the past week, on a four-point Likert scale (0-3). Total scores range from 0 to 63 (Muntingh et al., 2011). The BAI has been shown to have high internal validity (Cronbach's alpha ranged from 0.92 to 0.94) (Beck et al., 1988; Fydrich et al., 1992; Spinhoven et al., 2016) and acceptable test-retest reliability over one week, and good discriminant and convergent validity (Beck et al., 1988; Fydrich et al., 1992). ***Fear Questionnaire (FQ).*** When completing the 15-item FQ, participants reported how much they endorsed statements related to agoraphobia, blood/injury phobia, and social phobia. These subscales were previously shown to have moderate to high internal consistency (Cronbach's alpha ranged from 0.71 to 0.83) (Oei et al., 1991). Total scores range from 0 to 120 (Lee and Oei, 1994), with the total scale showing moderate high internal consistency (Cronbach's alpha = 0.83) and adequate reliability (Lee and Oei, 1994). The FQ has high retest reliability (Marks and Mathews, 1979).

Penn State Worry Scale (PSWQ). When completing the 16-item PSWQ, participants rated how much they endorsed statements related to their disposition to worry, including the frequency and burden of the worry complaints, on a five-point Likert scale (1-5) (Meyer et al., 1990). Total scores range from 16 to 80, with higher scores indicative of higher levels of worry. The PSWQ was previously found to have high moderate to high internal consistency (Cronbach's alpha ranged from 0.83 to 0.93) (Brown et al., 1992; Salarifar and Pouretemad, 2012) as well as good test-retest reliability and good convergent, discriminant, and criterion-related validity (Brown et al., 1992; Davey, 1993; Meyer et al., 1990).

World Health Organization Disability Assessment Schedule (WHO-DAS-II). We used the 32-item total score of the WHO-DAS-II and excluded the four-item 'work disability' scale considering the high percentage of patients without work in this population (Bokma et al., 2017; Rizvi et al., 2015; Verduijn et al., 2017; Zimmerman and Katon, 2005). Total scores ranged from 0 to 100, with higher scores reflecting greater disability (Bokma et al., 2017). The total WHO-DAS-II has been shown to have excellent internal consistency (Cronbach's alpha ranged from 0.92 and 0.95) (Bokma et al., 2017; Chwastiak and Von Korff, 2003) and convergent validity (Chwastiak and Von Korff, 2003).

II- Description of categorical trajectories

We used the following definitions for the trajectories of the main analysis ("primary trajectories"):

- ***Consistently recovered*** *(fixed diagnosis-free period)*: participants with no current six-month CIDI diagnosis at two-year follow-up nor at any subsequent follow-up assessment;
- ***Intermittently recovered*** *(variable diagnosis-free period)*: participants with either (A) no current six-month CIDI diagnosis at four-, six-, or nine-year follow-up, preceded by any CIDI diagnosis with LCI symptoms (i.e., in the recurrent or chronic trajectory) at the two-, four-, six-year follow-up period, respectively, or (B) no current six-month CIDI diagnosis at six- or nine-year follow-up, preceded by a delayed recovery course at the four- or six-year follow-up period, respectively;
- ***Intermittently recurrent*** *(variable diagnosis-present period):* either (A) any current six-month CIDI diagnosis at two-year follow-up with discontinuous (LCI) symptoms in the preceding period, or (B) any six-

month CIDI diagnosis at four-, six-, or nine-year follow-up assessments, which was preceded by a recurrent episode (see 3A) or recovery period at two-, four-, or six-year follow-up, respectively;

- ***Consistently chronic*** *(fixed diagnosis-present period):* any current six-month CIDI diagnosis at each follow-up assessment with continuous preceding (LCI) symptoms between all assessment periods.

The "fixed" trajectories integrated each subsequent assessment point from two- to nine-year follow-up, resulting in aggregate percentages of participants with a consistent course trajectory over time, and the "variable" trajectories allowed switching to another course trajectories across the assessment points (course variability).

