



Published in final edited form as:

Psychol Med. 2014 July ; 44(10): 2067–2076. doi:10.1017/S0033291713002900.

Depression is more than the sum-score of its parts: individual DSM symptoms have different risk factors

Eiko I. Fried^{1,2}, Randolph M. Nesse^{3,4}, Kara Zivin^{4,5,6}, Constance Guille⁷, and Srijan Sen^{4,*}

¹Cluster of Excellence “Languages of Emotion”, Freie Universität Berlin, Berlin, Germany

²Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

³Department of Psychology, University of Michigan, Ann Arbor, MI, USA

⁴Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI, USA

⁵Department of Veterans Affairs, National Serious Mental Illness Treatment Resource and Evaluation Center, Ann Arbor, MI, USA

⁶Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

⁷Medical University of South Carolina, Charleston, SC, USA

Abstract

Background—For diagnostic purposes, the nine symptoms that compose the DSM-5 criteria for Major Depressive Disorder (MDD) are assumed to be interchangeable indicators of one underlying disorder, implying that they should all have similar risk factors. The present study investigates this hypothesis, utilizing a population cohort that shifts from low to elevated depression levels.

Methods—We assessed the nine DSM-5 MDD criterion symptoms and seven depression risk factors (personal and family MDD history, sex, childhood stress, neuroticism, work hours, and stressful life events) in a longitudinal study of medical interns prior to and throughout internship ($n=1289$). We tested if risk factors varied across symptoms, and whether a latent disease model could account for heterogeneity between symptoms.

Results—All MDD symptoms increased significantly during residency training. Four risk factors predicted increases in unique subsets of PHQ-9 symptoms over time (depression history, childhood stress, sex, and stressful life events), while neuroticism and work hours predicted increases in all symptoms, albeit to varying magnitudes. MDD family history did not predict increases in any symptom. The strong heterogeneity of associations persisted after controlling for a latent depression factor.

Conclusions—The influence of risk factors varies substantially across DSM depression criterion symptoms. Since symptoms are etiologically heterogeneous, considering individual

*Address for correspondence: Srijan Sen, Rachel Upjohn Building, 4250 Plymouth Rd, Ann Arbor, MI 48109-5734, USA. srijan@umich.edu. .

Declaration of interest

None.

symptoms in addition to depression diagnosis might offer important insights obscured by symptom sum-scores.

Keywords

depression; depressive symptoms; DSM; heterogeneity; psychiatric diagnosis; residency

Introduction

Major depressive disorder (MDD) is common, burdensome, recurrent, and expensive (Solomon *et al.* 2000; Kessler *et al.* 2003; Kessler *et al.* 2005; Lopez *et al.* 2006). The Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013) criteria for MDD are met when an individual exhibits at least five out of nine psychiatric symptoms during the same two-week period, with at least one symptom being depressed mood or loss of interest. Interestingly, four of the nine symptoms are comprised of different or opposite symptoms (insomnia *or* hypersomnia, psychomotor agitation *or* retardation, weight/appetite loss *or* weight gain, feelings of worthlessness *or* inappropriate guilt). This leads to 1,497 potential unique symptom profiles that all qualify for the same diagnosis (Ostergaard *et al.* 2011), including profiles that do not have a single symptom in common.

Although variability in depression symptoms has been documented both across (Katschnig *et al.* 1986; Baumeister & Parker, 2012) and within individuals (Coryell *et al.* 1994; Oquendo *et al.* 2004), the DSM and depression screening instruments such as the Beck Depression Inventory (Beck *et al.*, 1988) and the Hamilton Rating Scale for Depression (Hamilton, 1960) build sum-scores of depressive symptoms that are used to distinguish healthy and depressed individuals. Sum-scores rely tacitly on the DSM-inherent assumption of symptom equivalence (Lux & Kendler, 2010): since criterion symptoms contribute equally to a diagnosis, they are considered to be equivalent to each other and thus interchangeable. Symptom equivalence, in turn, is the result of conceptualizing depression within the framework of reflective latent variable modeling (Bollen, 1989; Schmittmann *et al.* 2011): variation in the latent variable (depression) is understood to be the *cause* of variation of its observable indicators (DSM depressive symptoms), and as long as sufficient symptoms are present, a diagnosis can be given reliably.

Pronounced symptomatic variability of depression has led to a large number of proposed subtypes (e.g., psychotic depression, neurotic depression, anxious depression, and melancholic depression). However, no agreement has yet been reached as to the number or validity of these subtypes (Lichtenberg & Belmaker, 2010; Baumeister & Parker, 2012). And while symptomatic heterogeneity has been explored by extracting principle components or latent factors, factor solutions differ markedly across and within clinical screening instruments for depression (Shafer, 2006; Furukawa *et al.* 2005), appear to be arbitrary (Wood *et al.* 2010), not invariant across demographic variables like ethnicity (Crockett *et al.* 2005), sex (Baas *et al.* 2011), or age (Williams *et al.* 2007), and vary depending on the extraction method (Widaman, 1993). Due to these potential limitations of methods that reduce symptom space, this report examines depressive symptoms individually.

It has been suggested that individual symptoms are separate entities that have autonomous causal relevance (Staner, 2010; Borsboom *et al.* 2011; Schmittmann *et al.* 2013). This idea is supported by evidence that depressive symptoms are associated with specific genes in two small genetic reports (Jang *et al.* 2004; Myung *et al.* 2012), and by a larger study of 7,500 twins concluding that the DSM symptomatic criteria for depression do not reflect a single underlying genetic factor (Kendler *et al.* 2013). It has also been suggested that different symptoms might have different risk factors reflecting different etiologies (Hasler *et al.* 2004; Cramer *et al.* 2010; Hasler & Northoff, 2011), but much remains to be done to test the validity of the assumption of symptom equivalence.

