Psychological Medicine

cambridge.org/psm

Dynamic risk for first onset of depressive disorders in adolescence: does change matter?

Wenting Mu[1], Kaiqiao Li[2], Yuan Tian[2], Greg Perlman[3], Giorgia Michelini[3,4], David Watson[5], Hans Ormel[6], Daniel N. Klein[7] and Roman Kotov[2,3,4]

[1]Department of Psychology, Tsinghua University, Beijing, China; [2]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, USA; [3]Department of Psychiatry, Stony Brook University, Stony Brook, NY, USA; [4]Semel Institute for Neuroscience & Human Behavior, University of California Los Angeles, Los Angeles, CA, USA; [5]Department of Psychology, University of Notre Dame, Notre Dame, Indiana, USA; [6]Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands and [7]Department of Psychology, Stony Brook University, Stony Brook, NY, USA

Original Article

Cite this article: Mu W, Li K, Tian Y, Perlman G, Michelini G, Watson D, Ormel H, Klein DN, Kotov R (2023). Dynamic risk for first onset of depressive disorders in adolescence: does change matter? *Psychological Medicine* **53**, 2352–2360. https://doi.org/10.1017/S0033291721004190

Received: 31 March 2021
Revised: 27 August 2021
Accepted: 28 September 2021
First published online: 22 November 2021

Keywords:
Chronicity; dynamic risk; escalation; first-onset depression; prediction

Author for correspondence:
Wenting Mu,
E-mail: mwttwm@gmail.com;
Roman Kotov,
E-mail: Roman.Kotov@stonybrook.edu

Abstract

Background. Risk factors for depressive disorders (DD) change substantially over time, but the prognostic value of these changes remains unclear. Two basic types of dynamic effects are possible. The 'Risk Escalation hypothesis' posits that worsening of risk levels predicts DD onset above average level of risk factors. Alternatively, the 'Chronic Risk hypothesis' posits that the average level rather than change predicts first-onset DD.

Methods. We utilized data from the ADEPT project, a cohort of 496 girls (baseline age 13.5–15.5 years) from the community followed for 3 years. Participants underwent five waves of assessments for risk factors and diagnostic interviews for DD. For illustration purposes, we selected 16 well-established dynamic risk factors for adolescent depression, such as depressive and anxiety symptoms, personality traits, clinical traits, and social risk factors. We conducted Cox regression analyses with time-varying covariates to predict first DD onset.

Results. Consistently elevated risk factors (i.e. the mean of multiple waves), but not recent escalation, predicted first-onset DD, consistent with the Chronic Risk hypothesis. This hypothesis was supported across all 16 risk factors.

Conclusions. Across a range of risk factors, girls who had first-onset DD generally did not experience a sharp increase in risk level shortly before the onset of disorder; rather, for years before onset, they exhibited elevated levels of risk. Our findings suggest that chronicity of risk should be a particular focus in screening high-risk populations to prevent the onset of DDs. In particular, regular monitoring of risk factors in school settings is highly informative.

Identification of risk factors for psychopathology is essential for prevention efforts (e.g. defining the group to receive preventive intervention) and etiological models (e.g. providing insights about the processes leading toward psychopathology). The search for risk factors for mental disorders has identified numerous predictors but has generally assumed that risk is static, in that risk factors are typically assessed only once, rather than considering how risk changes with time (Fusar-Poli et al., 2013; Hankin, 2012; Klein, Kotov, & Bufferd, 2011; Nelson, McGorry, Wichers, Wigman, & Hartmann, 2017). However, many risk factors have been shown to change substantially over time (e.g. Roberts, Walton, & Viechtbauer, 2006). It is largely unknown what pattern of change indicates risk for psychopathology. At least two basic types of dynamic relationship are possible between risk factors and onset of psychopathology. The 'Risk Escalation hypothesis' posits that worsening of risk levels predicts disorder onset above the average level of the risk factor. In other words, among people with the same level of risk currently, those who were previously at low risk but worsened are more likely to experience onset than those who were at elevated risk all along. Alternatively, the 'Chronic Risk hypothesis' posits that average risk over time predicts DD onset, and fluctuations around the average are not informative for prediction. These hypotheses have not been systematically compared for any mental disorders. In this study, we seek to demonstrate a strategy for testing these hypotheses on a number of risk factors for adolescent-onset depressive disorders (DD; i.e. major depressive disorder, dysthymic disorder, and depressive disorder not otherwise specified).

Many risk factors have been identified for DD, including malleable vulnerabilities such as symptoms of anxiety and subclinical depression (Klein et al., 2013; Wang et al., 2014), certain personality traits (Bagby, Quilty, & Ryder, 2008; Jeronimus, Kotov, Riese, & Ormel, 2016), and social risk factors (Stice, Ragan, & Randall, 2004). Indeed, these characteristics have been found to change substantially over time (e.g. Roberts et al., 2006; Hankin, 2008; Nocentini, Menesini, & Salmivalli, 2013; Yaroslavsky, Pettit, Lewinsohn, Seeley, & Roberts, 2013; Nelemans, Hale, Branje, Hawk, & Meeus, 2014; Kopala-Sibley, Zuroff, Hankin, & Abela, 2015; Kendler & Aggen, 2017; Bleys, Soenens, Claes, Vliegen, & Luyten, 2018; Fernandes, Davidson, & Guthrie, 2018), especially during adolescence (e.g. Klimstra, Hale, Raaijmakers,

CAMBRIDGE UNIVERSITY PRESS

Branje, & Meeus, 2010). Hence, it is important to consider how change in risk factors predicts DD onset.

The 'Risk Escalation hypothesis' has received support in several longitudinal studies of depression. These studies found that increases in risk levels predict subsequent increases in depression symptoms (Mu, Luo, Rieger, Trautwein, & Roberts, 2019; Steiger, Allemand, Robins, & Fend, 2014) or DD onset (e.g. Laceulle, Ormel, Vollebergh, Van Aken, & Nederhof, 2014). However, these and most other studies tested escalation by analyzing baseline level and subsequent change in risk. This analytic approach cannot compare the two competing hypotheses, because high proximal risk should positively predict depression onset under both Risk Escalation and Chronic Risk scenarios. It would be more informative to compare the change in risk to the risk level most proximal to onset rather than to the distal baseline. The proximal assessment conveys more information about risk than the baseline assessment, which is often years before the proximal assessment. Indeed, past research has shown that the most recent assessment is most predictive of onset when multiple timepoints are available for a risk measure (e.g. Shanahan, Copeland, Costello, & Angold, 2011). In addition, modeling change while controlling for the proximal level of the risk factor is not only a sound analytic practice, but also aligns with clinical decision-making. When forecasting prognosis, clinicians first consider present illness and then its history, a practice best captured in models that include both the proximal assessment and change since baseline.

