Zainal *et al. BMC Psychiatry* (2026) 26:273
https://doi.org/10.1186/s12888-026-07902-8

BMC Psychiatry

RESEARCH Open Access

Who's at risk for emergent depression years later? Predictive modeling in a nine-year longitudinal cohort

Nur Hani Zainal[1*], Amy T. Peters[2,3], Nicholas C. Jacobson[4] and Kean J. Hsu[1*]

Abstract

Background Major depressive disorder (MDD) is prevalent in adulthood, but there remains a dearth of studies identifying predictors of emergent MDD in adulthood with high-dimensional biopsychosocial predictor sets. Our study thus examined how explainable artificial intelligence (XAI) models might accurately detect predictors of emergent MDD nine years later.

Methods Community adults who did not meet diagnostic criteria for MDD at Wave 1 (W1; 2004–2006) participated in the current study ($N = 931$). Forty-six W1 validated composite variables, including inflammation, childhood maltreatment, coping, emotion regulation, personality, and social support, were used to predict emergent MDD at Wave 2 (W2; 2013–2014). Six machine-learning models, each with four varying configurations of predictor set length and missingness handling strategies, were tested using five-fold nested cross-validation to determine which model had the best multivariable predictive performance. Shapley additive explanations (SHAP) analysis informed the sign and strength of each multivariable predictor.

Results Elastic net regression achieved the best classification accuracy (AUC = 0.724; 95% confidence intervals = 0.657–0.792), with moderate sensitivity and a high negative predictive value in predicting W2 emergent MDD, observed in 6.23% of the sample. Moderate-to-good calibration values were also observed, highlighting acceptable alignment between predicted probabilities and observed prevalences. Key psychosocial correlates of higher W2 emergent MDD risk included greater perceived stress, early life minimization and stress, as well as family and spousal strain. Other key correlates included fewer problem-focused coping strategies, lower self-acceptance, sense of control, and self-directedness, as well as greater tendencies for behavioral disengagement. Demographic correlates included younger age and racial minority identity. Comorbid mental health symptoms, especially higher W1 generalized anxiety disorder, panic disorder, and substance use disorder symptom severity, were also clinical correlates of greater W2 emergent MDD risk.

Conclusions XAI may inform clinically actionable distal risk modeling for emergent MDD using easily measurable, theory-driven variables. If externally validated, scalable multivariable predictive models could be integrated into

*Correspondence:
Nur Hani Zainal
hanizainal@nus.edu.sg
Kean J. Hsu
kean.hsu@nus.edu.sg

Full list of author information is available at the end of the article

healthcare systems to inform prevention strategies. These predictive models might inform tailored treatment strategies, including enhancing approach-focused coping, emotion regulation, and appraisals, as well as objective indicators of social support and stress.

Clinical trial number Not applicable.

Keywords Biopsychosocial, Coping, Emergent depression, Explainable artificial intelligence, Inflammation, Longitudinal, Machine learning, Predictive modeling, Prognosis, Risk factor

Introduction

Major depressive disorder (MDD) is a common psychiatric condition marked by anhedonia (deficits in pleasure or motivation), sleep difficulties, and related depressed mood symptoms across at least two weeks [1]. Based on meta-analyses of population-level epidemiological studies, estimated 12-month prevalence rates have ranged from 8% to 13% globally [2, 3], with recent upward trends [4]. The implications of unidentified and untreated MDD include executive dysfunction, role impairments at school or work, interpersonal problems, and physical health ailments, all of which compromise quality of life [5–7]. As recurrent MDD could be debilitating, identifying distal risk factors [i.e., long-term predictors in the order of years; 8] of *emergent* MDD is essential to plausibly inform universal, selective, or indicated prevention programs.