III- Comparison with previous NESDA study

Our procedure was based on that of Verduijn et al. (2017) but deviated in the following ways: we did not incorporate the LCI (hypo)mania or avoidance symptoms; we did not define an additional trajectory of recurrence with chronic episodes; we integrated comorbidity and diagnostic switching across the assessment points, defining different diagnostic groups at baseline; and we allowed transitioning or course variability between trajectories across the assessment points, with two variable 'intermittent recovery' and 'intermittent recurrence' trajectories. To compare with previous results of Verduijn et al. (2017), we classified three aggregated trajectories across the follow-up assessment points, based on the CIDI diagnoses and LCI symptom duration: 'recovered', 'recurrence', and 'chronicity' ("comparison trajectories"). For the purposes of comparison, we used the following definitions for the aggregated

trajectories, where each percentage was aggregated at each subsequent assessment point from two- to nine-year follow-up:

- ***Recovered***: participants with no current six-month CIDI diagnosis at two-year follow-up and at all subsequent follow-up assessments thereafter;
- ***Recurrent***: participants experiencing a recurrent six-month CIDI diagnosis at any follow-up assessment;
- ***Chronic***: participants with any current six-month CIDI diagnosis at each follow-up assessment with continuous or chronic preceding (LCI) symptoms between all assessment periods.

When calculating the aggregated trajectories based on Verduijn et al. (2017) and comparing the trajectories with those of the main analysis, we found at the nine-year follow-up in the any disorder group that there were slightly fewer participants with a chronic trajectory (6.2% vs 8.5%), and fewer with a sustained recovery trajectory (15.3% vs. 21.0%) (see *Figure A in Supplemental Material*). We also found that 78.5% of participants with any disorder experienced recurrence over a nine-year period. Therefore, at each subsequent follow-up, participants either did not meet the chronicity condition for the chronic trajectory nor the diagnosis-free condition for the sustained recovery trajectory.

IV- Description and interpretation criteria of the latent class growth analysis

For the latent class growth analysis (LCGA), all depression and anxiety severity scores were log-transformed to enable estimation of the log-linear relationships between severity and time. The scores were then standardized to

z-scores to allow comparison of the relative position of the scores within the distribution of each scale, after which these four z-scores were averaged per time point. Baseline values were centered to zero to remove the impact of baseline severity on the prognostic course, thus calculating the degree of change over the two-, four-, six- and nine-year follow-up period in relation to baseline. The aforementioned composite (pooled standardized mean) severity scores were analyzed using LCGA (Andruff et al., 2009; Jung and Wickrama, 2008; van de Schoot, 2015) to estimate one- to six-class models, using full information maximum likelihood, assuming that the missing data were missing at random (MAR). We examined a log-linear model as the primary model. This was compared to other alternative models with other trajectory shapes: quadratic, cubic, beta, and spline with three knots. Additionally, to test for the impact of missing data, we recalculated the LCGA classes without any missing data and compared the resulting trajectories. All models were estimated with 100 random starts to prevent solutions at a local maximum and 10 iterations in the optimization algorithm (Biernacki et al., 2003). Because LCGAs are susceptible to converging on local maximum, rather than global solutions, we conducted a grid search (Biernacki et al., 2003; Proust-Lima et al., 2017).

To determine which model best fitted the data, we inspected the log-likelihood, Akaike Information Criterion (AIC; Akaike, 1987), Bayesian Information Criterion (BIC; Schwarz, 1978), sample-size-adjusted BIC (SA-BIC; Sclove, 1987), entropy values (Ramaswamy et al., 1993), group membership probabilities, and posterior probabilities. Guidelines state that with reference to the AIC, BIC, and SA-BIC, the lower the values, the superior the fit (Armour et al., 2012; Jung and Wickrama, 2008; Muthén, 2003). As more classes are added to the model, fit tends to improve when the AIC, BIC, and SA-BIC values

continue to fall. If the difference between the values of one additional class is small, an additional class is said to add little to the model (DiStefano and Kamphaus, 2006). On the basis of parsimony, the solution with the fewest classes should be accepted (Armour et al., 2012). Moreover, Ramaswamy et al. (1993) reported that high entropy values indicate good classification, with a value of 1.0 indicating perfect model fit and classification. Interpretability and sufficient class size (all > 50) were also used as criteria. After identification of the best model using the aforementioned criteria, we examined how participants were assigned to classes by inspecting group membership probabilities and posterior probabilities using maximum likelihood techniques.

V- The compliance in this manuscript with guidelines for reporting latent trajectory study, according to GRoLTS Checklist (Van de Schoot et al., 2017):

1. Is the metric of time used in the statistical model reported? *Yes.*

2. Is information presented about the mean and variance of time within a wave? *No, but it is represented in the figure.*

3a. Is the missing data mechanism reported? *Yes.*

3b. Is a description provided of what variables are related to attrition/missing data? *Yes.*

3c. Is a description provided of how missing data in the analyses were dealt with? *Yes.*

4. Is information about the distribution of the observed variables included? *Yes.*

5. Is the software mentioned? *Yes.*

6a. Are alternative specifications of within-class heterogeneity considered (e.g., LGCA vs. LGMM) and clearly documented? *We did not explore within-class*

heterogeneity as the primary model. We did explore other models with alternative shapes.