The alternative 'Chronic Risk hypothesis' has been tested only indirectly. First, research has consistently shown that chronic stressors (e.g. chronic marital stress, chronic illness) are potent predictors of subsequent depression onset (e.g. Bey, Waring, Jesdale, & Person, 2018; Cuijpers, Van Straten, & Smit, 2005; Hammen, Hazel, Brennan, & Najman, 2012). Also, one study reported that adolescents with subclinical depressive symptoms at multiple waves are more likely to develop DD than adolescents with subclinical depressive symptoms at only one assessment (Klein, Shankman, Lewinsohn, & Seeley, 2009). Moreover, some studies have separated the stable portion of risk from temporary fluctuations around it and found that the stable fraction predicted subsequent change in depression (Naragon-Gainey, Gallagher, & Brown, 2013; Kendall, & Langer, 2015) and suicidality (Young et al., 1996). However, these studies did not directly compare the Chronic Risk *v.* Risk Escalation hypotheses. Moreover, most previous studies included only a small number of follow-ups, or failed to distinguish first onsets of depression from recurrent episodes, which confounds vulnerabilities to developing depression with processes that maintain depression after onset (Wilson, Vaidyanathan, Miller, McGue, & Iacono, 2014).

The current study aimed to provide the first direct test of these competing hypotheses – Risk Escalation and Chronic Risk – to predict the first onset of DD, addressing the aforementioned methodological limitations. We utilized data from a richly characterized sample of adolescent girls from the community who underwent five waves of assessment. We did not consider fixed and relatively fixed risk factors, such as childhood maltreatment and parental psychopathology, respectively, and discrete experiences (e.g. life events) which, by definition, cannot evolve. Indeed, most parents who are ever going to develop depression have already done so as most parents were in their 40s when they entered the study. We focused on well-established malleable risk factors for adolescent depression: symptoms of anxiety and subclinical depression (Klein et al., 2013; Wang et al., 2014), three personality traits (neuroticism, conscientiousness, and extraversion; Jeronimus et al., 2016; Mu, Luo, Nickel, & Roberts, 2016), three clinical traits indexing depressogenic cognitive or interpersonal styles (rumination, self-criticism, and dependency; Klein et al., 2011; Mahaffey, Watson, Clark, & Kotov, 2016), and four social risk factors (social support, school engagement, being bullied, and parental criticism; Burkhouse, Uhrlass, Stone, Knopik, & Gibb, 2012; Sachs-Ericsson, Verona, Joiner, & Preacher, 2006; Starr & Davila, 2008; Stice et al., 2004; Swearer, Song, Cary, Eagle, & Mickelson, 2001; Van Voorhees et al., 2008; Wilson et al., 2014).

Method

Participants

Data were collected as part of the Adolescent Development of Emotions and Personality Traits (ADEPT) project. Participants were 550 females aged 13.5–15.5 years at enrollment. This age range was targeted because of the sharp increase in DD incidence in girls during this period (Hankin et al., 1998). The sample was predominantly non-Hispanic White European (80.5%) and socioeconomically diverse (42.2% of families had neither parent with a bachelor's or higher degree). Exclusion criteria were intellectual disability and history of major depressive disorder or dysthymic disorder before enrollment. For the current analyses, we also excluded 44 participants because they developed DD before the second assessment wave and 10 because they were lost to follow-up before that wave; thus, 496 were included in the present analyses. Parents provided permission and adolescents provided assent. The study was approved by the Stony Brook University Institutional Review Board. Further details about the sample and recruitment can be found in Nelson, Perlman, Klein, Kotov, and Hajcak (2016).

Assessments

Participants completed five assessments of DD and risk factors every 9 months for 3 years.

Adolescent depression diagnosis

The Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children (K-SADS-PL; Kaufman et al., 1997) was used to assess DD based on the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV; American Psychiatric Association, 2020) in the interval since the previous assessment. DD included major depressive disorder, dysthymic disorder, and depressive disorder not otherwise specified. Of note, diagnosis of DD not otherwise specified required clinically significant distress or impairment. K-SADS interviews were conducted by staff trained and supervised by clinical psychologists (GP, DK, and RK). Interrater reliability was assessed through an independent rater deriving diagnoses from videotapes of 48 K-SADS interviews and was excellent ($\kappa = 0.81$ for any DD, 0.85 for DYS, 0.62 for DEPNOS, and 0.73 for MDD).

Depression and anxiety symptoms were measured using the expanded version of the Inventory of Depression and Anxiety Symptoms (IDAS-II; Watson et al., 2012). The IDAS-II contains 18 specific scales and a General Depression composite of items from the six depression symptom scales. We selected General Depression and another five scales that were most relevant to risk for DD onset: ill temper, panic, social anxiety, traumatic intrusions, and traumatic avoidance. We did not include the

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press

seven specific depression scales as they are redundant with General Depression, the claustrophobia scale because it did not predict DD onset, and the two mania and three obsessive-compulsive scales as these symptoms were rare in our sample.

Personality was assessed with the Big Five Inventory (BFI; John & Srivastava, 1999), specifically the neuroticism, conscientiousness, and extraversion scales.

Rumination was assessed with the Ruminative Responses Scale (RRS) of the Response Styles Questionnaire (Nolen-Hoeksema, 1987, 1991).

Self-criticism was assessed with Bagby, Parker, Joffe, & Buis's (1994) revised self-criticism subscale of the Depressive Experiences Questionnaire (DEQ; Blatt, D'Affitti, & Quinlan, 1976).

Dependency was measured using the emotional reliance subscale of the Interpersonal Dependency Inventory (IDI; Hirschfeld et al., 1977).

Social support was measured using total score on the Multidimensional Scale of Perceived Social Support (MSPSS; Zimet, Dahlem, Zimet, & Farley, 1988). Participants rated their perceptions of the general adequacy of social support received from family, friends, and a significant other.

School engagement was measured with three subscales of the School Attitude Assessment Survey – Revised (SAAS-R; McCoach & Siegle, 2003). The total score indicated school engagement in terms of attitudes toward school, attitudes toward teachers, and self-motivation/regulation.

Bullying was measured with the total of three victim subscales of the Revised Peer Experiences Questionnaire (RPEQ; De Los Reyes & Prinstein, 2004): overt, relational, and reputational.

