Acta Psychiatrica Scandinavica

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

van Eeden WA, van Hemert AM, Carlier IVE, Penninx BW, Giltay EJ. Severity, course trajectory, and within-person variability of individual symptoms in patients with major depressive disorder.

Background: Depression shows a large heterogeneity of symptoms between and within persons over time. However, most outcome studies have assessed depression as a single underlying latent construct, using the sum score on psychometric scales as an indicator for severity. This study assesses longitudinal symptom-specific trajectories and within-person variability of major depressive disorder over a 9-year period.
Methods: Data were derived from the Netherlands Study of Depression and Anxiety (NESDA). This study included 783 participants with a current major depressive disorder at baseline. The Inventory Depressive Symptomatology-Self-Report (IDS-SR) was used to analyze 28 depressive symptoms at up to six time points during the 9-year follow-up.
Results: The highest baseline severity scores were found for the items regarding energy and mood states. The core symptoms depressed mood and anhedonia had the most favorable course, whereas sleeping problems and (psycho-)somatic symptoms were more persistent over 9-year follow-up. Within-person variability was highest for symptoms related to energy and lowest for suicidal ideation.
Conclusions: The severity, course, and within-person variability differed markedly between depressive symptoms. Our findings strengthen the idea that employing a symptom-focused approach in both clinical care and research is of value.

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

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

Significant findings

- Depressive symptoms have heterogenetic longitudinal characteristics.
- Somatic/vegetative symptoms are less present at baseline but often exhibit a more persistent course trajectory.
- Mood and cognitive symptoms are more severe at baseline but show favorable course trajectories.

Limitations

- The first part of the symptom trajectories was subject to a 'regression to the mean' because patients were selected based on the criteria for MDD.
- Outcomes were based on analysis with single items.
- Because NESDA was an observational cohort study, other variables may have confounded our findings.

Introduction

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

Studies that did assess symptom-specific course trajectories have shown important differences between individual symptoms. Of the 12 studies that, to some extent, took symptom-specific courses within the adult population into account (2–13), sample sizes ranged from 51 (4) to 3278 participants (13). There were substantial differences in the methods and instruments that were used to assess individual symptoms. Studies used self-report measures (3, 4, 6, 10), clinician-rated measures (5, 8, 11), and structured interviews (2, 7, 9). Therefore, comparing these studies should be done with caution. Most studies featured a prospective design with the duration of follow-up ranging from 2 weeks (3) to 3 years (9). Researchers often focused on identifying residual symptoms, and only three studies specifically reported on relatively fast remitting symptoms (2, 3, 12). Four studies found that the two core symptoms, depressed mood and anhedonia, tended to persist as residual symptoms (2, 4, 5, 11), but sleep problems, energy loss, and cognitive problems were more often reported as residual symptoms (2, 5, 6, 12–14). Fast remitting symptoms were negative self-view and psychomotor problems (3, 7, 12). Some studies found no differences between individual symptoms (7, 8, 10).

The within-person variability of individual MDD symptoms over time has rarely been investigated.

Patients with MDD tend to show a recurring and chronic disease course, with fluctuating levels of severity (2, 15). Two studies found that a high variability of sum scores for severity was associated with an increased risk of relapse (16, 17), whereas another did not (7). Some depressive symptoms tend to show large changes over time in a single patient, whereas other symptoms tend to remain stable or are in steady decline. Based on the mean range of the Hamilton Depression Scale item scores (18), energy loss, loss of libido, and sleep problems showed considerable levels of variability during the 3-year follow-up of 114 patients with MDD (16). On the other hand, suicidal thoughts and psychomotor retardation have demonstrated a more stable course (16).

The present study assessed the longitudinal symptom-specific characteristics of MDD in a large cohort over a 9-year period. To gain more insight into the heterogeneity of MDD, it is important to know which symptoms feature clinically favorable characteristics and which show a more persistent course. Despite the common use of aggregate sum scores in most research, we hypothesized that MDD is a disorder with substantial heterogeneity between symptoms in terms of severity, within-person slopes, and variability. A primary aim was to address some of the methodological gaps in earlier studies by assessing within-person variability over time in which repeated measures are nested within persons (19). Therefore, we assessed baseline severity, course trajectory, and within-person variability of individual symptoms of depression over a 9-year period in a large sample of patients initially suffering from a current MDD.

Methods

Study sample and procedure

Participants were selected from the Netherlands Study of Depression and Anxiety (NESDA) cohort. A detailed description of the NESDA design and sampling procedures are published elsewhere (20). The aim of the NESDA is to investigate the course and consequences of depressive and anxiety disorders. The first wave (baseline) started in 2004 and ended in September 2007. The sixth wave of measurement at the 9-year follow-up finished in October 2016. The baseline measurement ($n = 2981$) consisted of demographic and personal characteristics, standardized diagnostic psychiatric interviews, and medical assessments (e.g., BMI and blood sampling). The 1-year follow-up consisted of a self-report questionnaire and

was completed by 2445 participants (82.0%). A face-to-face follow-up assessment was conducted at 2 years (n = 2596; 87.1%), 4 years (n = 2256; 80.6%), 6 years (n = 2256; 75.7%), and at 9 years postbaseline (n = 2069; 69.4% of the baseline sample).