Identified risk factors for depression include demographic variables such as age and sex (Piccinelli & Wilkinson, 2000; Kendler *et al.* 2005), personality traits such as neuroticism (Angst & Clayton, 1986; Kendler *et al.* 2004), early life adversity (Gilmer & McKinney, 2003; Gutman & Nemeroff, 2003), previous episodes of depression (Beekman *et al.* 1995; Colman *et al.* 2011), family history of depression (Nierenberg *et al.* 2007), and stressful life events (Mazure, 1998; Paykel, 2003). If all depressive symptoms are caused by an underlying disorder, symptoms should have similar risk factors because risk factors only influence the liability to develop depression, and depression in turn causes the symptoms. If, however, risk factors differ for different depression symptoms, this suggests possible substantial benefits from analyses of individual symptoms.

A recent comprehensive study (Lux & Kendler, 2010) investigated cross-sectional associations of all nine DSM criterion symptoms with 25 variables, including demographic information, personality traits, life events, history of depression, and lifetime comorbidities in a sample of 1015 individuals. The results revealed a complex association pattern. For instance, four symptoms were associated with years of education (sleep changes and fatigue with more education, psychomotor problems and suicidal ideation with less education), two symptoms were associated with family income (depressed mood with lower income, concentration problems with higher income), and two symptoms (depressed mood and psychomotor problems) had a significant positive correlation with current age. Lux and Kendler (2010) concluded that the surprising degree of covert heterogeneity is difficult to reconcile with the assumption of symptom equivalence.

Longitudinal investigation of a population that shifts from low to elevated depression levels allows a more detailed examination of the influence of risk factors on specific symptoms. Here, we use medical residency as a prospective stress model. Medical residency is known to be an intense stressor (Butterfield, 1988; Duffy, 2005), with residents facing long work hours, sleep deprivation, loss of autonomy, as well as extreme emotional situations (Shanafelt & Habermann, 2002). In a previous longitudinal study of medical residents, depression levels increased from 3.9% at baseline to 25.7% during residency (Sen *et al.* 2010). Medical residency thus offers the rare opportunity to assess depressive symptoms and risk factors shortly before the onset of a severe and chronic stressor that drastically increases depression symptoms, allowing us to test the null hypothesis that increases in DSM depression criterion symptoms are predicted by similar risk factors.

Methods

Sample

4005 interns entering residency programs in the USA during the 2009-2011 academic years were invited to participate in the study; fifty-eight percent (2455/4005) accepted the invitation. The institutional review boards at participating hospitals approved the study. Participating subjects provided electronic informed consent, and were given \$50 in gift certificates.

Assessment

All surveys were conducted through a secure online website designed to maintain confidentiality. Depressive symptoms were measured using the Patient Health Questionnaire (PHQ-9) (Spitzer *et al.* 1999). The PHQ-9 is a self-report component of the PRIME-MD inventory designed to screen for the DSM-IV (American Psychiatric Association, 2000) criterion symptoms of depression (DSM-IV MDD symptoms are identical to the DSM-5 symptoms). For each of the nine symptoms, subjects indicated whether, during the previous 2 weeks, the symptom had bothered them “not at all,” “several days,” “more than half the days,” or “nearly every day.” Each item yields a score of 0, 1, 2 or 3. The nine symptoms assessed by the PHQ-9 are: little interest or pleasure in doing things (interest), feeling depressed or hopeless (depressed), sleep problems (sleep), feeling tired (fatigue), appetite problems (appetite), feeling bad about yourself/that you are a failure (self-blame), trouble concentrating on things (concentration), moving or speaking slowly/being fidgety or restless (psychomotor), and suicidal ideation (suicide).

Subjects completed a baseline survey 1-2 months prior to commencing internship that assessed general demographic factors (age, sex), personal factors (baseline PHQ-9 depressive symptoms, self-reported history of depression, self-reported family history of depression, childhood stress (Risky Families Questionnaire (Taylor *et al.* 2006)), and neuroticism (NEO-Five Factor Inventory (Costa & McCrae, 1997)).

Participants were contacted via email at months 3, 6, 9 and 12 of their internship year and asked to complete the PHQ-9 again. They were also queried regarding work hours over the past week and the occurrence of a series of stressful non-internship life events (marriage, childbirth, serious illness, death or serious illness in close family or friend, financial problems, end of a serious relationship, or becoming a victim of crime or domestic violence,) during the past three months. Because the number of life events reported in each category was too low to use in subsequent analyses, we computed a sum score of all events for each measurement point per subject and subsequently used it as predictor.

Statistical Analysis

We compared symptom severity at baseline with average symptom severity during the four measurements across the residency. This approach has been used in previous publications based on this dataset (Sen *et al.* 2010; Sen *et al.* 2013), and has the advantage of increased reliability of symptom assessment within-internship through repeated measurement. When averaging the within-internship symptom scores, 1166 of the 2455 subjects (47.5%) were

dropped via listwise deletion because they had missing data on two or more timepoints, leaving 1289 subjects in the analytic sample. Average scores for the two risk factors that were assessed after baseline (work hours and number of stressful life events) were constructed analogous to symptoms.

Overall, three analyses were conducted. First, we used a series of paired *t*-tests comparing baseline levels of each symptom to the average symptom level during residency to determine which symptoms increased significantly.