6b. Are alternative specifications of the between-class differences in variance–covariance matrix structure considered and clearly documented? *In the LCGA the variance and covariance were assumed to be fixed to be zero.*

7. Are alternative shape/functional forms of the trajectories described? *Yes.*

8. If covariates have been used, can analyses still be replicated? *No covariates.*

9. Is information reported about the number of random start values and final iterations included? *Yes.*

10. Are the model comparison (and selection) tools described from a statistical perspective? *Yes.*

11. Are the total number of fitted models reported, including a one-class solution? *Yes.*

12. Are the number of cases per class reported for each model (absolute sample size, or proportion)? *Yes, proportion.*

13. If classification of cases in a trajectory is the goal, is entropy reported? *Yes.*

14a. Is a plot included with the estimated mean trajectories of the final solution? *Yes.*

14b. Are plots included with the estimated mean trajectories for each model? *Not for each model due to limited number of figures.*

14c. Is a plot included of the combination of estimated means of the final model and the observed individual trajectories split out for each latent class? *We only show the individual trajectories as this is our goal in the paper.*

15. Are characteristics of the final class solution numerically described (i.e., means, SD/SE, n, CI, etc.)? *No, due to the word- and table limits of our manuscript. However, the figure and table show the final class solution.*

16. Are the syntax files available (either in the appendix, supplementary materials, or from the authors)? *They are available from the first author (Solis, corresponding author).*

VI- Alternative models using latent class growth analysis

Comparing the primary log-linear model to the alternative models resulted in high percentages of similarity across the classes, with approximately 96% similarity for class 1, 92% similarity for class 2, and 92% for class 3 in the quadratic and cubic models, and approximately 84% similarity for class 1, 81% similarity of class 2, and 100% for class 3 in the beta and 3-knot spline models. We thus retained the log-linear model.

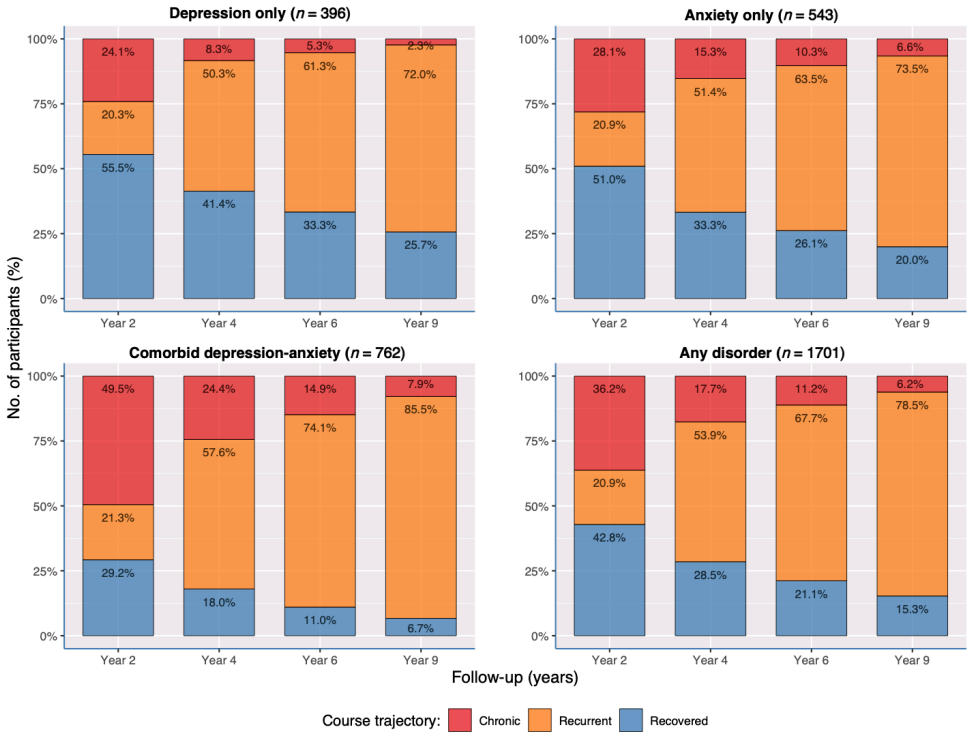

Figure A. Categorical trajectories of aggregate diagnoses at two-, four-, six-, and nine-year assessments, based on any CIDI affective disorder and LCI data, following a procedure similar to Verduijn et al (2017). Depression only = depressive disorders such as major depressive disorder or dysthymia. Anxiety only = anxiety disorders such as social phobia and generalized anxiety disorder.