The cohort was recruited from the community (n = 564; 18.9%), general practice (n = 1610; 54.0%), and secondary mental health care (n = 807; 27.1%; 21). For the present analysis, we only included patients with an 1-month diagnosis of MDD—the excluded participants did not have a mood disorder at the time of baseline assessment (67.3%), had dysthymia without MDD (2.1%) or a minor depression (2.9%). This resulted in a final study sample of 783 participants.

Measures

We used the Composite International Diagnostic Interview (CIDI; WHO version 2.1) to assess the presence of depressive disorders according to the DSM-IV. The CIDI is a fully standardized diagnostic interview with extensively validated psychometric characteristics (20, 21).

Chronic depression and chronic somatic disease at baseline were measured for the purpose of post hoc sensitivity analyses. Depression history was assessed using the Life Chart Interview method—a standardized interview designed to retrospectively assess the course of psychopathology (22). The Life Chart Interview uses age- and calendar-linked life events that occurred over the course of a patient's past 4 years and then assesses the presence and the severity of symptoms during this period. Participants who were depressed for 24 months or more during this period of 48 months (i.e., >50% of the time) were defined as being chronically depressed (22).

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

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

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

In our study sample, the Cronbach's alphas were 0.83, 0.89, 0.89, 0.90, 0.90, and 0.90 for the six time points, respectively, from baseline to 9 years. Because Items 11 and 12 (increased/decreased appetite) and Items 13 and 14 (weight gain/weight loss) contained opposite features, these item pairs were combined into one ordinal item in order to maintain psychometric similarity between the items, which yielded 28 items for the present analyses (23). In order to enhance interpretability, we grouped the symptoms by symptom clusters, which were previously identified across various studies in the figures (28). The symptom clusters had no role in computing our outcome variables—only in how they were grouped in the figures. The clusters include 10 mood symptoms (capacity for pleasure, general interest, quality of mood, reactivity of mood, feeling anxious or tense, feeling irritable, feeling sad, interpersonal sensitivity, leaden paralysis, panic/phobic symptoms), 14 somatic/vegetative symptoms (aches and pains, constipation/diarrhea, mood in time of the day, waking up early, low energy, sympathetic arousal, problems falling asleep, sleep during the night, psychomotor agitation, psychomotor retardation, interest in sex, sleeping too much, weight gain/loss, increased/decreased appetite), and four cognitive problems (concentration/decision-making, view of my future, view of myself, suicidal thoughts; (28)).

Statistical analysis

Multiple steps were taken to assess the longitudinal MDD symptom characteristics. The outcome measures (baseline item score, slope, and fraction of

variance unexplained) were summarized and presented with a 95% CI (represented by error bars) in forest plots, which were sorted by the size of each mean effect estimate. All analyses were computed using R, version 3.4.1, with main packages mixor (29), mirt (30), tidyverse (31), ggplot2 (32), and ggrepel (33).

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

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

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

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

Fraction of variance unexplained. We calculated the fraction of variance unexplained (FVU) per item as a measure of within-person variability. A high FVU represents a variable course with more fluctuation throughout the follow-up years. A low FVU represents a stable course, that is, symptoms with a steady decline or a stable persistent course or symptoms that, if not present, are not likely to be present in the future. FVU was calculated using a simple linear regression analysis. We computed the regression analyses per person and per item, resulting in a total of 15 624 modeled regression lines (i.e., n. of participants * no. of items). As the steep slope within the first year at follow-up had disproportional large impact on the FVU measure and we were interested in the FVU as a function of within-person variability over time, and not as a function of recovery, we decided to exclude the baseline measurements when calculating the FVU. When patients did not fulfill all five follow-up IDS assessments, regression analyses were computed based on the remaining time points (at least three). This approach of modeling course variability per individual has been used in other fields of medical research, for example, blood pressure variability (37).

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

Results

Demographics

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

Mean values over time

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

Baseline severity

The baseline mean (with standard error) is presented in Table 2 and Fig. 1. Baseline mean with the 95% CI and baseline mean in relation to the 1-year slope are presented in the Supporting

Table 1. Sociodemographic and clinical characteristics

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

TCA, tricyclic antidepressants; SSRI, selective serotonin reuptake inhibitors; AD, antidepressants; IDS-SR, The Inventory Depressive Symptomatology-Self-Report.
*Depressed for 24 months or more at baseline.
†The following diseases were asked: asthma, chronic bronchitis or pulmonary emphysema, heart disease, diabetes, stroke or CVA, osteoarthritis, cancer, stomach or intestinal ulcers, intestinal disorders, liver disease, epilepsy, or thyroid gland disease, and other chronic disease.
‡Refers to mental health care.

information (see Figs S1 and S2). The baseline mean of all items combined was 1.29 and ranged from 0.69 (Item 4 'sleeping too much') to 1.93 (Item 30 'leaden paralysis'). The highest baseline severity was found for items concerning energy and depressed mood (Items 30, 10, 20, 5), followed by 'low self-esteem' (Item 16), 'sleep during the night' (Item 2), 'concentration' (Item 15), 'feeling anxious or tense' (Item 7), and 'sensitivity' (Item 29). Interestingly, the mean of Items 20 'energy level' and 2 'sleep during the night' showed a much higher baseline mean level compared to most other symptoms within the somatic/vegetative cluster. Other items within the somatic/vegetative domain were less severe at baseline. The lowest mean baseline values were found for Item 4 'sleeping too much', Item 9a 'mood in time of the day', Item 3 'waking up too early', and 'thoughts of death or suicide'.