Parental criticism was measured with the criticism subscale of the Network of Relationships Inventory (NRI, Furman & Buhrmester, 2009). Participants answered three items each about their mother figure and father figure; the mean of ratings across the two figures was used to index parental criticism. Items were rated for how much each behavior occurred in each relationship.

Specifics of the measures (i.e. rating scale, sample items, number of items, rating time frame, Cronbach's α, and stability coefficients) are presented in online Supplementary Tables S1 and S2.

Statistical analyses

Outcomes were whether or not DD onset occurred in the interval. After first onset, outcomes were censored. We labeled the wave when DD was diagnosed for the first time as Wave$_n$, and analyses tested whether first-onset DD occurred between Wave$_{n-1}$ and Wave$_n$. We refer to Wave$_{n-1}$ as the 'proximal wave', as it is closest to onset, and Wave$_{n-2}$ as the 'baseline wave', as it is the baseline used to calculate the change score. Change was operationalized as proximal minus baseline score. We also calculated the mean score, averaging across all waves preceding the proximal wave. For example, if Wave$_n$ was Wave 5, the model tested whether first onset occurred between Waves 4 and 5, the proximal wave was Wave 4, the baseline wave was Wave 3, and the mean score was based on Wave 1 to Wave 3.

Next, we conducted three sets of Cox regression analyses with time-varying covariates to predict DD onset. Analysis 1 tested the Risk Escalation hypothesis; Analysis 2 tested the Chronic Risk hypothesis; and Analysis 3 tested both hypotheses in the same model. In Analysis 1, we examined the effect of change by entering the change score alongside the proximal score as time-varying predictors. This was needed because change is often confounded with level, in that people with high levels of scores tend to show greater levels of change. However, unlike prior research that invariably controlled for baseline assessment (e.g. Laceulle et al., 2014), we controlled for the proximal score, as the proximal score is the most informative single indicator of risk and therefore controlling for the proximal assessment provides the most rigorous test of dynamic effects. The Risk Escalation hypothesis would be supported by a positive relationship between the change score and first DD onset. In Analysis 2, we entered the mean and proximal scores as time-varying predictors. The Chronic Risk hypothesis would be supported by a positive relationship between the mean score and DD onset. In Analysis 3, we entered change and mean scores simultaneously as time-varying predictors. The change score positively predicting DD onset would support the Risk Escalation hypothesis, whereas the mean score positively predicting onset would support the Chronic Risk hypothesis. These analyses were conducted for each risk factor separately. We calculated the C-statistic for each model, which equals the area under the curve of receiver operating characteristics and indicates the predictive accuracy of the model.

Before completing the Cox regression analyses, we examined the correlations among the predictors (proximal, change, mean) for each risk factor to identify any multicollinearity. We found no substantial multicollinearity, as correlations across risk factors ranged 0.34 to 0.51 for proximal with change, 0.44 to 0.80 for proximal with mean, and −0.42 to −0.03 for change with mean (see Table S4), so in all models tolerance was >0.35, which is well within the acceptable range (Belsley, Kuh, & Welsch, 2005). Moreover, although absolute change is the most interpretable index of change (Rogosa, 1995), some applications use residual scores to represent change. Accordingly, we performed sensitivity analyses, repeating Analysis 1 with the residual score (the baseline score regressed on the proximal score) instead of the change score, and Analysis 3 with the residual score (the proximal score regressed on the mean score) instead of the change score. To confirm that the performance of the mean score is not due to adjustment for other predictors, we also performed analyses with the mean score as the only predictor.

Analyses were carried out using the R 3.5.0 package 'Survival' (2.42-3). We used grand mean standardization – standardizing each variable across subjects and across time – to improve interpretability; thus, the hazards ratio (HR) reflected the difference in risk per standard deviation.

Missing data

For each scale, if fewer than 25% of the items were missing, we used ipsative mean imputation (Schafer & Graham, 2002) to replace the missing data before computing the scale total; otherwise the score was considered missing. Survival analyses used all available data.

Results

Adolescent depression

Sixty-six participants had first onset of DD after Wave 2 (Table 1). The descriptive statistics for each risk factor at each wave are presented in online Supplementary Table S3. The scores of each risk factor from Wave 1 to Wave 4 for participants who had onsets at different waves are presented in Fig. 1. The no-onset group showed low or decreasing levels of risk throughout the interval.

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press

Table 1. Number of first onset of depressive disorders

	Depression type			
Time of onset	DYS	DEPNOS	MDD	Total
Interval 3	4	14	12	30
Interval 4	3	9	4	16
Interval 5	1	7	12	20
Total	5	28	27	66

Note. DYS = Dysthymia; DEPNOS = Major Depressive Not Otherwise Specified; MDDs = Major Depressive Disorder; Interval 3 = Between Waves 2 & 3; Interval 4 = Between Waves 3 & 4; Interval 5 = Between Waves 4 & 5).

In the other groups, the ranking of initial levels of risk generally followed the order of onset, with higher risk in groups that had an earlier onset. However, the trajectories within these groups did not show a clear pattern, increasing in some cases and decreasing in others before onset.

Effect of change while controlling for proximal assessment (Analysis 1)

For all risk factors, the proximal score significantly and positively predicted first-onset DD (Table 2). For eight of 16 risk factors, a *decrease or smaller increase* in risk from baseline to proximal assessment significantly predicted onset after controlling for the proximal score. Non-significant HRs for the other variables were in the same direction. These findings are inconsistent with the Risk Escalation hypothesis as it posits that a *larger increase* in risk predicts onset. Predictive accuracy (C-statistic) of models ranged from 0.53 to 0.73, which is low to moderate.

Effect of mean while controlling for proximal assessment (Analysis 2)

For nine of 16 risk factors, the mean value significantly predicted first DD onset in the expected direction, even controlling for the proximal score (Table 2). Thus, looking back from the pre-onset wave, participants who developed DD had higher risk scores throughout the entire course of the study, consistent with the Chronic Risk hypothesis. When the mean was controlled, the proximal score did not predict onset for the majority (nine of 16) of risk factors. Predictive accuracy (C-statistic) of models ranged from 0.60 to 0.77, which is low to moderate.