Second, we investigated whether risk factors differentially predicted symptom change over time. We estimated a structural equation model (SEM) encompassing nine linear regressions, one regression per symptom. In each regression, severity of a given symptom during residency was predicted by five baseline risk factors (sex, history of depression, family history of depression, neuroticism, and childhood stress) and two within-internship risk factors (work hours and number of stressful events), controlling for baseline severity of the symptom. To avoid overfitting, we used identical predictors in all nine regression, and did not drop insignificant predictors to reach the best possible fit for each symptom. We did, however, decide to exclude predictors from the analysis in case they were not related with *any* of the nine symptoms. We then compared two models: in the *homogeneity* model, regression weights were constrained to be equal for each risk factor across all symptoms. Specifically, a regression coefficient was estimated for each risk factor separately, but this coefficient was constrained to have the same effect on each symptom. The homogeneity model represents the hypothesis that depression is the common cause for all symptoms, with each risk factor affecting all symptoms equally. In the *heterogeneity* model, all regression coefficients were freely estimated. The models were compared using a χ^2 test, and model fit was assessed using the root mean square error of approximation (RMSEA; RMSEA = 0.06 indicating good fit) as well as the comparative fit index (CFI; CFI = 0.95 indicating good fit) (Hu & Bentler, 1999). If the constrained homogeneity model would fit worse than the heterogeneity model, this would indicate that risk factors vary in their effects across symptoms. Both models were estimated using the Maximum Likelihood Estimator, and results of the model with better fit were visualized using the R-package *qgraph* (Epskamp *et al.* 2011).

While our second analysis provides information about potential heterogeneity on the basis of observable symptoms, it does not rule out the possibility that heterogeneity disappears in a latent disease model. Thus, we investigated whether controlling for a latent depression variable eliminates heterogeneity, using a longitudinal multiple-indicator multiple-cause (MIMIC) model. MIMIC models (Jöreskog & Goldberger, 1975) encompass one or more latent variables that have both a number of observable indicators (in our case symptoms) and variables that influence the latent (in our case risk factors). Following the approach described by Jones (2006), we compared two versions of a MIMIC model. The first model (model I) allows for risk factors to directly influence symptoms, while the second model (model II) represents the hypothesis that risk factors only affect the latent depression variable which, in turn, influences individual depressive symptoms (Figure 1). Evidence that these direct paths improve model fit significantly would indicate that symptom level effects of risk factors exist and are not mediated by a latent variable. Model I was constructed in an

iterative process. First, all risk factors were allowed to have direct effects on all symptoms, except for the last PHQ-9 symptom suicidal ideation (for the purpose of model identification). Second, non-significant direct paths were removed and significant ones added for suicidal ideation until only significant paths from risk factors to symptoms remained. Model comparison and fit of the two MIMIC models were assessed analogous to analysis two. A complication to the MIMIC model is the longitudinal nature of our data. Similar to our second analysis, we thus predicted the second measurement point by risk factors, controlling for the first. While work hours and number of stressful life events that were assessed during residency naturally only affected the within-internship latent, baseline risk factors (sex, history of depression, neuroticism, and childhood stress) were allowed to additionally affect the baseline latent variable. In both models, the residual of each symptom at baseline was allowed to be correlated with the residual of the same symptom during residency. Both MIMIC models were estimated using the Maximum Likelihood Estimator.

Analysis one was performed using *R 2.13.0* (R Development Core Team, 2008), analyses two and three were conducted in *MPLUS 7.0* (Muthén & Muthén, 2012).

Results

1289 interns were included in the final analyses. The mean age of study participants was 27.6 ($SD = 3.02$) and 47.9% were male (Table 1). Participants that had to be dropped due to missing values did not differ significantly from the retained participants in sex, neuroticism, childhood stress, personal and family history of depression, work hours, and number of stressful life events (all $p > 0.05$).

Effects of residency on depressive symptoms

All depression symptoms increased significantly during residency training (t (1288) ranging from 6.77 to 40.79, all $p < 0.001$). On average, symptoms increased by 169% of their baseline means, with a range of 48% (sleep problems) to 270% (psychomotor problems).

Impact of risk factors on symptoms

Family history of depression was the only risk factor that was not related to changes of *any* of the nine symptoms and was excluded from subsequent analyses. The homogeneity model in which each risk factor was constrained to have equal impact on all symptoms fit the data significantly worse than the heterogeneity model in which the effects of risk factors on symptoms were freely estimated ($p < 0.001$). The heterogeneity model fit the data well, and the highly significant χ^2 difference test indicated strong differential impact of risk factors on symptoms (Table 2).

Full results of the heterogeneity model are summarized in Table 3, and visualized in Figure 2. Of the various findings, three are particularly worthy of note. First, four risk factors (prior history of depression, higher number of stressful life events, childhood stress, and sex) predicted worsening of unique subsets of three to eight of the symptoms. Second, two predictors (neuroticism and work hours) had significant impact on all nine PHQ-9 symptoms over time. Third, female residents showed increases in sleep problems, appetite problems, and fatigue during residency, while male residents reported increased suicidal ideation.

To assess whether the two uniform predictors neuroticism and work hours varied in their predictive strength across symptoms, we compared confidence intervals (CI) of the regression coefficients across symptoms. Of 36 symptom pairs, 14 CI (38.9%) did not overlap for neuroticism, and 14 CI (38.9%) did not overlap for work hours (see supplementary materials). This implies considerable variability in the predictive strength of neuroticism and work hours across symptoms.

Latent disease model

In this analysis, we compared the fit of two longitudinal MIMIC models. Since preliminary analyses revealed that factor loadings were not invariant across time, we allowed factor loadings to vary across measurement points in both MIMIC models. Due to high modification indices, residuals of the following symptoms were allowed to be correlated in both models: symptoms sleep and fatigue at baseline, and symptoms depressed and appetite, depressed and self-blame, fatigue and appetite, as well as concentration and psychomotor within-internship. As result of the iterative fitting process described in the Methods section, model I retained 19 significant direct effects of risk factors on symptoms after controlling for the latent depression factor.