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

When assessing the six items that in previous studies were found to represent a unidimensional melancholia construct (34, 35), component loadings ranged from 0.244 (item 16 'view of myself') to 0.605 (item 5 'feeling sad'). Of the six items, 2 items had rather weak component loadings below 0.400, that is, item 16 'view of myself' (loading = 0.244), and item 23 'psychomotor retardation' (loading = 0.338). Only item 5 'feeling sad' (a = 1.293) had a discrimination parameter above 1 and three items had partial credit model parameters that were not ordered in accordance with the item scales, that is, item 16 'View of myself', item 19 'General interest', and item 23 'Psychomotor retardation'. More detailed results regarding component loadings, the discriminative properties and item-specific partial credit model parameter estimates thresholds can be found in Table S1 and Fig. S3. In sum, our findings from the IRT analyses are not in support of the idea of a single coherent latent construct of depression.

Table 2. IDS symptoms during 9 years follow-up

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

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

Slope during the first year

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

The symptom course (slope) in relation to baseline severity (mean item score) is shown in Fig. S2. Items with a high baseline mean tended to show a stronger decrease over time. Two items with steep slopes fell within the 95% CI: Item 30 ('leaden paralysis') and Item 10 ('quality of mood'). Two items with slopes close to zero also had a small mean baseline item score (within the 95% CI): Item 4 ('sleeping too much') and Item 28 ('constipation/diarrhea'). The regression line with a 95% CI provides insight into the association between baseline severity and slope and symptoms that do not fulfill this association.

Slope adjusted for IDS sum scores

The adjusted slopes over 9 years are presented in Table 2 and Fig. 2b. Twelve symptoms had a slope that was significantly different from 0, indicating a larger or smaller decrease than the overall sum score. Of the six items with a relatively larger decline, three were in the mood symptom cluster: Item 5 ('feeling sad'), Item 8 ('response of mood'), and Item 10 ('quality of mood'); one was a cognitive symptom (Item 15 'concentration'); and two items were in the somatic/vegetative symptom cluster: Item 11 ('appetite') and Item 24 ('psychomotor agitation'). Four symptoms with a smaller decrease than the overall sum score were in the somatic/vegetative symptom cluster: Item 2 ('sleep during the
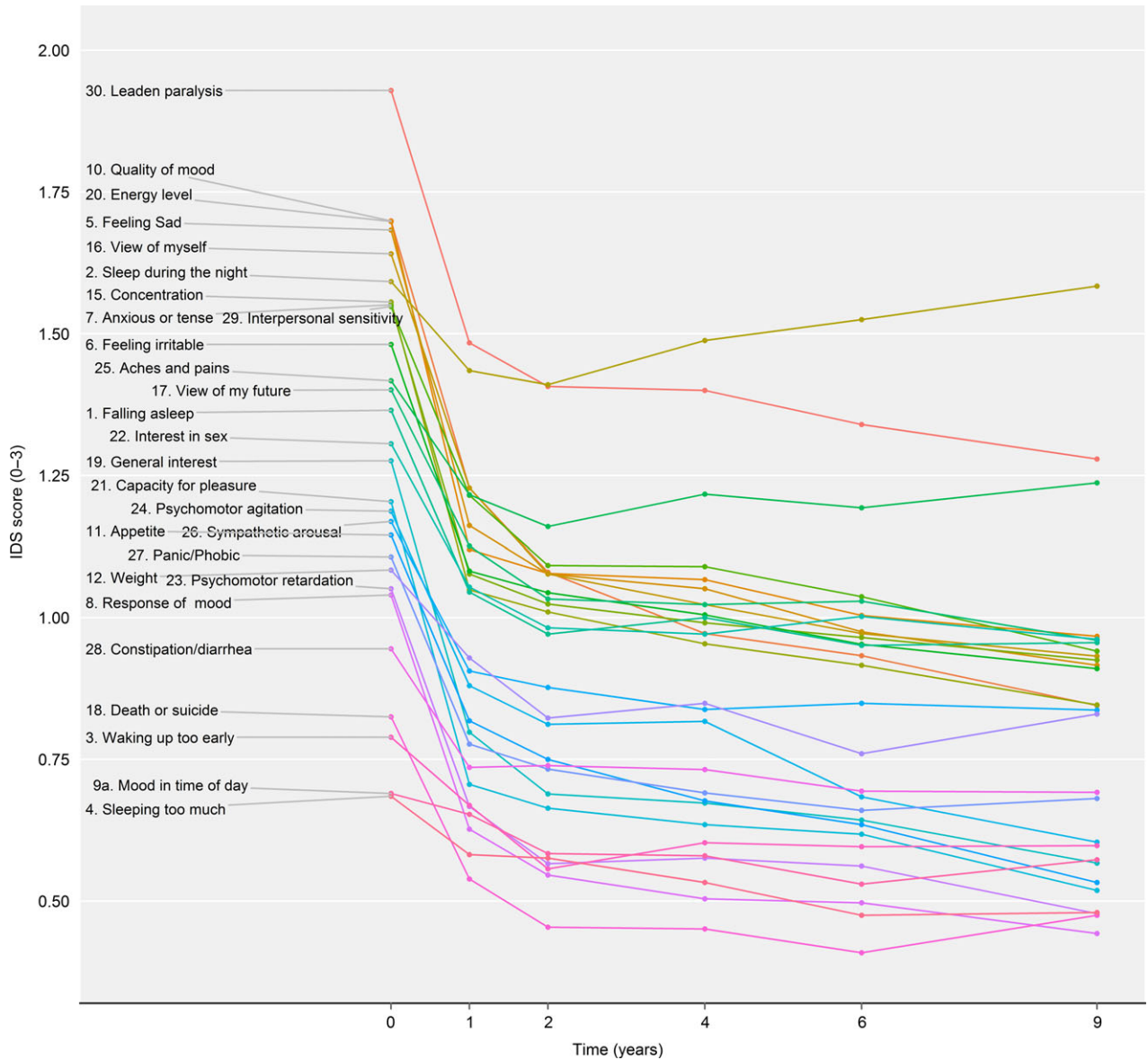
Fig. 1. Group-level mean item scores over the course of 9 years. [Colour figure can be viewed at wileyonlinelibrary.com]