Direct comparison of risk escalation and chronic risk hypotheses (Analysis 3)

When mean risk and change in risk were both included in the model, the mean significantly predicted first DD onset for all risk factors (Table 2). In contrast, an increase in risk significantly predicted onset for only four out of 16 risk factors: social anxiety, traumatic intrusions, self-criticism, and bullying. These findings provide consistent support for the Chronic Risk hypothesis for all risk factors, and support for the Risk Escalation hypothesis for only a limited set (25%) of risk factors. Differences in mean and change plotted as a function of subsequent onset status showed the same pattern (Fig. 2). Predictive accuracy (C-statistic) of models ranged from 0.59 to 0.76, which is low to moderate.

Sensitivity analyses

To evaluate the robustness of the findings, we repeated Analysis 1 using the residual score instead of the difference score, and the pattern of significant effects was unchanged (Table S6). We also repeated Analysis 3 with the residual rather than difference score. It produced four additional significant effects for the residual, for a total of eight effects, and all 16 effects for mean risk remained significant (Table S7). Overall, analyses that operationalized change using residual scores were consistent with analyses of change scores and both supported the Chronic Risk hypothesis. Moreover, when change was removed from the model, mean risk continued to predict DD in all 16 models (Table S8).

Discussion

The current study is the first direct and rigorous test of the Risk Escalation and Chronic Risk hypotheses. We demonstrated a general approach to evaluating these hypotheses using a number of malleable risk factors for first DD onset. We found that chronically elevated risk (i.e. the mean across multiple waves up to, but not including, the most proximal assessment) predicted first-onset DD, even when recent escalation (from the next-to-most to the most proximal assessment) was included in the model. This pattern, predicted by the Chronic Risk hypothesis, held across all risk factors examined: prior anxiety and subclinical depression symptoms, personality traits (neuroticism, conscientiousness, extraversion), clinical traits (rumination, dependency, and self-criticism), and social factors (social support, school engagement, bullying, parental criticism). In contrast, the Risk Escalation hypothesis was not supported for the majority (75%) of risk factors, as change in risk from baseline to proximal assessment (a) rarely predicted DD onset above mean risk level and (b) predicted in the opposite direction (i.e. *less change* was associated with increased likelihood of first onset) when proximal risk was controlled.

Our findings shed light on the nature and developmental course of risk for first DD onset. The Risk Escalation hypothesis has intuitive appeal, yet we found minimal support among the risk markers examined; instead the Chronic Risk hypothesis received consistent support. It appears that the likelihood of DD onset reflects the mean level of the risk factor over years, rather than change in the months before onset. This aligns with prior research that found stable levels of a risk factor to be highly predictive (Kendall, & Langer, 2015; Klein et al., 2009; Naragon-Gainey et al., 2013). Overall, these findings suggest that risk tends to be present years before onset, and short-term alterations often reflect transient fluctuations rather than a lasting change. This pattern raises the question of why DD onset had not happened earlier. One possibility is that long-standing risk might make individuals especially vulnerable to precipitating factors, such as a major life event or maturation, and when those occur, DD is triggered (Slavich & Irwin, 2014). It is also possible that long-standing risk factors, especially those related to personality, symptoms, and clinical traits, evoke stress, as a result of a complex interaction between these risk factors and enduring environmental contexts (Kushner, Bagby, & Harkness, 2017; Liu & Alloy, 2010). Future research should examine the interplay of chronic and discrete risk factors (e.g. negative life events) in eliciting first DD onset (Hammen, Kim, Eberhart, & Brennan, 2009). Last but not least, one other possibility is that the impact of a

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press

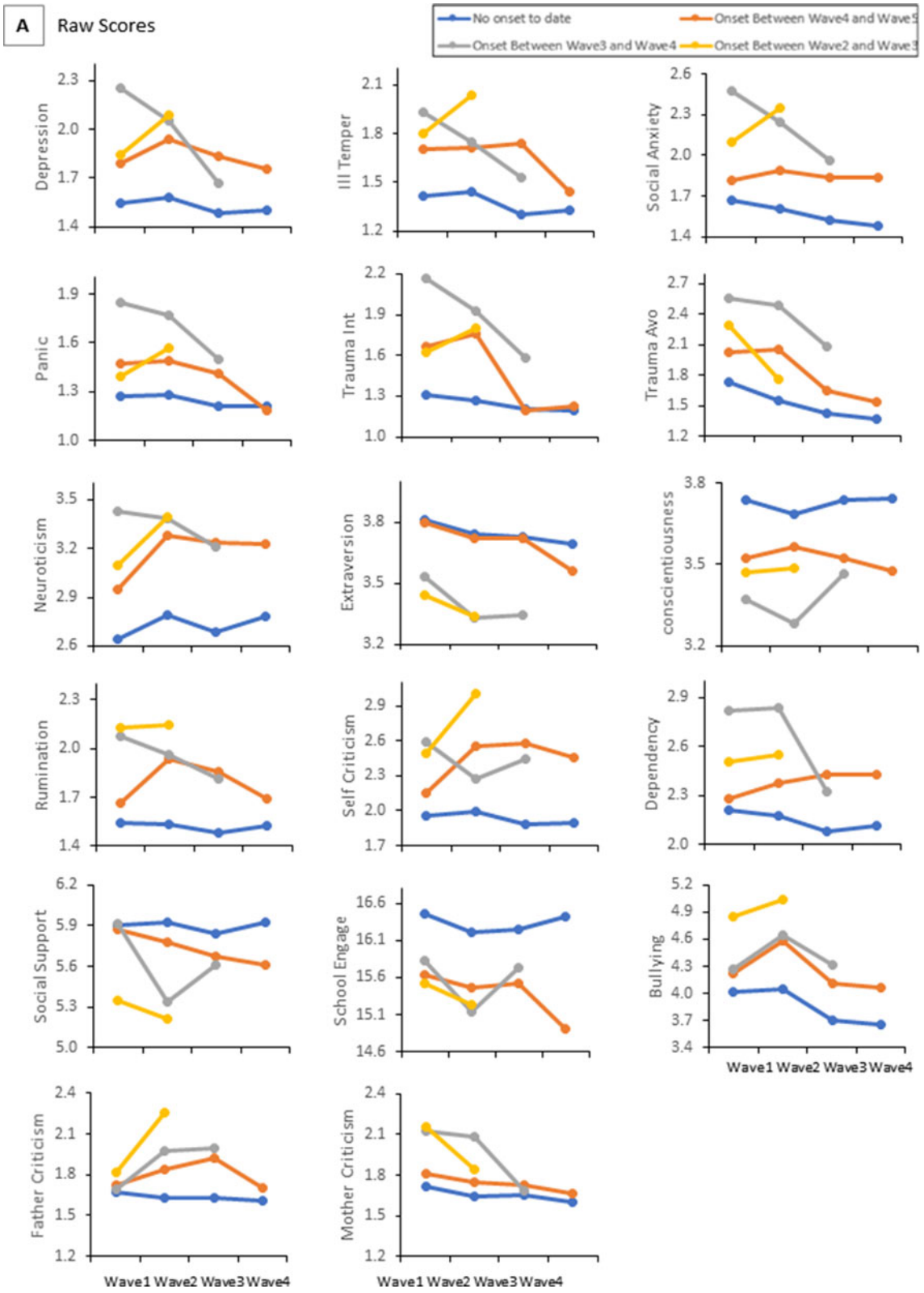

Fig. 1. Raw scores of each risk marker from wave 1 to wave 4 by onset group at different waves.
Note. Trauma Int = TraumaticIntrusion;TraumaAvo = TraumaticAvoidance;