Myriad biological markers and psychosocial indicators might precede long-term emergent MDD. Proinflammatory activity, characterized by high levels of chronic, low-grade inflammation markers such as interleukin-6, could sustain sedentary lifestyles and sickness-related behaviors that predispose individuals to long-term emergent MDD [9–11]. Relatedly, obesity indicators, such as high body mass index (BMI), might fuel vicious cycles of fatigue, sleep, and appetite disturbances that heighten the risk for emergent MDD over long durations [12]. Key psychological risk factors of emergent MDD are manifold and could include prior child abuse, emotion dysregulation, social support deficits, perceived stress, and mental health comorbidities. Higher child abuse and neglect could trigger and maintain avoidant behaviors by modeling self-sabotaging patterns from caregivers and eroding individuals' sense of control and resilience factors across time [13, 14]. Emotion dysregulation might predict long-term emergent MDD via suboptimal coping tactics, such as deficits in goal-directed behaviors, positive reappraisal, and purpose in life [15]. Interpersonally, perceiving more support from family, friends, spouses, or partners equips individuals with resilience factors and resources to cope with stressors, protecting against MDD in the long run [16]. Together, frameworks such as learned helplessness-hopelessness [17], social support stress-buffering [18], transactional stress-coping [19], and triple vulnerability models [20] suggest that these biopsychosocial variables could be distal risk factors of long-term emergent MDD.

The past five decades have witnessed a veritable growth in studies examining these biopsychosocial variables as correlates or risk factors of emergent and recurrent MDD, examined in isolation, without exploring them holistically within a single high-dimensional predictor set. First, across 27 prospective studies, higher levels of proinflammatory acute-phase proteins and IL-6 were associated with increased future depressive symptoms [21]. Relatedly, another meta-analysis of 19 studies showed that adults with higher (vs. lower) BMI displayed an 18% increased likelihood of depression later on [22]. Second, more frequent child abuse and neglect experiences raised the probability of adulthood MDD by 1.4 to 7.1 times [cf. meta-analysis by 23]. Third, meta-analytic data indicated that low positive reappraisal, self-esteem, and fewer optimal coping approaches predicted more future depressive symptoms in primarily young adults [24, 25]. Moreover, reduced perceived control was correlated with higher depressive symptoms during the COVID-19 pandemic [26]. Another meta-analysis showed that comorbid anxiety disorders and symptoms predicted more future depressive disorders in diverse populations [27]. Fourth, small yet practically meaningful effect sizes that indicated inverse longitudinal relations between social support and depressive symptoms have been consistently observed in both clinical and community samples [28, 29]. Collectively, across studies, the above biopsychosocial variables could, in isolation, precede the emergence of MDD.

However, examining biopsychosocial variables in isolation as distal risk factors of emergent MDD might oversimplify reality and hinder the development of prognostic calculators [e.g., 30] that could inform prevention and treatment targets. A more optimal predictive method would be to test high-dimensional predictor sets that comprise myriad biopsychosocial and theory-driven variables mentioned earlier. However, the mainstay approach to identifying distal risk factors, ordinary least squares (OLS) regression, hinders progress on this topic because it struggles to handle multicollinearity in high-dimensional datasets [31]. Machine learning (ML), a subset of precision medicine methods, addresses multicollinearity [32] while it enables the detection of potential

nonlinearities and higher-order interactions [33]. More accurate estimates could also be achieved through ML-based approaches, such as nested cross-validation (NCV), which trains pattern detection on specific data subsets (or folds) while testing if those patterns generalize to independent subsets [i.e., unseen data; 34].

Ample studies have examined the use of multivariable ML models in predicting depression outcomes, albeit without focusing on predicting emergent MDD per se. Biological models testing genetic and neuroimaging variables, often with small sample sizes, have observed that these high-dimensional predictor sets yield modest to good accuracy rates, ranging from 68.0% to 85.0%, in classifying depression status [cf. systematic review by [35]. However, genetic and neuroimaging data are costly and infrequently assessed in routine care settings, underscoring the need to test the prognostic value of easily measurable, less expensive endocrine and immune markers, such as BMI and proinflammatory activity. Psychosocial variables based on clinical and demographic data, relative to biological variables, also showed greater promise, offering acceptable predictive power (i.e., accuracy thresholds ≥ 70%) for detecting postpartum depression [36]. These patterns were similarly observed in treatment: predictor sets comprising clinical variables yielded external validation accuracies ranging from 57.0% to 74.0% when classifying depression treatment remission, resistance, and response [37]. Together, it is plausible that a multivariable model predicting distal, biopsychosocial risk factors of emergent MDD would have acceptable predictive accuracy.