night'), Item 3 ('waking up too early'), Item 25 ('aches and pains'), Item 22 ('interest in sex'), and Item 26 ('sympathetic arousal'). One item with a small decrease fell within the cognitive symptom cluster: Item 17 ('view of my future').

Variability

Patients with four or more IDS-SR assessments were included for the FVU analysis, which resulted in a sample size of *n* = 498. Excluded patients (with less than four assessments; *n* = 244) were less likely to be of northern European heritage (86.9% vs. 95.3%; *P* < 0.006) and had lower mean number years of education (*M* = 10.4 vs. *M* = 11.7 years; *P* < 0.004). We found similar characteristics

between included and excluded patients for the remaining variables mentioned in Table 1, such as gender (64.7% female), antidepressants (3.2% TCA; 29.9% SSRI; 10.9% other antidepressants; 56.8% no antidepressants), chronic depression (36%), chronic somatic disease (61.2%), and IDS-SR sum score (34.7; SD = 11.3).

The within-person FVU for each symptom is presented in Table 2 and Fig. 3. The overall FVU of all items combined was 0.498, ranging from 0.339 (Item 18 'death or suicide') to 0.591 (Item 30 'leaden paralysis'). Among the items with high within-person variability, all three symptom clusters were equally represented. Item 30 ('leaden paralysis') was the most unstable followed by Item 11 ('weight'), Item 7 ('anxious or tense'), Item 25
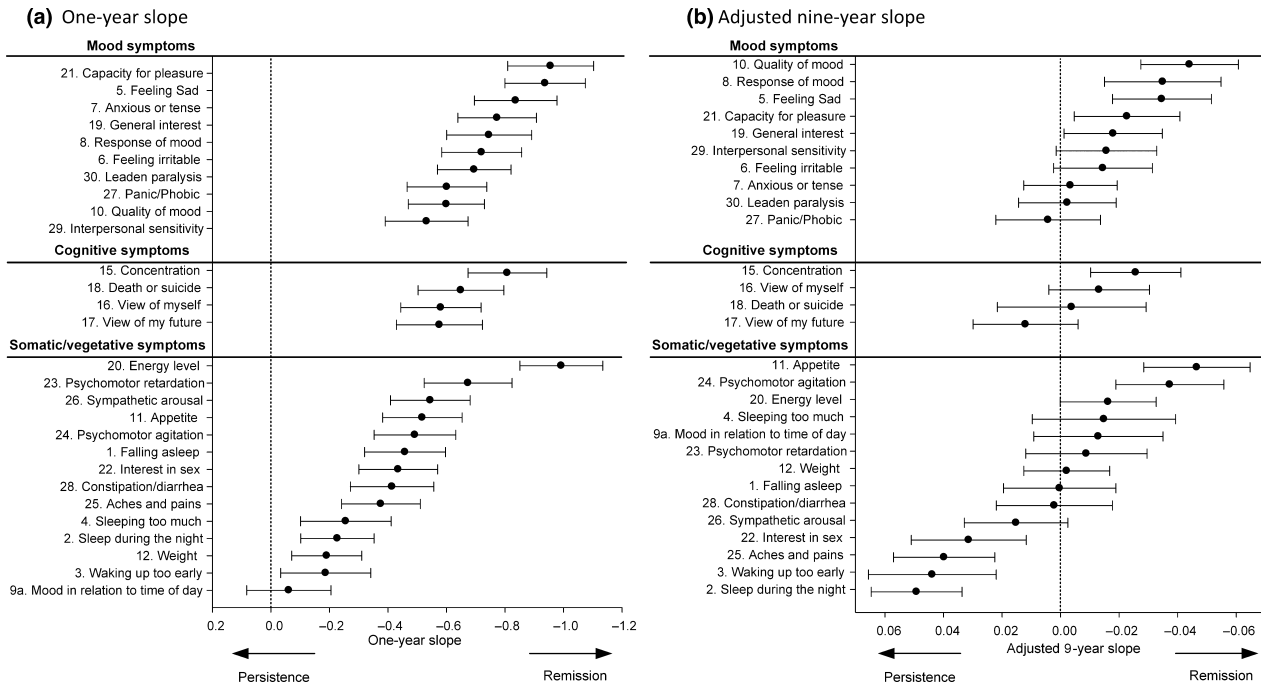
Fig. 2. Unadjusted 1-year slope represents the decrease in symptom severity after the first year of follow-up. Negative values represent a steeper decline. Sum score adjusted 9-year slope represents the decline in symptom severity in relation to the sum score. Negative values mean that the symptom had a steeper decline compared to the overall sum score.