risk factor needs to accumulate until it passes a threshold before first onset is triggered. In other words, people with higher means will have onsets when they are younger than people with lower but still elevated means. In fact, we see this pattern in Fig. 1. Unfortunately, we cannot test it formally in the current study due to lack of enough data points. Studies with more frequent assessments over a longer time span would be ideal to explore this possibility.

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press

Table 2. Hazards ratio of risk factors for first DD onset using three prediction models

	Analysis 1 (proximal + change)			Analysis 2 (proximal + mean)			Analysis 3 (change + mean)		
	Proximal Hazards ratio	Change Hazards ratio	C	Proximal Hazards ratio	Mean Hazards ratio	C	Change Hazards ratio	Mean Hazards ratio	C
Depression and anxiety symptoms									
1. General depression	1.70**	0.76*	0.70	1.15	1.55**	0.72	1.19	1.74**	0.71
2. Ill temper	1.57**	0.81*	0.68	1.25	1.28*	0.68	1.21	1.60**	0.68
3. Social anxiety	1.78**	0.88	0.69	1.49**	1.21	0.68	1.44**	1.76**	0.68
4. Panic	1.44**	0.80	0.66	1.14	1.32**	0.68	1.16	1.52**	0.67
5. Trauma intrusions	1.59**	0.86	0.67	1.31*	1.32**	0.71	1.45**	1.76**	0.70
6. Trauma avoidance	1.59**	0.64**	0.69	1.05	1.64**	0.68	1.04	1.72**	0.68
Personality traits									
7. Neuroticism	2.13**	0.77*	0.69	1.47*	1.48*	0.70	1.30	2.12**	0.70
8. Conscientiousness	0.67**	1.24	0.61	0.93	0.73	0.62	0.97	0.68**	0.62
9. Extraversion	0.68**	1.14	0.62	0.84	0.81	0.62	0.91	0.69**	0.62
Depressogenic cognitive/personality styles									
10. Rumination	2.00**	0.65**	0.72	1.18	1.68**	0.73	1.14	1.94**	0.73
11. Self-criticism	2.05**	0.99	0.73	2.01**	1.03	0.73	1.65**	1.92**	0.73
12. Dependency	1.80**	0.71*	0.65	1.25	1.37	0.65	1.11	1.66**	0.65
Social risk factors									
13. Social support	0.59**	1.35*	0.64	0.78*	0.80*	0.63	0.86	0.67**	0.63
14. School engagement	0.73**	1.22	0.59	0.84	0.85	0.60	0.95	0.76*	0.59
15. Bullying	1.71**	0.80*	0.65	1.36**	1.25*	0.64	1.27*	1.62**	0.64
16. Parental criticism	1.51**	0.94	0.61	1.42*	1.03	0.61	1.27	1.41**	0.60

IDAS-II, Expanded Inventory of Depression and Anxiety Symptoms; C, Concordance Index. Two-sided statistical tests were performed at a level of significance of 5%.
*$p < 0.05$; **$p < 0.01$.

Further support for the Chronic Risk hypothesis stems from our findings that change in risk factors from the baseline to proximal risk assessment significantly and *negatively* predicted first-onset DD when the proximal assessment was controlled, which is in the opposite direction than when the baseline score was controlled (Table S5). In other words, for two people with a given proximal score, the person who had high baseline but de-escalated was at greater risk than the person who had low baseline but escalated. This suggests that participants reverted back to their mean risk levels after the change observed in the proximal assessment (e.g. the person who experienced a recent decrease in the risk factor then returned to high mean level).

Our findings are inconsistent with past evidence supporting the Risk Escalation hypothesis (e.g. Laceulle et al., 2014), possibly due to several factors. First, prior studies controlled for baseline rather than proximal assessment. A model that controls for baseline cannot compare Risk Escalation and Chronic Risk scenarios, because both imply that increase from baseline predicts onset. Indeed, increase from baseline may indicate either a persisting new increase in risk or a return to a high mean risk after a transient improvement.

Second, prior studies assessed risk change over a lengthy period (e.g. 5 years; Laceulle et al., 2014) when more consistent change in risk has accumulated than change during the 9-month intervals examined here. Future studies may clarify the optimal schedule of follow-up assessment intervals to maximize the predictive power of change. That said, we did find some support for the Risk Escalation hypothesis. For four risk factors – social anxiety, traumatic intrusions, self-criticism, and bullying – both change and mean scores independently predicted first-onset DD, providing evidence for both the Chronic Risk and Risk Escalation hypotheses. However, evidence for escalation was inconsistent. Change in three of these four risk factors (i.e. social anxiety, traumatic intrusions, self-criticism) was not significant when the proximal assessment was controlled, and change in bullying changed sign, indicating lower likelihood of DD onset. Therefore, the dynamic effects in these variables require further study.

Of note, for the majority of risk factors, the proximal risk score did not predict DD onset above the mean of previous timepoints. Given that these risk factors are all well-established in the literature, this finding underscores the limitations of cross-sectional risk assessments and suggests that the aggregation of risk over multiple time points can improve prediction (e.g. Fusar-Poli et al., 2013). Echoing calls for dynamic prediction models in psychopathology research (Klein et al., 2011; Nelson et al., 2017), our findings suggest that chronic vulnerability is important to examine in such studies.