Results of fitting the two models to the data are presented in Table 2. The χ^2 difference test indicated that the 19 direct paths from risk factors to symptoms in model I improved the fit significantly ($p < 0.001$). This means that heterogeneity of risk factors and symptoms cannot be explained by a latent depression variable.

Discussion

In this study, we tested the hypothesis that the nine DSM-5 depression symptoms have similar risk factors. The evidence from a longitudinal population cohort of 1289 medical interns is inconsistent with this hypothesis. Instead, the results support the alternative hypothesis that risk factors have differential impact on depression symptoms. Four risk factors predicted increases of unique subsets of symptoms. Neuroticism and work hours did predict increases of all symptoms, albeit to varying magnitudes. We also tested whether the pronounced heterogeneity could be explained by a latent depression factor; this was not the case.

These results support the notion that depressive symptoms have different characteristics and properties. Studies of depression typically use sum-scores to describe severity, a strategy that may obfuscate crucial information about the nature of depression symptoms and causes. For instance, in this report, women were more likely to report worsening of sleep and appetite problems as well as fatigue during internship, whereas men reported increased suicidal ideation – information that would not have been available from assessing depression sum-scores alone. Furthermore, the use of sumscores groups individuals with dramatic inter-individual symptom differences together in the same diagnostic group. This covert heterogeneity might help to explain recent studies reporting “disappointing” results such as questionable reliability of depression diagnosis (Regier *et al.* 2012), and the striking lack of success of GWAS and other genetic studies in identifying loci definitively associated with depression and treatment response (Tansey *et al.* 2012; Hek *et al.* 2013; see also: Power *et*

al. 2012). Our results are consistent with the growing chorus of voices suggesting that covert heterogeneity of DSM diagnostic criteria in research may be inhibiting progress in elucidating the biological roots of mental illness (Lux & Kendler, 2010; Kendler *et al.* 2013; National Institute of Mental Health, 2013).

Assessing depression symptoms individually promises to reveal new findings. Special attention is needed for the four compound symptoms (insomnia or hypersomnia, psychomotor agitation or retardation, weight loss or weight gain, feelings of worthlessness or inappropriate guilt). For example, chronic stress is more strongly associated with weight gain than weight loss (Keller *et al.* 2007), and it is plausible that these opposite symptoms might arise from different etiologies. In addition, the symptoms chosen as criteria for DSM depression are a small subset of possible depression symptoms (McGlinchey *et al.* 2006; Zimmerman *et al.* 2006a) and were determined largely by clinical consensus instead of empirical evidence (Zimmerman *et al.* 2006b; Kendler & Zachar, 2008; Lux & Kendler, 2010). For instance, helplessness and hopelessness as compound symptom have been shown to perform more strongly than some of the DSM-IV criterion symptoms in distinguishing depressed from healthy individuals (McGlinchey *et al.* 2006).

Furthermore, prior work shows that different life stressors provoke different symptoms in patterns that are somewhat consistent (Keller & Nesse, 2005; Keller & Nesse, 2006; Keller *et al.* 2007; Cramer *et al.* 2011), supporting the need for close attention to heterogeneity of depression symptoms.

While it is unclear at this time why risk factors predict certain depressive symptoms, research informed about differences between individual symptoms will facilitate the search for etiologically and symptomatically homogenous groups, and thus serve our main goal of increasing treatment efficacy.

Limitations

These findings need to be interpreted in the light of four limitations. First, medical residents are relatively homogenous in terms of age, education, and socioeconomic status, so our results may not generalize to other populations.

Second, the PHQ-9 does not contain detailed information about bi-directional symptoms. It is thus unclear whether, for instance, the increasingly severe sleep problems in female residents manifested themselves as insomnia or hypersomnia, or whether the increases of psychomotor problems reported by interns that experienced a higher number of stressful life events were due to psychomotor agitation or psychomotor retardation.

Third, in preliminary analyses conducted prior to MIMIC modeling, a one factor solution emerged for the first measurement point, while two factors fit the second measurement point better. Seeing that the MIMIC models were not interpretable from a substantive point of view with different numbers of factors across time, we estimated one factor per measurement point, but allowed factor loadings to be freely estimated. This approach penalizes the fit of both MIMIC models equally and should not favor one over the other.

Fourth, because multivariate models that contained nine dependent variables, six risk factors, five measurement points, and controlled for a latent depression factor were not interpretable, our analyses are limited to two timepoints. As a result, if the relationship between specific variables changed during internship, these changes may have been masked. However, correlations of symptoms with time-varying risk factors at each individual timepoint were fairly stable across the four post stress onset measurements, and comparable to the averaged timepoint used in our analyses. Our approach therefore does not seem to substantially distort the relationships of these variables.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Martin Schultze and Dr Kerby Shedden for their valuable statistical input.

Eiko Fried is supported by fellowships from the Cluster of Excellence “Languages of Emotion” (grant no. *EXC302*), the German Research Foundation, and the Dahlem Research School Berlin. Data collection was supported by *MH095109* from the National Institute of Mental Health.