Simultaneously, distal risk factors could retain predictive power across long durations when they capture potential mechanisms that accrue, recur, or impact long-term behavioral processes. Biological pathways, such as higher BMI and stronger proinflammatory activity, suggest the buildup of allostatic load stemming from chronic stress and lifestyle patterns that raise MDD risk over time [38]. Psychosocial mechanisms, including social support and stress appraisal, vary within individuals and include trait-level components that indicate reliable variations in coping and interpersonal tendencies over time [39–41]. Personality-focused aspects, such as perceived control and self-esteem, showed strong, stable longitudinal patterns, shaping how people choose, perceive, and change their circumstances and influencing their exposure to stressors over time [26, 42, 43]. Therefore, a nine-year duration is, in theory, suitable for identifying accruing biological pathways as well as behavioral and psychological processes central to midlife. Such duration might provide practical predictive value for early detection and prevention. Although this duration aligns with the timeframe offered by the MIDUS project, prior longitudinal research with timeframes over nine years also supported the utility of these biopsychosocial variables for providing essential distal risk prediction within this timeframe [e.g., 11, 44].

Given the theory, empirical research, and logic outlined, the present study examined the predictive accuracy of a high-dimensional predictor set comprising 46 validated composite biopsychosocial variables in predicting the emergence of MDD nine years later. The study's aims were twofold. First, as a prerequisite for discerning the strength and direction of multivariable predictors of nine-year emergent MDD, we hypothesized that our predictive model would achieve good classification accuracy. Second, we expected that the pattern of linear and possibly nonlinear relations and interactions in the multivariable predictors would be consistent with the aforementioned theories.

In summary, we examined three regularized multivariable logistic regression ML models. These models (least absolute shrinkage and selection operator [LASSO], ridge, and elastic net) assume linear main effects and handle collinearity by shrinking coefficients to zero to varying degrees [45]. We also assessed the performance of classification and regression trees (CART), which divide observations into decision trees while accommodating interactions and nonlinearities, but may lead to high variance and non-generalizable patterns (overfitting). Random forest (RF), an ensemble of decorrelated CART algorithms with stopping rules that reduce variance and improve stability on the test set, was also evaluated [46]. Lastly, the support vector machine (SVM) with a radial kernel, which typically excels at distinguishing between outcome classes using a high-dimensional predictor set [47], was assessed. Regularized linear models typically offer greater interpretability and better calibration (i.e., the likelihood that predicted probabilities match actual proportions), whereas CART, RF, and SVM offer greater flexibility at the cost of intuitive appeal [48]. Thus, our approach comprehensively compares interpretable regularized models with flexible nonparametric methods when predicting low prevalence of emergent MDD in a community sample.

Method

Study design

Community-based adults ($N = 931$) offered informed consent to participate in the Midlife Development in the United States (MIDUS) study [49, 50]. The MIDUS project received ethics approval from four participating universities. Given the nature of this secondary data analysis, no additional ethics approval was necessary. The present dataset comprised participants who provided relevant data at Wave 1 [W1; 2004–2006; 50] and Wave 2 [W2; 2013–2014; 49], addressing the focal research aims. We limited our sample to individuals who did not have

past-year MDD at W1, and we defined past-year MDD at W2 as the outcome; lifetime episodes or history between W1 and W2 were not measured.

Participant attributes

Table S1 in the online supplemental materials (OSM) summarizes the descriptive statistics of the original 65 predictor variables, prior to deriving the 46 validated composite predictors, and one outcome variable (W2 emergent MDD). Mean age was 55.93 years (*SD* = 11.91). Gender distribution comprised 52.20% women (486/931) and 47.80% men (445/931). Educational levels included formal college, university, or postgraduate education in 45.30% (422/931). Most participants (93.80%) identified as White (873/931), while the remaining participants self-described themselves as African American, Asian, Multiracial, Native American, or as belonging to other racial groups. The observed prevalence rate of emergent MDD at W2 was 6.23% (58/931).