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press

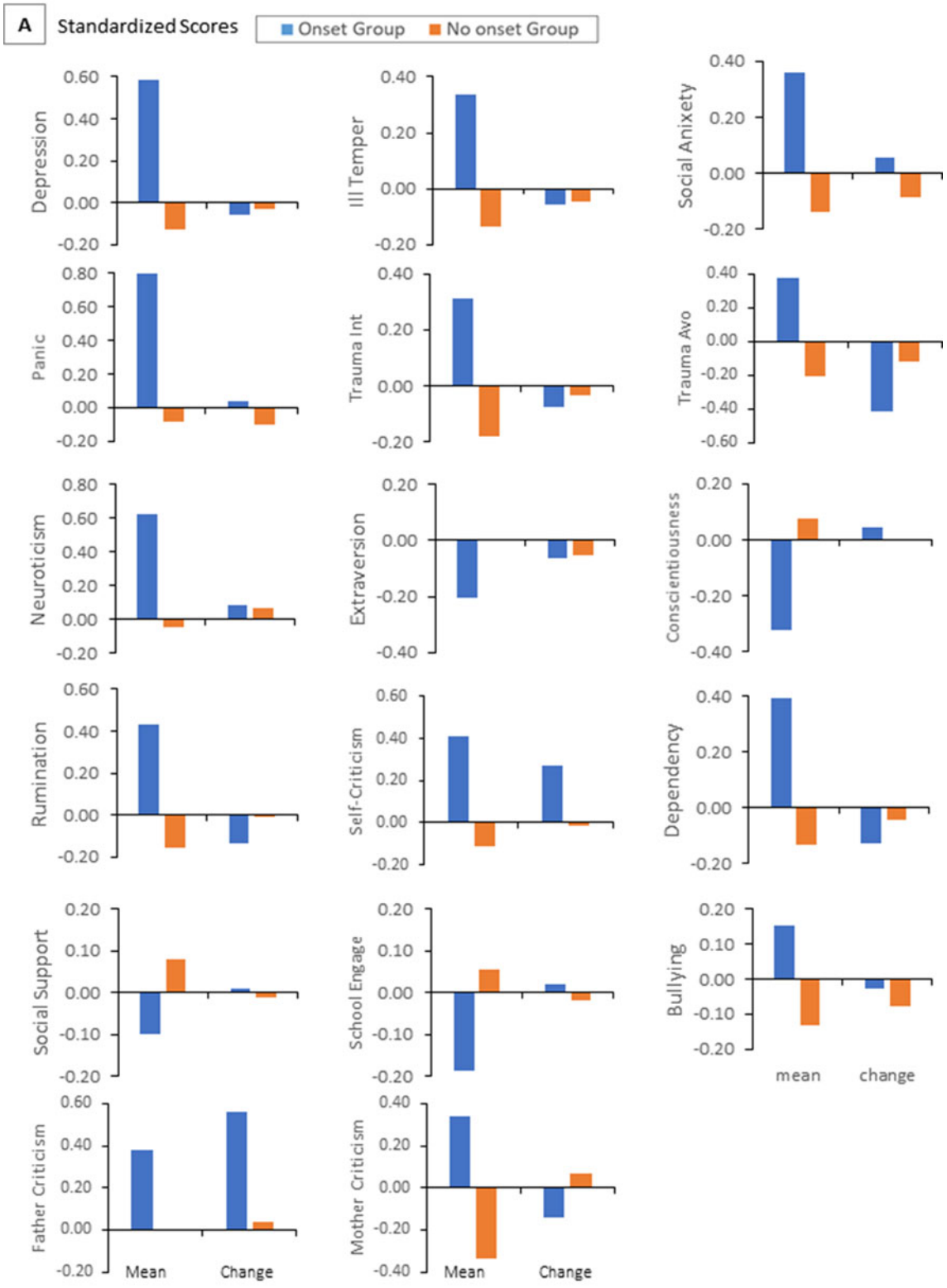

Fig. 2. Standardized mean an change scores of each risk marker by onset group across waves.
Note. The Scores were aggregated over multiple outcome waves (weighted by the number of people in the corresponding group at that time) and were standardized based on the first observation (wave 1). Trauma Int = Traumatic Intrusion; Trauma Avo = Traumatic Avoidance;

Our findings have implications for efforts to screen high-risk populations to prevent depression. Clinicians have been encouraged to consider recent escalation in risk factors for DD (e.g. Steiger et al., 2014). However, we found little support for this recommendation. Instead, prevention efforts should prioritize individuals who have elevated levels of risk on repeated assessments. In clinical practice, clinicians should pay close attention to history of risk and anticipate that recent changes in risk may

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press

revert to the baseline. However, some changes require attention even if they are likely to be transient, such as the emergence of acute stress (e.g. major life events) or those with a high probability of negative consequences (e.g. self-injury).

Our study had several limitations. First, the sample was limited to adolescent girls from the community, so the results may not generalize to males, younger children, adults, or clinical populations. Moreover, 80% of our sample are European White, and future studies should explore if the current findings hold among other populations such as but not limited to Asian Americans, African Americans, etc. Second, similar to most of the literature, when assessing vulnerability, we relied exclusively on self-report inventories. Indeed, self-report provides reasonably accurate assessments of psychopathology (e.g. Babor, Brown, & Del Boca, 1990) and other informants have limited insight into the emotional states of participants. Nevertheless, future studies should improve vulnerability assessment by employing multiple methods (e.g. informant reports, behavioral observations, laboratory measures). Third, some of our risk measures may be assessing the depression prodrome, thereby confounding the risk factor and the outcome. However, we ruled out this confounding effect by controlling for proximal risk, which would have captured any effects of the prodrome on the risk measure. Fourth, we have not examined negative life events, which are a major risk factor for DD, but events are discrete rather than developing over years and require a different analytic approach. Future research should consider the interplay between life events and trajectories of risk factors. Fifth, we operationalized change as a difference score, which can be noisy. Difference scores were used in prior dynamic research (Jacobson & Truax, 1991; Laceulle et al., 2014) and we followed this practice. However, we also repeated analyses with change operationalized as residual scores and obtained very similar results. Other operationalizations of change were not feasible with just five assessment waves, given that onsets could occur during any interval and analyses that use more intervals for modeling dynamic risk would leave only a small window to observe onsets. Future studies should collect more time points and explore more sophisticated analytic methods. Finally, the change rate of risk factors may vary depending upon age. However, we only examined three and a half years of life development, limiting our ability to examine how age interacts with risk factors to predict depression onset. Future studies should include a longer age span to allow for more sophisticated analytical approaches to examine the influence of age on our proposed models.

Our findings indicate that alternative models of dynamic risk are testable and are important targets for research. Across multiple well-established risk factors for DD, chronically elevated levels of risk, rather than recent escalation in vulnerability, best predicted first onsets. Longitudinal studies should consider mean and proximal scores as alternatives to distal baseline and change designs when predicting future outcomes. Our findings suggest that chronicity of risk should be a particular focus in screening high-risk populations to prevent the onset of DDs. In particular, regular monitoring of risk factors in school settings would be highly informative.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291721004190.