Funding information: Cluster of Excellence “Languages of Emotion” (LOE) - *EXC302*; National Institute of Mental Health (NIMH) - *MH095109*.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed.. American Psychiatric Association; Washington, DC: 2000.
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed.. American Psychiatric Association; Washington, DC: 2013.
- Angst J, Clayton P. Premorbid personality of depressive, bipolar, and schizophrenic patients with special reference to suicidal issues. *Comprehensive Psychiatry*. 1986; 27:511–532. [PubMed: 3780193]
- Baas KD, Cramer AOJ, Koeter MWJ, van de Lisdonk EH, van Weert HC, Schene AH. Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). *Journal of Affective Disorders*. 2011; 129:229–35. [PubMed: 20888647]
- Baumeister H, Parker JD. Meta-review of depressive subtyping models. *Journal of Affective Disorders*. 2012; 139:126–140. [PubMed: 21885128]
- Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: 25 years of evaluation. *Clinical Psychology Review*. 1988; 8:77–100.
- Beekman ATF, Deeg DJH, van Tilburg T, Smit JH, Hooijer C, van Tilburg W. Major and minor depression in later life: a study of prevalence and risk factors. *Journal of Affective Disorders*. 1995; 36:65–75. [PubMed: 8988267]
- Bollen, KA. Structural equations with latent variables. Wiley; New York: 1989.
- Borsboom D, Cramer AOJ, Schmittmann VD, Epskamp S, Waldorp LJ. The small world of psychopathology. *PloS one*. 2011; 6:e27407. [PubMed: 22114671]
- Butterfield PS. The stress of residency: a review of the literature. *Archives of Internal Medicine*. 1988; 148:1428–1435. [PubMed: 3288162]
- Colman I, Naicker K, Zeng Y, Ataullahjan A, Senthilselvan A, Patten SB. Predictors of long-term prognosis of depression. *Canadian Medical Association Journal*. 2011; 183:1969–1976. [PubMed: 22025655]
- Coryell W, Winokur G, Shea T, Maser JD, Endicott J, Akiskal HS. The long-term stability of depressive subtypes. *American Journal of Psychiatry*. 1994; 151:199–204. [PubMed: 8296889]

- Costa PT Jr, McCrae RR. Stability and change in personality assessment: the revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment*. 1997; 68:86–94. [PubMed: 9018844]
- Cramer AOJ, Waldorp LJ, van der Maas HLJ, Borsboom D. Comorbidity: a network perspective. *The Behavioral and Brain Sciences*. 2010; 33:137–150. [PubMed: 20584369]
- Cramer AOJ, Borsboom D, Aggen SH, Kendler KS. The pathoplasticity of dysphoric episodes: differential impact of stressful life events on the pattern of depressive symptom inter-correlations. *Psychological Medicine*. 2011; 42:957–965. [PubMed: 22093641]
- Crockett LJ, Randall BA, Shen Y-L, Russell ST, Driscoll AK. Measurement equivalence of the center for epidemiological studies depression scale for Latino and Anglo adolescents: a national study. *Journal of Consulting and Clinical Psychology*. 2005; 73:47–58. [PubMed: 15709831]
- Duffy TP. Glory days: what price glory? *The Pharos of Alpha Omega Alpha-Honor Medical Society*. 2005; 68:22–30. [PubMed: 16342456]
- Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. QGRAPH: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*. 2012; 48:1–18.
- Furukawa TA, Streiner DL, Azuma H, Higuchi T, Kamijima K, Kanba S, Ozaki N, Aoba A, Murasaki M, Miura S. Cross-cultural equivalence in depression assessment: Japan–Europe–North American study. *Acta Psychiatrica Scandinavica*. 2005; 112:279–85. [PubMed: 16156835]
- Gilmer WS, McKinney WT. Early experience and depressive disorders: Human and non-human primate studies. *Journal of Affective Disorders*. 2003; 75:97–113. [PubMed: 12798250]
- Gutman DA, Nemeroff CB. Persistent central nervous system effects of an adverse early environment: Clinical and preclinical studies. *Physiology and Behavior*. 2003; 79:471–478. [PubMed: 12954441]
- Hamilton M. A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry*. 1960; 23:56–62.
- Hek K, Demirkan A, Lahti J, Terracciano A, Teumer A, Cornelis MC, Amin N, Bakshis E, Baumert J, Ding J, Liu Y, Marcianti K, Meirelles O, Nalls MA, Sun YV, Vogelzangs N, Yu L, Bandinelli S, Benjamin EJ, Bennett DA, Boomsma D, Cannas A, Coker LH, de Geus E, De Jager PL, Diez-Roux AV, Purcell S, Hu FB, Rimm EB, Hunter DJ, Jensen MK, Curhan G, Rice K, Penman AD, Rotter JJ, Sotoodehnia N, Emeny R, Eriksson JG, Evans DA, Ferrucci L, Fornage M, Gudnason V, Hofman A, Illig T, Kardina S, Kelly-Hayes M, Koenen K, Kraft P, Kuningas M, Massaro JM, Melzer D, Mulas A, Mulder CL, Murray A, Oostra BA, Palotie A, Penninx B, Petersmann A, Pilling LC, Psaty B, Rawal R, Reiman EM, Schulz A, Shulman JM, Singleton AB, Smith AV, Sutin AR, Uitterlinden AG, Völzke H, Widen E, Yaffe K, Zonderman AB, Cucca F, Harris T, Ladwig KH, Llewellyn DJ, Räikkönen K, Tanaka T, van Duijn CM, Grabe HJ, Launer LJ, Lunetta KL, Mosley TH Jr, Newman AB, Tiemeier H, Murabito J. A Genome-Wide Association Study of Depressive Symptoms. *Biological Psychiatry*. 2013; 73:667–678. [PubMed: 23290196]
- National Institute of Mental Health. Transforming Diagnosis. Apr 29. 2013 Retrieved from <http://www.nimh.nih.gov/about/director/2013/transforming-diagnosis.shtml>
- Jang KL, Livesley WJ, Taylor S, Stein MB, Moon EC. Heritability of individual depressive symptoms. *Journal of Affective Disorders*. 2004; 80:125–133. [PubMed: 15207925]
- Jöreskog K, Goldberger A. Estimation of a model of multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*. 1975; 10:631–639.
- Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling. *Medical Care*. 2006; 44:124–133. [PubMed: 16434911]
- Hasler G, Drevets WC, Manji HK, Charney DS. Discovering endophenotypes for major depression. *Neuropsychopharmacology*. 2004; 29:1765–1781. [PubMed: 15213704]
- Hasler G, Northoff G. Discovering imaging endophenotypes for major depression. *Molecular psychiatry*. 2011; 16:604–619. [PubMed: 21602829]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55.