Procedures

At W1, participants underwent multifaceted data collection, including biological data collection [51] and psychosocial self-report assessments [52]. Participants attended a two-day overnight protocol for the biological data collection at the MIDUS General Clinical Research Center [GCRC; 53]. They fasted in the evening before blood was drawn by a trained phlebotomist the following morning to examine proinflammatory activity and related biomarkers using standardized procedures. Participants also filled out a series of self-reports assessing various theory-informed psychosocial predictors. Simultaneously, trained research personnel clinically interviewed participants to assess the presence of MDD and comorbid psychiatric disorders and symptoms. The following sections elaborate on the measures embedded throughout these procedures. Table S2 presents descriptive statistics for the 46 psychometrically validated composite variables, based on prior literature, coupled with their respective internal consistency values, where applicable.

Measures

W1 and W2 MDD

The Composite International Diagnostic Interview-Short Form (CIDI-SF) was administered to assess past-year MDD symptoms linked to anhedonia or depressed mood for at least two weeks, encountered nearly every day or daily [54]. The MDD symptoms assessed included appetite changes, concentration problems, easily fatigued, feelings of worthlessness, sleep disturbances, and suicidal ideation. This assessment was aligned with the Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised [DSM-III-R; [55]. Previous work has shown that the CIDI-SF MDD scale had high sensitivity, strong specificity, and good construct validity [44, 56]. Emergent MDD was conceptualized as the presence of past-year CIDI-SF-based MDD criteria at W2 among participants who did not have past-year MDD at W1. Major depressive episodes (MDEs) that occurred and remitted fully between W1 and W2 time points were not measured; thus, emergent MDD cases at W2 could indicate either initial-onset or repeated MDEs.

W1 Biological predictors

Participants' blood specimens from the W1 biomarker data collection protocol were assayed to assess the levels of proinflammatory marker activity levels, including CRP, IL-6, IL-6 receptor (IL-6R), fibrinogen, E-selectin, intracellular adhesion molecule-1 (ICAM-1), Meso scale diagnostics (MSD) IL-6, tumor necrosis factor-alpha (TNF-α), interleukin-8 (IL-8), and interleukin-10 [IL-10; 57]. To clarify, MSD IL-6 is defined as IL-6 measured on the MSD electrochemiluminescence machine, which varies from the IL-6 value derived separately, as it is the same cytokine but quantified with a unique assay and detection limits [58]. The research team also used calibrated measurement tools to assess anthropometric data, including height and weight. BMI was computed using the accurate formula (weight/height^2 [kg/m^2]), which reflected an indicator of obesity risk [59].

W1 Psychosocial predictors

Participants also completed a series of self-report assessments to evaluate a range of theory-based psychosocial predictors of emergent MDD. These constructs included perceived stress in the past month [Perceived Stress Scale; 60] and retrospective childhood maltreatment [e.g., emotional and physical abuse and neglect; Childhood Trauma Questionnaire; CTQ; 61]. Trait-level coping strategies, such as behavioral disengagement, denial, using food to cope [52], and emotion regulation [e.g., reappraisal; 62], were also measured. Personality and self-concept measures, for example, included autonomy, self-acceptance, and self-esteem [63]. Cognitive functioning tests that assessed diverse executive functioning (EF) and episodic memory (EM) dimensions were also administered [64]. Further, various dimensions of social support were assessed, including perceived social support and strain from family members, friends, and spouses or partners [65]. Mental health comorbidities, including generalized anxiety disorder (GAD) and panic disorder (PD), and symptom severity, were also assessed using the CIDI-SF [54], given research on prospective comorbidities [66, 67]. Based on prior research [68], we also examined alcohol use disorder [AUD; 69] and substance use disorder (SUD) symptom severity [70] as potential predictors of W2 emergent MDD.