Financial support. This work was supported by the National Institute of Mental Health of the National Institutes of Health (R01MH093479 to RK).

Conflict of interest. None.

References

American Psychiatric Association. (2020). *Diagnostic and statistical manual of mental disorders (DSM-IV)*. Am Psychiatric Assoc.

Babor, T. F., Brown, J., & Del Boca, F. K. (1990). Validity of self-reports in applied research on addictive behaviors: Fact or fiction? *Behavioral Assessment*, *12*, 5–31.

Bagby, R. M., Parker, J. D., Joffe, R. T., & Buis, T. (1994). Reconstruction and validation of the Depressive Experiences Questionnaire. *Assessment*, *1*(1), 59–68.

Bagby, R. M., Quilty, L. C., & Ryder, A. C. (2008). Personality and depression. *The Canadian Journal of Psychiatry*, *53*(1), 14–25.

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity* (Vol. *571*). New York: Wiley.

Bey, G. S., Waring, M. E., Jesdale, B. M., & Person, S. D. (2018). Gendered race modification of the association between chronic stress and depression among Black and White US adults. *American Journal of Orthopsychiatry*, *88*(2), 151.

Blatt, S. J., Afflitti, J. P., & Quinlan, D. M. (1976). Experiences of depression in normal young adults. *Journal of Abnormal psychology*, *85*(4), 383.

Bleys, D., Soenens, B., Claes, S., Vliegen, N., & Luyten, P. (2018). Parental psychological control, adolescent self-criticism, and adolescent depressive symptoms: A latent change modeling approach in Belgian adolescents. *Journal of clinical psychology*, *74*(10), 1833–1853.

Burkhouse, K. L., Uhrlass, D. J., Stone, L. B., Knopik, V. S., & Gibb, B. E. (2012). Expressed emotion-criticism and risk of depression onset in children. *Journal of Clinical Child & Adolescent Psychology*, *41*(6), 771–777.

Cuijpers, P., Van Straten, A., & Smit, F. (2005). Preventing the incidence of new cases of mental disorders: A meta-analytic review. *The Journal of Nervous and Mental Disease*, *193*(2), 119–125.

De Los Reyes, A., & Prinstein, M. J. (2004). Applying depression-distortion hypotheses to the assessment of peer victimization in adolescents. *Journal of Clinical Child and Adolescent Psychology*, *33*(2), 325–335.

Fernandes, S., Davidson, J. G., & Guthrie, D. M. (2018). *Changes in social engagement and depression predict incident loneliness among seriously ill home care clients. Palliative and supportive care*, *16*, 170–179.

Furman, W., & Buhrmester, D. (2009). Methods and measures: The network of relationships inventory: Behavioral systems version. *International Journal of Behavioral Development*, *33*(5), 470–478.

Fusar-Poli, P., Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rössler, A., Schultze-Lutter, F., … Valmaggia, L. (2013). The psychosis high-risk state: A comprehensive state-of-the-art review. *JAMA Psychiatry*, *70*(1), 107–120.

Hammen, C., Hazel, N. A., Brennan, P. A., & Najman, J. (2012). Intergenerational transmission and continuity of stress and depression: Depressed women and their offspring in 20 years of follow-up. *Psychological Medicine*, *42*(5), 931–942.

Hammen, C., Kim, E. Y., Eberhart, N. K., & Brennan, P. A. (2009). Chronic and acute stress and the prediction of major depression in women. *Depression and Anxiety*, *26*(8), 718–723.

Hankin, B. L. (2008). Stability of cognitive vulnerabilities to depression: a short-term prospective multiwave study. *Journal of abnormal psychology*, *117*(2), 324.

Hankin, B. L. (2012). Future directions in vulnerability to depression among youth: Integrating risk factors and processes across multiple levels of analysis. *Journal of Clinical Child & Adolescent Psychology*, *41*(5), 695–718.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology*, *107*(1), 128.

Hirschfeld, R. M., Klerman, G. L., Gouch, H. G., Barrett, J., Korchin, S. J., & Chodoff, P. (1977). A measure of interpersonal dependency. *Journal of Personality Assessment*, *41*(6), 610–618.

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19.

Jeronimus, B. F., Kotov, R., Riese, H., & Ormel, J. (2016). Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly

decays with time: A meta-analysis on 59 longitudinal/prospective studies with 443 313 participants. *Psychological Medicine*, *46*(14), 2883–2906.
John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, *2*(1999), 102–138.
Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., … Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988.
Kendall, T., & Langer, A. (2015). Critical maternal health knowledge gaps in low-and middle-income countries for the post-2015 era. *Reproductive Health*, *12*(1), 55.
Kendler, K. S., & Aggen, S. H. (2017). Symptoms of major depression: Their stability, familiality, and prediction by genetic, temperamental, and childhood environmental risk factors. *Depression and anxiety*, *34*(2), 171–177.
Klein, D. K., Glenn, C. R., Kosty, D. B., Seeley, J. R., Rohde, P., & Lewinsohn, P. M. (2013). Predictors of first lifetime onset of major depressive disorder in young adulthood. *Journal of Abnormal Psychology*, *122*, 1–6.
Klein, D. N., Kotov, R., & Bufferd, S. J. (2011). Personality and depression: Explanatory models and review of the evidence. *Annual Review of Clinical Psychology*, *7*, 269–295.
Klein, D. N., Shankman, S. A., Lewinsohn, P. M., & Seeley, J. R. (2009). Subthreshold depressive disorder in adolescents: Predictors of escalation to full-syndrome depressive disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(7), 703–710.
Klimstra, T. A., Hale III, W. W., Raaijmakers, Q. A., Branje, S. J., & Meeus, W. H. (2010). Identity formation in adolescence: Change or stability?. *Journal of Youth and Adolescence*, *39*(2), 150–162.
Kopala-Sibley, D. C., Zuroff, D. C., Hankin, B. L., & Abela, J. R. (2015). The development of self-criticism and dependency in early adolescence and their role in the development of depressive and anxiety symptoms. *Personality and Social Psychology Bulletin*, *41*(8), 1094–1109.
Kushner, S. C., Bagby, R. M., & Harkness, K. L. (2017). Stress generation in adolescence: Contributions from five-factor model (FFM) personality traits and childhood maltreatment. *Personality Disorders: Theory, Research, and Treatment*, *8*(2), 150.
Laceulle, O. M., Ormel, J., Vollebergh, W. A., Van Aken, M. A., & Nederhof, E. (2014). A test of the vulnerability model: Temperament and temperament change as predictors of future mental disorders – the TRAILS study. *Journal of Child Psychology and Psychiatry*, *55*(3), 227–236.
Liu, R. T., & Alloy, L. B. (2010). Stress generation in depression: A systematic review of the empirical literature and recommendations for future study. *Clinical Psychology Review*, *30*(5), 582–593.
Mahaffey, B. L., Watson, D., Clark, L. A., & Kotov, R. (2016). Clinical and personality traits in emotional disorders: Evidence of a common framework. *Journal of Abnormal Psychology*, *125*(6), 758.
McCoach, D. B., & Siegle, D. (2003). The school attitude assessment survey-revised: A new instrument to identify academically able students who underachieve. *Educational and Psychological Measurement*, *63*(3), 414–429.
Mu, W., Luo, J., Nickel, L., & Roberts, B. W. (2016). Generality or specificity? Examining the relation between personality traits and mental health outcomes using a bivariate bi-factor latent change model. *European Journal of Personality*, *30*(5), 467–483.
Mu, W., Luo, J., Rieger, S., Trautwein, U., & Roberts, B. (2019). The relationship between self-esteem and depression when controlling for neuroticism. *Collabra: Psychology*, *5*(1), 11.
Naragon-Gainey, K., Gallagher, M. W., & Brown, T. A. (2013). Stable 'trait' variance of temperament as a predictor of the temporal course of depression and social phobia. *Journal of Abnormal Psychology*, *122*(3), 611.
Nelemans, S. A., Hale, W. W., Branje, S. J., Hawk, S. T., & Meeus, W. H. (2014). Maternal criticism and adolescent depressive and generalized anxiety disorder symptoms: A 6-year longitudinal community study. *Journal of abnormal child psychology*, *42*(5), 755–766.
Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T., & Hartmann, J. A. (2017). Moving from static to dynamic models of the onset of mental disorder: A review. *JAMA Psychiatry*, *74*(5), 528–534.
Nelson, B. D., Perlman, G., Klein, D. N., Kotov, R., & Hajcak, G. (2016). Blunted neural response to rewards as a prospective predictor of the development of depression in adolescent girls. *American Journal of Psychiatry*, *173*(12), 1223–1230.
Nocentini, A., Menesini, E., & Salmivalli, C. (2013). Level and change of bullying behavior during high school: A multilevel growth curve analysis. *Journal of adolescence*, *36*(3), 495–505.
Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: evidence and theory. *Psychological bulletin*, *101*(2), 259.
Nolen-Hoeksema, S. (1991). *Responses to depression questionnaire. Unpublished manuscript*.
Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *132*(1), 1.
Rogosa, D. R. 1995. Myths and methods: 'Myths about longitudinal research' plus supplemental questions. In J. M. Gottman (Ed.), *The analysis of change* (pp. 3–66). Mahwah, NJ: Erlbaum.
Sachs-Ericsson, N., Verona, E., Joiner, T., & Preacher, K. J. (2006). Parental verbal abuse and the mediating role of self-criticism in adult internalizing disorders. *Journal of Affective Disorders*, *93*(1–3), 71–78.
Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147.
Shanahan, L., Copeland, W. E., Costello, E. J., & Angold, A. (2011). Child-, adolescent-and young adult-onset depressions: Differential risk factors in development? *Psychological Medicine*, *41*(11), 2265–2274.
Slavich, G. M., & Irwin, M. R. (2014). From stress to inflammation and major depressive disorder: A social signal transduction theory of depression. *Psychological Bulletin*, *140*(3), 774.
Starr, L. R., & Davila, J. (2008). Differentiating interpersonal correlates of depressive symptoms and social anxiety in adolescence: Implications for models of comorbidity. *Journal of Clinical Child & Adolescent Psychology*, *37*(2), 337–349.
Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology*, *106*(2), 325.
Stice, E., Ragan, J., & Randall, P. (2004). Prospective relations between social support and depression: Differential direction of effects for parent and peer support?. *Journal of Abnormal Psychology*, *113*(1), 155.
Swearer, S. M., Song, S. Y., Cary, P. T., Eagle, J. W., & Mickelson, W. T. (2001). Psychosocial correlates in bullying and victimization: The relationship between depression, anxiety, and bully/victim status. *Journal of Emotional Abuse*, *2*(2–3), 95–121.
Van Voorhees, B. W., Paunesku, D., Kuwabara, S. A., Basu, A., Gollan, J., Hankin, B. L., … Reinecke, M. (2008). Protective and vulnerability factors predicting new-onset depressive episode in a representative of US adolescents. *Journal of Adolescent Health*, *42*(6), 605–616.
Wang, J., Sareen, J., Patten, S., Bolton, J., Schmitz, N., & Birney, A. (2014). A prediction algorithm for first onset of major depression in the general population: Development and validation. *Journal of Epidemiology & Community Health*, *68*(5), 418–424.
Watson, D., O'Hara, M. W., Naragon-Gainey, K., Koffel, E., Chmielewski, M., Kotov, R., … Ruggero, C. J. (2012). Development and validation of new anxiety and bipolar symptom scales for an expanded version of the IDAS (the IDAS-II). *Assessment*, *19*(4), 399–420.
Wilson, S., Vaidyanathan, U., Miller, M. B., McGue, M., & Iacono, W. G. (2014). Premorbid risk factors for major depressive disorder: Are they associated with early onset and recurrent course?. *Development and Psychopathology*, *26*, 1477.
Yaroslavsky, I., Pettit, J. W., Lewinsohn, P. M., Seeley, J. R., & Roberts, R. E. (2013). Heterogeneous trajectories of depressive symptoms: Adolescent predictors and adult outcomes. *Journal of affective disorders*, *148*(2-3), 391–399.
Young, M. A., Fogg, L. F., Scheftner, W., Fawcett, J., Akiskal, H., & Maser, J. (1996). Stable trait components of hopelessness: Baseline and sensitivity to depression. *Journal of Abnormal Psychology*, *105*(2), 155.
Zimet, G. D., Dahlem, N. W., Zimet, S. G., & Farley, G. K. (1988). The multidimensional scale of perceived social support. *Journal of Personality Assessment*, *52*(1), 30–41.

https://doi.org/10.1017/S0033291721004190 Published online by Cambridge University Press