- Katschnig, H.; Pakesch, G.; Egger-Zeidner, E. Life stress and depressive subtypes: A review of present diagnostic criteria and recent research results. In: Katschnig, H., editor. *Life events and psychiatric disorders: Controversial issues*. Cambridge University Press; Cambridge: 1986. p. 201-245.
- Keller MC, Neale MC, Kendler KS. Association of different adverse life events with distinct patterns of depressive symptoms. *American Journal of Psychiatry*. 2007; 164:1521–1529. [PubMed: 17898343]
- Keller MC, Nesse RM. Is low mood an adaptation? Evidence for subtypes with symptoms that match precipitants. *Journal of Affective Disorders*. 2005; 86:27–35. [PubMed: 15820268]
- Keller MC, Nesse RM. The evolutionary significance of depressive symptoms: different adverse situations lead to different depressive symptom patterns. *Journal of Personality and Social Psychology*. 2006; 91:316–30. [PubMed: 16881767]
- Kendler KS, Kuhn J, Prescott CA. The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry*. 2004; 161:631–636. [PubMed: 15056508]
- Kendler KS, Myers J, Prescott CA. Sex differences in the relationship between social support and risk for major depression: A longitudinal study of opposite-sex twin pairs. *American Journal of Psychiatry*. 2005; 162:250–256. [PubMed: 15677587]
- Kendler, KS.; Zachar, P. The incredible insecurity of psychiatric nosology. In: Kendler, KS.; Parnas, J., editors. *Philosophical Issues in Psychiatry*. The Johns Hopkins University Press; Baltimore, MD: 2008. p. 368-383.
- Kendler KS, Aggen SH, Neale MC. Evidence for Multiple Genetic Factors Underlying DSM-IV Criteria for Major Depression. *American Journal of Psychiatry*. 2013 doi:10.1001/jamapsychiatry.2013.751.
- Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS. The epidemiology of major depressive disorder: results from the national comorbidity survey replication. *Journal of the American Medical Association*. 2003; 289:3095–3105. [PubMed: 12813115]
- Kessler RC, Chiu WT, Demler O, Walters EE. Prevalence, severity, and comorbidity of twelve-month DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry*. 2005; 62:617–627. [PubMed: 15939839]
- Lichtenberg P, Belmaker RH. Subtyping major depressive disorder. *Psychotherapy and Psychosomatics*. 2010; 79:131–135. [PubMed: 20185969]
- Lopez, AD.; Mathers, CD.; Ezzati, M.; Jamiso, DT.; Murray, CJL. *Global Burden of Disease and Risk Factors*. Oxford University Press and The World Bank; New York: 2006.
- Lux V, Kendler KS. Deconstructing major depression: a validation study of the DSM-IV symptomatic criteria. *Psychological Medicine*. 2010; 40:1679–1690. [PubMed: 20059797]
- Mazure CM. Life Stressors as Risk Factors in Depression. *Clinical Psychology: Science and Practice*. 1998; 5:291–313.
- McGlinchey JB, Zimmerman M, Young D, Chelminski I. Diagnosing major depressive disorder VIII: are some symptoms better than others? *The Journal of Nervous and Mental Disease*. 2006; 194:785–90. [PubMed: 17041292]
- Muthén, LK.; Muthén, BO. *Mplus User's Guide*. seventh edition. Muthén & Muthén; Los Angeles, CA: 1998-2012.
- Myung W, Song J, Lim S-W, Won H-H, Kim S, Lee Y, Kang HS, Lee H, Kim J-W, Carroll BJ, Kim DK. Genetic association study of individual symptoms in depression. *Psychiatry Research*. 2012; 198:400–406. [PubMed: 22429480]
- Nierenberg AA, Trivedi MH, Fava M, Biggs MM, Shores-Wilson K, Wisniewski SR, Balasubramani GK, Rush AJ. Family history of mood disorder and characteristics of major depressive disorder: a STAR*D (sequenced treatment alternatives to relieve depression) study. *Journal of Psychiatric Research*. 2007; 41:214–221. [PubMed: 16690084]
- Oquendo MA, Barrera A, Ellis SP, Li S, Burke AK, Grunebaum M, Endicott J, Mann JJ. Instability of symptoms in recurrent major depression: a prospective study. *American Journal of Psychiatry*. 2004; 161:255–261. [PubMed: 14754774]