W1 Data sources

To clarify the assessment time points, the W1 data were derived from unique MIDUS 2 substudies. Primary psychosocial measures and CIDI-SF data were gathered during the MIDUS survey protocol from 2004 to 2006. A subsample from this participant pool underwent the MIDUS 2 Biomarker Project procedures described above. The mean duration between the survey protocol and Biomarker Project procedures has been reported to be about two years [51]. BMI and markers of proinflammatory activity were drawn from the GCRC visit. The remaining predictors were derived from the MIDUS 2 survey protocol or the telephone-delivered Brief Test of Adult Cognition by Telephone (BTACT) protocol, which assessed cognitive functioning domains at the same W1 time point [64].

Data analyses

A series of multivariable ML analyses were conducted to predict emergent W2 MDD diagnosis. To predict emergent W2 MDD diagnosis, participants with W1 MDD diagnosis were removed, resulting in a dataset with 931 participants. All analyses were conducted using the *R* software [71]. Initial steps included RF imputation of missing data using the *missRanger* package [72], normalizing continuous variables, and one-hot encoding nominal variables [48] in the training folds of the nested cross-validation (NCV) multivariable models to prevent data leakage. Additionally, since our objective was risk modeling rather than causal effect estimation, the predictor set was not limited by assumed mediational pathways. We measured all predictors at W1 before measuring the W2 outcome.

Assumption checks linked to the initial steps outlined above were conducted. Regarding missingness, we assessed the missingness pattern using Little's MCAR test (χ^2([degrees of freedom] [*df*] = 200) = 268.01, $p < .001$); RF imputation was constrained to the predictors while consistently excluding the outcome. Continuous predictors were normalized; the data distribution, both before and after scaling, did not reveal any outliers [48]. Categorical predictors were one-hot encoded. Sensitivity analyses were conducted to compare the main results of RF imputation with those from the complete-case analysis, as well as to examine results using unique predictor sets focused on distal risk factors or mechanisms.

To address potential attrition and survivorship bias resulting from the 9-year interval, we estimated inverse probability weights (IPWs) using logistic regression of W1 variables to predict W2 completion status [73]. IPW weights every participant by the inverse of their expected likelihood of the W2 outcome, thereby upweighting minority profiles and producing estimates that closely approximate the entire W1 sample rather than the completer subsample. These IPW values were applied to all ML models. Table S3 details the completer vs. non-completer analyses that generated the IPWs.

Our multivariable ML approaches followed the Transparent Reporting of a multivariable prediction model of Individual Prognosis or Diagnosis (TRIPOD) recommendations [74, 75]. Six ML algorithms were examined [least absolute shrinkage and selection operator [LASSO], ridge, elastic net, classification and regression trees [CART], random forest [RF], and support vector machine [SVM]; 34]. Table S4 provides more details on ML models. Linear models (LASSO, ridge, elastic net with the learning rate [α] set at 0.5) tested how a sparse predictor set or uniform shrinkage patterns produced improved generalization. The penalty terms [λs] were tuned in the inner training folds. Regarding tree-based models, we assessed how a single-pruned CART algorithm performed compared to an RF ensemble of CARTs, which typically reduces model variance. We also evaluated the performance of an SVM algorithm with a radial basis function kernel. A grid search was conducted to select the best hyperparameters in the inner loop. Given the rarity of the W2 emergent MDD outcome, two sample-balancing strategies were compared: the synthetic minority oversampling technique (SMOTE) and random subsampling. Model performance tests of all six models were repeated across four different configurations of distinct predictor sets: (i) all 65 W1 variables (Table S1); (ii) a bivariate-correlations screened subset of 22 W1 variables that significantly correlated with W2 emergent MDD; (iii) 10 principal components analysis (PCA) factors; (iv) 46 composite W1 theory-informed variables validated based on the psychometric literature (Table S2). Together, these configurations were selected to balance variance, bias, and interpretability, and draw robust inferences about predictor set selection, regularization, and sampling.