('aches and pains'), Item 29 ('interpersonal sensitivity'), and Item 20 ('low energy level'). The most stable items fell within the somatic/vegetative symptom cluster, with the exception of Item 18 ('thinking of death or suicide'). Other particularly stable items were Item 3 ('waking up too early'), Item 23 ('psychomotor retardation'), and Item 4 ('sleeping too much'). Note that many of the stable symptoms had low baseline severity. This means that when symptoms were not present at baseline, they were often unlikely to be present at the follow-up, except for Item 1 'falling asleep'.

Sensitivity analysis

We conducted several post hoc sensitivity analyses in which we assessed the effects on baseline severity, slope, and FVU. These results are presented in Table S2. We assessed baseline severity (i.e., the mean baseline item score) in the subgroup of patients with a history of chronic depression, chronic somatic disease, and antidepressant users. Overall, the mean baseline severity was slightly lower when we excluded chronic depressed patients (delta = -0.09; $M = 1.20$), patients with chronic somatic diseases (delta = -0.05; $M = 1.25$), and antidepressant users (delta = -0.08; $M = 1.21$). When taking individual items into account, no meaningful differences were found because only two items had a delta (i.e., unadjusted mean minus

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

For the next sensitivity analyses, 1-year slope and 9-year slope findings were tested for robustness. Therefore, models were adjusted for a history of chronic depression, chronic somatic diseases, the use of antidepressants, and age. This again resulted in similar findings. Sorting on effect sizes did not change the order (see Table S2).

Finally, sensitivity analyses for the FVU hardly affected our findings. Only three items showed slight changes when patients with a history of chronic depression were excluded (Item 15 'concentration', Item 17 'view of my future', and Item 25 'aches and pains') and three items when antidepressant users were excluded (Item 2 'sleep during the night', Item 15 'concentration', and Item 17 'view of my future'; see Table S2).

Discussion

Our study confirms the existence of substantial heterogeneity between depressive symptoms in terms of symptom severity at baseline, slopes over time, and within-person variability over time.
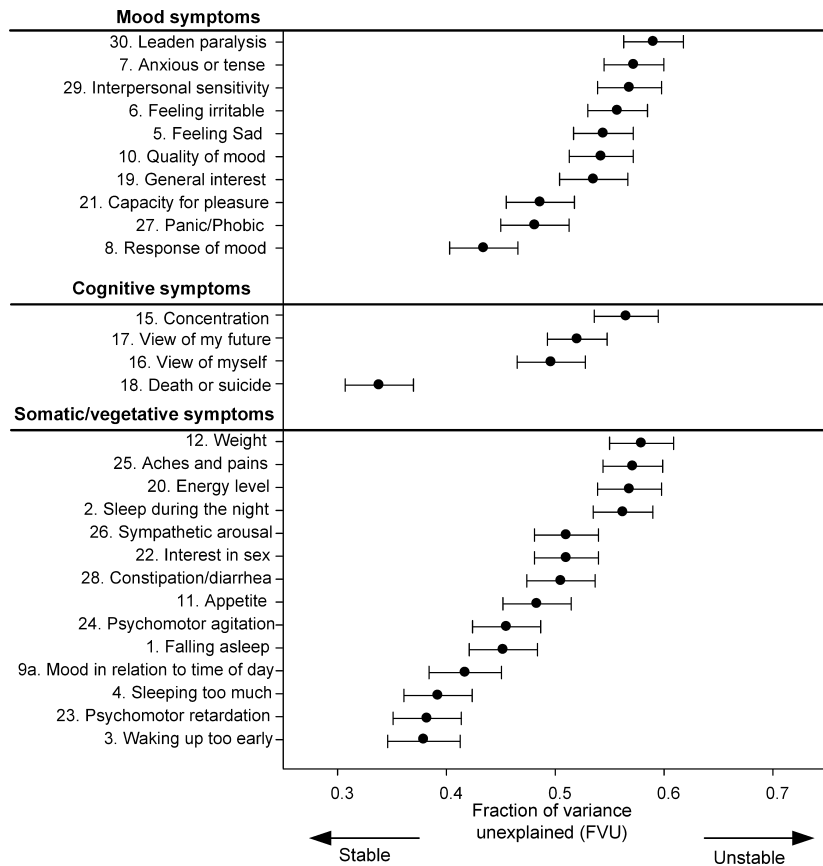
Fig. 3. Within-person variability based on 8 years follow-up (baseline excluded)

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

For all items in our study, the largest (mean) recovery took place within the first study year. When the diagnostic criteria for MDD were assessed 2 years postbaseline, 70% of the patients had recovered from MDD. However, other studies report that, although 50–90% recovered within the first year, many patients still experienced residual symptoms or relapsed after initial remission (15, 38).

Research on the symptom-specific characteristics during and directly following a depressive episode is scarce. In our group of MDD patients, a depressed mood and low energy level were among the most severe symptoms at baseline, which is in line with most other reports (7, 12). In our population, in contrast to others (2, 5, 11), the mood symptoms (e.g., depressed mood and anhedonia) showed a more favorable course. Somatic/vegetative symptoms, such as sleep and somatic complaints, often had more persistent course trajectories. The persistent course of insomnia is in line with most other studies (5, 39–41), with two exceptions (7, 13). The generally low severity at baseline, but persistent nature of multiple somatic symptoms associated with depression, has been documented in earlier studies (342, 43). These studies suggested that patients who experience these symptoms may represent a separate subgroup of MDD (42, 43).