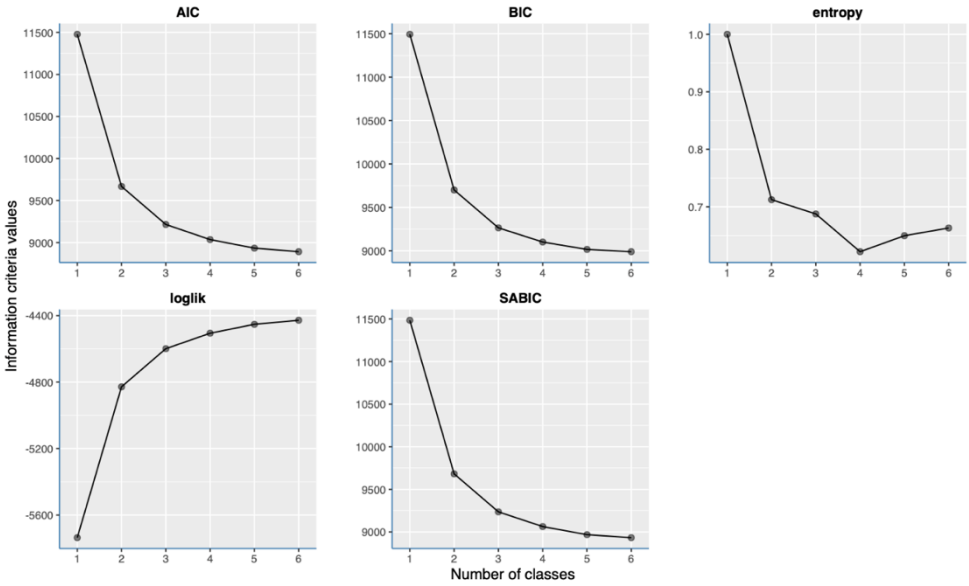

Figure B. Elbow plots of the information criteria values of the Latent Class Growth Analysis comparing one to six classes. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; entropy = entropy values; Loglik = log-likelihood; SABIC = sample-size-adjusted BIC.

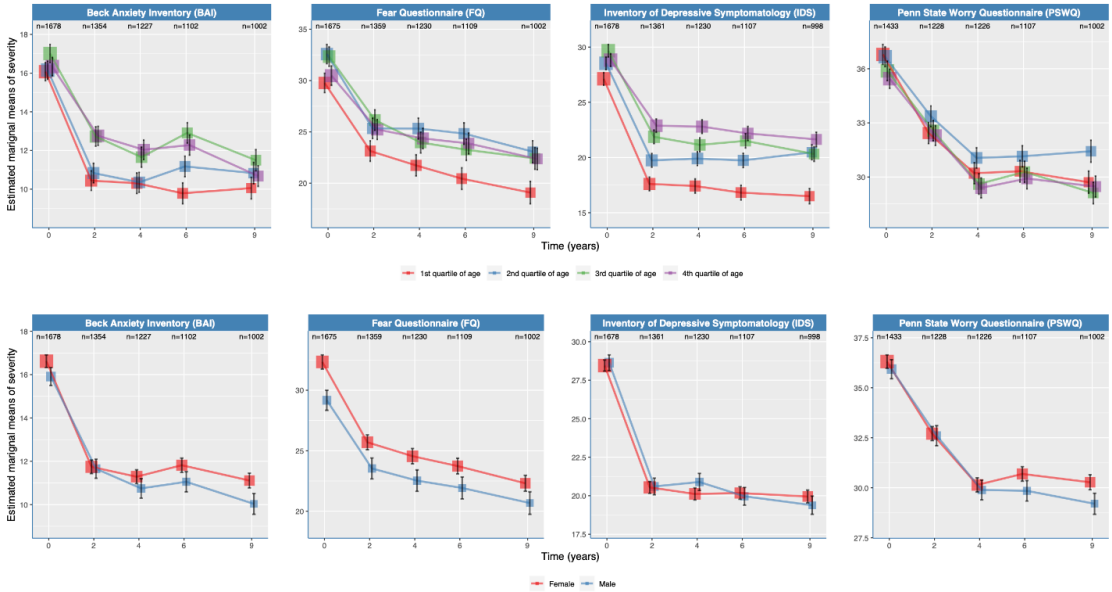

Figure C. Trajectories of depressive, anxiety, fear, and worry symptoms, split into age groups and gender, across the nine-year follow-up period. Error bars represent standard errors, and the box size is proportional to the number of participants.