To reduce overfitting and optimize generalizability, we used a five-fold NCV approach with the *nestedcv* package for each examined algorithm [34]. Hyperparameter tuning, model selection, and training were conducted in the inner CV loops, whereas model testing was performed in the outer CV loops to avoid data leakage. Multicollinearity was addressed with regularized classifiers [76]. The most optimal model, elastic net, applied binomial logistic regression and tuned the regularization parameters to 0.5 to set an optimal balance between LASSO and ridge penalties. Random subsampling was used in each training fold to address class imbalance (i.e., 6.23% of cases with vs. without emergent MDD). After fitting the model to the training data, the model generated predictive performance metrics with the held-out test data in the outer CV folds, aggregated across the five outer test folds to attain stable estimates.

Model discrimination was assessed using the area under the receiver operating characteristic curve (AUC), along with its 95% confidence intervals (CI), computed with the *pROC* package [77]. The AUC indicated that the model ranked a randomly chosen participant with vs. without W2 emergent MDD higher. An AUC value of 0.5 indicated no difference from chance performance, 1.0 denoted perfect classification, and AUC values above 0.7 were considered acceptable in predicting clinical outcomes [78]. The confusion matrix was extracted from all aggregated predictions, weighted by IPW values. Specifically, the confusion matrix data were used to calculate sensitivity (also called recall), specificity, positive predictive value (PPV; also called precision), and negative predictive value [NPV; 79], and the *binom* package was used to calculate their 95% CIs [80]. Sensitivity (recall) indicated the percentage of true W2 emergent MDD cases accurately detected. Specificity was defined as the percentage of true W2 emergent MDD cases accurately ruled out. PPV denoted the likelihood that a case predicted to develop W2 emergent MDD actually did, whereas NPV reflected the likelihood that a case predicted not to develop W2 emergent MDD actually did not. PPV and NPV were sensitive to prevalence rates, so PPV values were expected to be low herein, given the low prevalence of W2 emergent MDD (6.23% [58/931]).

Model calibration analyses, which assessed the consistency between predicted probabilities and observed prevalences, were conducted to predict the emergence of W2 MDD. The following point estimates and their bootstrapped 95% CIs were computed for three calibration metrics using the *boot* package [81]: *Brier score* (global prediction accuracy derived by computing the mean squared difference between predicted probabilities and observed prevalences); *expected calibration error* [ECE; mean absolute score difference between predicted probabilities and observed prevalences across bins of model predictions; [82]; *integrated calibration index* [ICI; mean absolute score difference between predicted probabilities and observed prevalences from a calibration curve with smoothing features; [83]. Lower Brier scores, ICI, and ECE values indicated better model calibration.

To improve model interpretability, we implemented the Shapley Additive explanations [SHAP; 84] method for the elastic net model to identify the essential predictors of emergent MDD with the *fastshap* package [85]. SHAP coefficients indicated the effect of each multivariable predictor on the model output at the participant level [86]. As SHAP analysis estimates the marginal effect of a multivariable predictor across all possible permutations of predictor subsets, it provides a holistic approach to examining the relative importance of predictors, consistent with game-theoretic principles [84].

To implement the SHAP approach, SHAP values were calculated using an NCV out-of-sample approach. The elastic net was tuned and trained in the inner loop, and SHAP values were computed on the corresponding held-out outer loop test folds using a model-agnostic Monte Carlo method (number of simulations = 500) through the *fastshap* package [85]. The *glmnet* prediction wrapper was applied. Individual-level SHAP values were aggregated across outer folds, and predictors were organized according to the weighted-average absolute SHAP values computed using the IPW values mentioned earlier.

The SHAP bee swarm plot was created to visualize both the strength and sign of each predictor's contribution to the model output, with horizontal jitter indicating between-participant variability and color schemes denoting predictor strength [blue for lower scores and red for higher scores; [87]. The vertical position signaled the predictor's relative importance, but the horizontal density plots showed the variability in SHAP coefficients that relay the overall sign between the predictor and outcome, including potential nonlinearities and interactions. Importantly, SHAP values indicated predictive importance without permitting causal inferences.