- Ostergaard SD, Jensen SOW, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatrica Scandinavica*. 2011; 124:495–496. [PubMed: 21838736]
- Paykel ES. Life events and affective disorders. *Acta Psychiatrica Scandinavica*. 2003; 108:61–66. [PubMed: 12807378]
- Piccinelli M, Wilkinson G. Gender differences in depression. Critical review. *British Journal of Psychiatry*. 2000; 177:486–492. [PubMed: 11102321]
- Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. *Psychotherapy and Psychosomatics*. 2010; 79:267–279. [PubMed: 20616621]
- Power RA, Keers R, Ng MY, Butler AW, Uher R, Cohen-Woods S, Ising M, Craddock N, Owen MJ, Korszun A, Jones L, Jones I, Gill M, Rice JP, Hauser J, Henigsberg N, Maier W, Zobel A, Mors O, Placentino AS, Rietschel M, Souery D, Kozel D, Preisig M, Lucae S, Binder EB, Aitchison KJ, Tozzi F, Muglia P, Breen G, Craig IW, Farmer AE, Müller-Myhsok B, McGuffin P, Lewis CM. Dissecting the genetic heterogeneity of depression through age at onset. *American journal of medical genetics. Part B, Neuropsychiatric Genetics: the Official Publication of the International Society of Psychiatric Genetics*. 2012; 159B:859–868.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna: 2008.
- Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, Kupfer DJ. DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. *American Journal of Psychiatry*. 2012; 170:59–70. [PubMed: 23111466]
- Schmittmann VD, Cramer AOJ, Waldorp LJ, Epskamp S, Kievit RA, Borsboom D. Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*. 2013; 31:43–53.
- Sen S, Kranzler HR, Krystal JH, Speller H, Chan G, Gelernter J, Guille C. A prospective cohort study investigating factors associated with depression during medical internship. *Archives of General Psychiatry*. 2010; 67:557–565. [PubMed: 20368500]
- Sen S, Kranzler HK, Didwania AK, Schwartz AC, Amarnath S, Kolars JC, Dalack GW, Nichols B, Guille C. Effects of the 2011 Duty Hour Reforms on Interns and Their Patients: A Prospective Longitudinal Cohort Study. *Journal of the American Medical Association Internal Medicine*. 2013; 173:657–662. [PubMed: 23529201]
- Shafer AB. Meta-analysis of the Factor Structures of Four Depression Questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology*. 2006; 62:123–146. [PubMed: 16287149]
- Shanafelt T, Habermann T. Medical residents' emotional well-being. *Journal of the American Medical Association*. 2002; 288:1846–1847. [PubMed: 12377076]
- Solomon DA, Keller MB, Leon AC, Mueller TI, Lavori PW, Shea MT, Coryell W, Warshaw M, Turvey C, Maser JD, Endicott J. Multiple recurrences of major depressive disorder. *American Journal of Psychiatry*. 2000; 157:229–233. [PubMed: 10671391]
- Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study, Primary Care Evaluation of Mental Disorders, Patient Health Questionnaire. *Journal of the American Medical Association*. 1999; 282:1737–1744. [PubMed: 10568646]
- Staner L. Comorbidity of insomnia and depression. *Sleep medicine reviews*. 2010; 14:35–46. [PubMed: 19939713]
- Tansey KE, Guipponi M, Perroud N, Bondolfi G, Domenici E, Evans D, Hall SK, Hauser J, Henigsberg N, Hu X, Jerman B, Maier W, Mors O, O'Donovan M, Peters TJ, Placentino A, Rietschel M, Souery D, Aitchison KJ, Craig I, Farmer A, Wendland JR, Malafosse A, Holmans P, Lewis G, Lewis CM, Stensbøl TB, Kapur S, McGuffin P, Uher R. Genetic predictors of response to serotonergic and noradrenergic antidepressants in major depressive disorder: a genome-wide analysis of individual-level data and a meta-analysis. *PLOS Medicine*. 2012; 9:e1001326. [PubMed: 23091423]
- Taylor SE, Way BM, Welch WT, Hilmert CJ, Lehman BJ, Eisenberger NI. Early family environment, current adversity, the serotonin transporter promoter polymorphism, and depressive symptomatology. *Biological Psychiatry*. 2006; 60:671–676. [PubMed: 16934775]

- Widaman K. Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*. 1993; 28:263–311.
- Williams CD, Taylor TR, Makambi K, Harrell J, Palmer JR, Rosenberg L, Adams-Campbell LL. CES-D four-factor structure is confirmed, but not invariant, in a large cohort of African American women. *Psychiatry Research*. 2007; 150:173–80. [PubMed: 17291596]
- Wood AM, Taylor PJ, Joseph S. Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. *Psychiatry Research*. 2010; 177:120–123. [PubMed: 20207424]
- Zimmerman M, Chelminski I, McGlinchey JB, Young D. Diagnosing major depressive disorder X: can the utility of the DSM-IV symptom criteria be improved? *The Journal of Nervous and Mental Disease*. 2006a; 194:893–897. [PubMed: 17164626]
- Zimmerman M, McGlinchey JB, Young D, Chelminski I. Diagnosing major depressive disorder I: A psychometric evaluation of the DSM-IV symptom criteria. *The Journal of Nervous and Mental Disease*. 2006b; 194:158–163. [PubMed: 16534432]

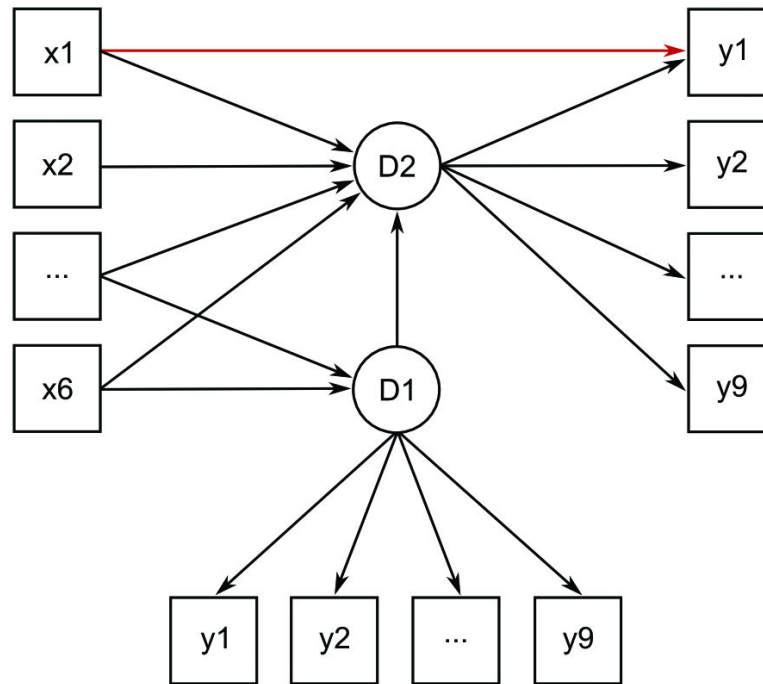


Figure 1. Visualization of the longitudinal MIMIC models used in this study. D1, latent depression factor at baseline; D2, latent depression factor during residency; y1-y9, depressive symptoms; x1-x2, within-internship risk factors; x3-x6, baseline risk factors; the red arrow is an example of a significant direct effect of a risk factor on a symptom (in this case, x1 on y1) that is not mediated by D2.