We found significant differences between symptoms regarding within-person variability. Suicidal ideation tended to be stable and showed less fluctuation within patients over time. If patients had suicidal ideations, they were likely to keep on having these ideations during the subsequent years of

follow-up. If patients did not have suicidal ideations during their depressive episode at baseline, they were unlikely to experience them in the future. Suicidality is described in the literature as being related to a specific cognitive response pattern of hopelessness; this pattern is continually present throughout an individual's life (44). From a psychometric perspective, we could argue that the latent thresholds for scoring 0, 1, 2, or 3 on the item 'energy level' are much lower than those on the item 'suicidal ideation' (45). A 1-point change in an unstable item, such as 'energy level', is clinically of less importance than a 1-point change in a stable item, such as 'suicidal ideation'. Our results on variability are in line with those of Karp et al. (16) who found energy loss to be an unstable symptom and suicidal thoughts to be a stable symptom among 114 patients with MDD (aged 21–65 years) during a follow-up lasting 3 years. More research is needed on the topic of within-person variability. Beside group-level changes of individual items, the within-person variability may have additional predictive and/or clinical value.

Drawing inferences about changes in depression severity is an imperfect process because severity cannot be measured directly (19). Outcome measurements are generally based on a questionnaire sum score in which the same weight is given to each item. This method would be valid if MDD was a unified construct and all its symptoms contributed equally to its latent construct (1, 15). However, MDD is unlikely to be a distinct illness that causes all of its symptoms (1, 9, 46). Instead, MDD is more like a complex system in which symptoms are connected by a dynamic network of causality (47–49). The symptom-specific diversity in mean item scores, slopes, and variability shows that symptoms are not diagnostically equivalent and are not interchangeable (50). The persistent use of merely a sum score to estimate depression severity may obscure insight into both patient and symptom-specific characteristics and can lead to misinterpretations regarding depressive severity over time (1, 51). For example, a patient who recovers by feeling less depressed will show a similar change in the depressive severity measure as a patient whose recovery takes place in another symptom domain, such as sleep. Even when there is a significant change in the sum score, a clinically important change might be obscured by more trivial changes on other items. It is therefore advised to assess individual symptoms in addition to sum scores when testing a patient's (longitudinal) depressive characteristics.

Research on personalized medicine in mental health care (15, 52, 53) and treatment of specific (residual) symptoms has highlighted that a symptom-specific approach may be beneficial (54, 55). In general, depression treatment focuses mainly on the core symptoms of depression. However, other symptoms (e.g., sleeping problems) are more persistent and can indicate a risk factor for relapse; therefore, these symptoms deserve particular attention as a focus for treatment (39, 40). Moreover, because a causal relationship exists between symptoms on group level (47–49), targeting the key symptoms (i.e., more central in the causal network of depressive symptoms) in clinical care may benefit a patient's recovery. Symptom-specific treatment of, for example, sleeping problems are widely available (41, 56, 57). For instance, cognitive behavioral therapy and pharmacological treatment for insomnia appear to have a positive effect on depression (56, 57). It seems that our currently applied treatments warrant a more symptom-specific approach in order to also take the persistent (somatic/vegetative) symptoms into account.

The present study has several strengths. A large sample of MDD patients was included and followed for up to 9 years, whereas many earlier studies featured shorter follow-up periods or cross-sectional designs. Using a per-person, per-item method allowed us to compute a measure for within-person variability. Although the use of this method is relatively rare in the field of psychiatric research, it is often used in other fields of medicine (37).

The study also has some limitations. First, because all patients were initially selected to fulfill criteria for MDD, the first part of the symptom trajectories was subject to a 'regression to the mean' effect (58). Therefore, baseline severity needed to be taken into account when interpreting the slope measures. Furthermore, because the steep decline within the first year had a large effect on the variance within patients, we calculated the FVU and excluded the baseline measure. Second, the FVU measure may be affected by the design of the IDS-SR with severity measured on a nominal scale. When participants scored a baseline severity of 0 on a particular item, there would only be room for change toward the higher scores. On the other hand, a baseline ordinal score of 3 is the highest score and scores above that point cannot be measured, this again limits the ability of the instrument to detect variability. Third, assessing individual symptoms based on single items presents psychometric difficulties. Single items are more strongly affected by random error than sum scores of items, which may have particularly affected our FVU measures. Moreover, we did not assess the reliable change indices because the focus of our study was

not on the clinical impact of a one-point ordinal scale change in each item. Finally, because the NESDA was an observational cohort study, several variables may have confounded our findings. We performed multiple sensitivity analyses to test other variables, such as pharmacological treatment (e.g., antidepressants). We found that our results remained robust and that only minimal changes occurred after adjusting for other variables.

In this study, we examined within-person trajectories over time of different depressive symptoms measured using the IDS-SR. The severity, course, and variability differed markedly between the depressive symptoms and between patients, which further supports the idea that MDD is a heterogeneous disease, rather than a singular construct, when studied over time (1, 50). We recommend the advancement of symptom-specific and personalized approaches for both interventional and observational research. The sum scores of symptom questionnaires might obscure too much information potentially yielded by the individual symptoms. Moreover, a symptom-specific study approach may help the development of symptom-specific treatment strategies.

Acknowledgements

Declaration of interest

None.