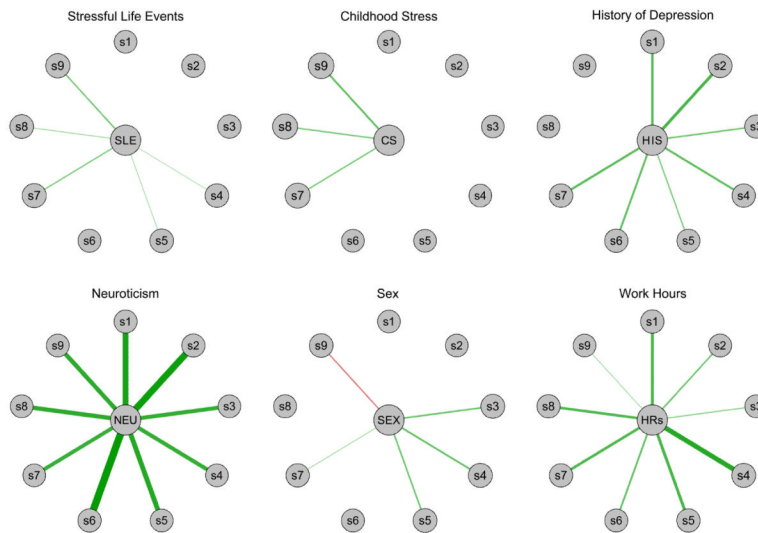


Figure 2. Standardized regression coefficients of six risk factors on change of nine depression symptoms during residency, as estimated by the heterogeneity model. Thickness of lines indicates strength of regression weights. Green lines represent positive regression weights, red lines negative ones; sex was coded 0=female, 1=male. s1, interest; s2, depressed; s3, sleep; s4, fatigue; s5, appetite; s6, self-blame; s7, concentration; s8, psychomotor; s9, suicide.

Table 1
Demographic characteristics of study participants

Variable	Number (%)
Sex	
Male	617 (47.9)
Female	672 (52.1)
Age, y	
25	221 (17.1)
26-30	867 (67.3)
31-35	111 (8.6)
>35	28 (2.2)
History of depression	
Yes	568 (44.1)
No	721 (55.9)
Family history of depression	
Yes	682 (52.9)
No	607 (47.1)
Specialty	
Internal Medicine	481 (37.3)
Other	158 (12.3)
Pediatrics	150 (11.6)
General Surgery	113 (8.8)
Psychiatry	93 (7.2)
Emergency Medicine	81 (6.3)
Family Medicine	62 (4.8)
Obstetrics/gynecology	47 (3.6)
Internal medicine/pediatrics	35 (2.7)
Neurology	32 (2.5)
Transitional	27 (2.1)
Missing	10 (0.8)

Table 2
Goodness-of-fit statistics and χ^2 difference tests for the two model comparisons
(heterogeneity vs. homogeneity, MIMIC models I vs. II)

	χ^2	df	RMSEA	CFI	χ^2_{diff}	df _{diff}	<i>p</i>
Test for heterogeneity							
Model I ^a	197.3	72	0.04	0.98			
Model II ^b	617.4	120	0.06	0.93	420.1	48	< 0.001
MIMIC models							
Model I ^c	868.5	199	0.05	0.94			
Model II ^d	1041.5	218	0.05	0.93	173.0	19	< 0.001

df, degrees of freedom; RMSEA, root mean square error of approximation; CFI, comparative fit index; χ^2_{diff} , χ^2 statistic of the χ^2 difference test; df_{diff}, degrees of freedom of the χ^2 difference test; *p*, *p*-value of the χ^2 difference test;

^a heterogeneity model;

^b homogeneity model;

^c 19 significant direct paths from risk factors to symptoms;

^d no direct paths from risk factors to symptoms.

Table 3
PHQ-9 symptoms predicted by risk factors (heterogeneity model)

	Female sex		Neuroticism		Childhood stress		History of depression		Stressful life events		Work hours		R ²
	β		β		β		β		β		β		
Interest	-0.01		0.29	***	0.05		0.13	***	0.05		0.15	***	0.20
Depressed	0.01		0.32	***	0.04		0.15	***	0.03		0.09	***	0.23
Sleep	0.09	**	0.20	***	0.02		0.09	**	0.04		0.07	**	0.13
Fatigue	0.10	***	0.20	***	0.05		0.11	***	0.05	*	0.24	***	0.20
Appetite	0.09	***	0.24	***	0.04		0.08	**	0.05	*	0.15	***	0.20
Self-blame	0.04		0.36	***	0.03		0.11	***	0.02		0.10	***	0.26
Concentration	0.05	*	0.19	***	0.08	**	0.12	***	0.08	**	0.14	***	0.17
Psychomotor	-0.04		0.22	***	0.08	**	0.03		0.06	*	0.14	***	0.11
Suicide	-0.08	**	0.22	***	0.10	***	0.04		0.08	**	0.05	*	0.14

β , standardized regression coefficients of risk factors on symptoms. R², adjusted R-squared values. Family history of depression is not displayed because none of the effects was significant.

* $p < 0.05$,

**

$p < 0.01$,

*** $p < 0.001$.