References

1. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med 2015;**13**:72.
2. Conradi H, Ormel J, De Jonge P. Symptom profiles of the DSM-IV-defined remission, recovery, relapse, and recurrence of depression: the role of the core symptoms. Depress Anxiety 2012;**29**:638–645.
3. Sakurai H, Uchida H, Abe T et al. Trajectories of individual symptoms in remitters versus non-remitters with depression. J Affect Disord 2013;**151**:506–513.
4. Young MA, Watel LG, Lahmeyer HW, Eastman CI. The temporal onset of individual symptoms in winter depression: differentiating underlying mechanisms. J Affect Disord 1991;**22**:191.
5. Romera I, Perez V, Ciudad A et al. Residual symptoms and functioning in depression, does the type of residual symptom matter? A post-hoc analysis BMC Psychiatry 2013;**13**:51.
6. Madhoo M, Levine SZ. Initial severity effects on residual symptoms in response and remission: a STAR* D study during and after failed citalopram treatment. J Clin Psychopharmacol 2015;**35**:450–453.
7. Minor KL, Champion JE, Gotlib IH. Stability of DSM-IV criterion symptoms for major depressive disorder. J Psychiatr Res 2005;**39**:415–420.
8. Paykel ES, Ramana R, Cooper Z et al. Residual symptoms after partial remission: an important outcome in depression. Psychol Med 1995;**25**:1171–1180.
9. Wardenaar KJ, Monden R, Conradi HJ, De Jonge P. Symptom-specific course trajectories and their determinants in primary care patients with major depressive disorder: evidence for two etiologically distinct prototypes. J Affect Disord 2015;**179**:38–46.
10. Rabin AS, Kaslow NJ, Rehm LP. Changes in symptoms of depression during the course of therapy. Cognit Ther Res 1984;**8**:479–487.
11. Harada E, Satoi Y, Kikuchi T et al. Residual symptoms in patients with partial versus complete remission of a major depressive disorder episode: patterns of painful physical symptoms in depression. Neuropsychiatr Dis Treat 2016;**12**:1599–1607.
12. Conradi HJ, Ormel J, De JP. Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study. Psychol Med 2011;**41**:1165–1174.
13. Nil R, Lütolf S, Seifritz E. Residual symptoms and functionality in depressed outpatients: A one-year observational study in Switzerland with escitalopram. J Affect Disord 2016;**197**:245–250.
14. Menza M, Marin H, Opper RS. Residual symptoms in depression: can treatment be symptom-specific? J Clin Psychiatry 2003;**64**:516–523.
15. Musliner KL, Munk-Olsen T, Eaton WW, Zandi PP. Heterogeneity in long-term trajectories of depressive symptoms: patterns, predictors and outcomes. J Affect Disord 2016;**192**:199–211.
16. Karp JF, Buysse DJ, Houck PR et al. Relationship of variability in residual symptoms with recurrence of major depressive disorder during maintenance treatment. Am J Psychiatry 2004;**161**:1877–1884.
17. Torres LD. Understanding the variability of depression symptoms in recovery: life stress as context and consequence in the course of recurrent depression. Dissert Abstr Int B Sci Eng 2008;**68**:8417.
18. Hamilton M. Hamilton depression scale. Göttingen: Internationale Skalen für Psychiatrie Beltz Test; 1996: pp 93–96.
19. Curran PJ, Bauer DJ. The disaggregation of within-person and between-person effects in longitudinal models of change. Annu Rev Psychol 2011;**62**:583–619.
20. Penninx BW, Beekman AT, Smit JH et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Meth Psychiatr Res 2008;**17**:121–140.
21. Wittchen H-U. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): a critical review. J Psychiatr Res 1994;**28**:57–84.
22. Lyketsos CG, Nestadt G, Cwi J, Heithoff K. The Life Chart Interview: a standardized method to describe the course of psychopathology. Int J Meth Psychiatr Res 1994;**4**:143–155.
23. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The inventory of depressive symptomatology (IDS): psychometric properties. Psychol Med 1996;**26**:477–486.

24. RUSH AJ, TRIVEDI MH, IBRAHIM HM et al. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003;**54**:573–583.

25. RUSH AJ, TRIVEDI MH, CARMODY TJ et al. Self-reported depressive symptom measures: sensitivity to detecting change in a randomized, controlled trial of chronically depressed, nonpsychotic outpatients. Neuropsychopharmacology 2005;**30**:405–416.

26. HAMILTON M. Hamilton depression scale. Group 2004;**1**:4.

27. BECK AT, STEER RA, BROWN GK. Beck depression inventory-II. San Antonio 1996;**78**:490–498.

28. SCHAAKXS R, COMIJS H, LAMERS F, BEEKMAN A, PENNINX B. Age-related variability in the presentation of symptoms of major depressive disorder. Psychol Med 2017;**47**:543–552.

29. ARCHER KJ, HEDEKER D, NORDGREN R, GIBBONS RD. mixor: an R package for longitudinal and clustered ordinal response modeling, 2015.

30. CHALMERS RP. Mirt: a multidimensional item response theory package for the R environment. J Stat Soft 2012;**48**:1–29.

31. WICKHAM H, GROLEMUND G. R for data science: import, tidy, transform, visualize, and model data. Sebastopol, CA: O'Reilly Media, Inc.; 2016.

32. WICKHAM H. ggplot2: elegant graphics for data analysis. Basel, Switzerland: Springer; 2016.

33. SLOWIKOWSKI K. ggrepel: repulsive text and label geoms for 'ggplot2'. R package version 06, 2016; **5**.

34. BECH P, FAVA M, TRIVEDI MH, WISNIEWSKI SR, RUSH AJ. Factor structure and dimensionality of the two depression scales in STAR* D using level 1 datasets. J Affect Disord 2011;**132**:396–400.

35. ØSTERGAARD SD, BECH P, TRIVEDI MH, WISNIEWSKI SR, RUSH AJ, FAVA M. Brief, unidimensional melancholia rating scales are highly sensitive to the effect of citalopram and may have biological validity: Implications for the Research Domain Criteria (RDoC). J Affect Disord 2014;**163**:18–24.

36. BECKER WE, KENNEDY PE. A graphical exposition of the ordered probit. Econ Theory 1992;**8**:127–131.

37. WEBB AJ, FISCHER U, MEHTA Z, ROTHWELL PM. Effects of antihypertensive-drug class on interindividual variation in blood pressure and risk of stroke: a systematic review and meta-analysis. Lancet 2010;**375**:906–915.

38. RHEBERGEN D, BEEKMAN AT, DE GRAAF R et al. The three-year naturalistic course of major depressive disorder, dysthymic disorder and double depression. J Affect Disord 2009;**115**:450–459.

39. BUYSSE DJ, ANGST J, GAMMA A et al. Prevalence, course, and comorbidity of insomnia and depression in young adults. Sleep 2008;**31**:473.

40. LUSTBERG L, REYNOLDS CF. Depression and insomnia: questions of cause and effect. Sleep Med Rev 2000;**4**:253–262.

41. BALDWIN DS, PAPAKOSTAS GI. Symptoms of fatigue and sleepiness in major depressive disorder. J Clin Psychiatry 2006;**67**:9–15.

42. WANDERS RBK, WARDENAAR KJ, PENNINX BWJH, MEIJER RR, DE JONGE P. Data-driven atypical profiles of depressive symptoms: Identification and validation in a large cohort. J Affect Disord 2015;**180**:36.

43. MONDEN R, WARDENAAR KJ, STEGEMAN A, CONRADI HJ, DE JONGE P. Simultaneous decomposition of depression heterogeneity on the person-, symptom-and time-level: the use of three-mode principal component analysis. PLoS ONE 2015;**10**:e0132765.

44. WILLIAMS J, VAN DER DOES A, BARNHOFER T, CRANE C, SEGAL Z. Cognitive reactivity, suicidal ideation and future fluency: preliminary investigation of a differential activation theory of hopelessness/suicidality. Cognit Ther Res 2008;**32**:83–104.

45. OORT FJ, VISSER MR, SPRANGERS MA. Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. J Clin Epidemiol 2009;**62**:1126–1137.

46. SJÖHOLM L, LAVEBRATT C, FORSELL Y. A multifactorial developmental model for the etiology of major depression in a population-based sample. J Affect Disord 2009;**113**:66–76.

47. BEARD C, MILLNER AJ, FORGEARD MJ et al. Network analysis of depression and anxiety symptom relationships in a psychiatric sample. Psychol Med 2016;**46**:3359–3369.

48. BORKULO CV, BOSCHLOO L, BORSBOOM D et al. Association of symptom network structure with the course of longitudinal depression (Report). JAMA Psychiatry 2015;**72**:1219.

49. FRIED EI, EPSKAMP S, NESSE RM, TUERLINCKX F, BORSBOOM D. What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. J Affect Disord 2016;**189**:314–320.

50. CRAMER AOJ, VAN BORKULO CD, GILTAY EJ et al. Major depression as a complex dynamic system. PLoS ONE 2016;**11**:e0167490.

51. SNAITH P. What do depression rating scales measure? Br J Psychiatry 1993;**163**:293–298.

52. ARNOW BA, BLASEY C, WILLIAMS LM et al. Depression subtypes in predicting antidepressant response: a report from the iSPOT-D trial. Am J Psychiatry 2015;**172**:743–750.

53. CHEKROUD AM, GUEORGUIEVA R, KRUMHOLZ HM et al. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. JAMA Psychiatry. 2017;**74**:370–378.

54. FAVA M. Pharmacological approaches to the treatment of residual symptoms. J. Psychopharmacol. 2006;**20**:29–34.

55. RAFANELLI C, PARK SK, FAVA GA. New psychotherapeutic approaches to residual symptoms and relapse prevention in unipolar depression. Clin Psychol Psychother 1999;**6**:194–201.

56. BRITTON WB, HAYNES PL, FRIDEL KW, BOOTZIN RR. Mindfulness-based cognitive therapy improves polysomnographic and subjective sleep profiles in antidepressant users with sleep complaints. Psychother Psychosom 2012;**81**:296–304.

57. MANBER R, EDINGER J, GRESS J et al. Cognitive behavioral therapy for insomnia enhances depression outcome in patients with comorbid major depressive disorder and insomnia. Sleep 2008;**31**:489–495.

58. STREINER DVL. Regression toward the mean: its etiology, diagnosis, and treatment. Can J Psychiatry 2001;**46**:72–76.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article:
Figure S1. Supplementary material. Baseline mean item score.
Figure S2. Supplementary material. One-year slope in relation to level of severity (intercept).
Figure S3. Supplementary material. Item response theory analysis of the IDS-SR 6.
Table S1. Item response theory analysis of the IDS-SR 30.
Table S2. Sensitivity analysis, values adjusted for a history of chronic depression, chronic somatic disease, age, or the use of antidepressants.