Results

Model performance metrics in multivariable models predicting W2 emergent MDD

Discrimination

The configuration with the best performance was the multivariable ML models that used 46 validated W1 composite predictors while using RF imputation and IPW to handle missing data and random subsampling to manage class imbalance (OSM Tables S5 to S10 and Table 1). Table 1 summarizes the multivariable model performance metrics for the 46 validated W1 composite predictors of emergent W2 MDD, including AUC, sensitivity, specificity, PPV, NPV, and their 95% CIs. Table 2 displays the weighted confusion matrices of the six five-fold NCV models in that configuration. Elastic net classification produced the best discrimination performance (AUC = 0.724, 95% CI = 0.657–0.792), followed by the ridge classification model (AUC = 0.723, 95% CI = 0.655–0.790). Furthermore, the elastic net showed moderate sensitivity (0.683, 95% CI = 0.552–0.790) and a high NPV (0.970, 95% CI = 0.953–0.981). Together, none of the ML models that accommodated nonlinearities and higher-order interactions (CART, RF, and SVM) performed as well as the regularized classification models.

Calibration

Table 3 summarizes the model calibration metrics. The elastic net classification model consistently showed low values for Brier score (0.199, 95% CI = 0.189–0.207), ICI (0.366, 95% CI = 0.347–0.383), and ECE (0.365, 95%

Table 1 Multivariable ML performance metrics for the model with 46 W1 validated composite predictors of emergent W2 MDD using random subsampling

Model	Metric	Estimate	LCI	UCI
LASSO	AUC	0.691	0.623	0.759
	Sensitivity (SN)	0.488	0.362	0.616
	Specificity (SP)	0.806	0.777	0.833
	Positive Predictive Value (PPV)	0.152	0.107	0.212
	Negative Predictive Value (NPV)	0.957	0.939	0.970
Ridge	AUC	0.723	0.655	0.790
	Sensitivity (SN)	0.627	0.495	0.741
	Specificity (SP)	0.744	0.712	0.773
	Positive Predictive Value (PPV)	0.148	0.109	0.199
	Negative Predictive Value (NPV)	0.966	0.948	0.977
ENR	AUC	0.724	0.657	0.792
	Sensitivity (SN)	0.683	0.552	0.790
	Specificity (SP)	0.739	0.707	0.768
	Positive Predictive Value (PPV)	0.157	0.116	0.208
	Negative Predictive Value (NPV)	0.970	0.953	0.981
CART	AUC	0.646	0.578	0.714
	Sensitivity (SN)	0.646	0.515	0.759
	Specificity (SP)	0.534	0.499	0.568
	Positive Predictive Value (PPV)	0.090	0.066	0.122
	Negative Predictive Value (NPV)	0.955	0.931	0.971
RF	AUC	0.662	0.593	0.731
	Sensitivity (SN)	0.641	0.510	0.754
	Specificity (SP)	0.665	0.632	0.698
	Positive Predictive Value (PPV)	0.120	0.088	0.162
	Negative Predictive Value (NPV)	0.963	0.944	0.976
SVM	AUC	0.624	0.551	0.696
	Sensitivity (SN)	0.584	0.454	0.704
	Specificity (SP)	0.574	0.539	0.608
	Positive Predictive Value (PPV)	0.089	0.064	0.122
	Negative Predictive Value (NPV)	0.951	0.928	0.967

Note. ML, machine learning; W1, wave 1 (2004–2006); W2, wave 2 (2013–2014); MDD, major depressive disorder; LCI, lower bound of the 95% confidence intervals (CIs); UCI, upper bound of the 95% CIs; AUC, area under the receiver operating characteristic curve; SN, sensitivity; SP, specificity; PPV, positive predictive value; NPV, negative predictive value; LASSO, least absolute shrinkage and selection operator; ENR, elastic net regularization; CART, classification and regression trees; RF, random forest; SVM, support vector machine. Bold values denote the algorithm with the best specific multivariable ML model performance metric. Inverse probability weights (IPWs) are applied in each multivariable ML model

Table 2 Weighted confusion matrix of the five-fold nested CV model with 46 W1 validated composite predictors of emergent W2 MDD

Model	True Positives (TP)	False Positives (FP)	True Negatives (TN)	False Negatives (FN)
LASSO	27.242	151.814	632.538	28.577
Ridge	34.972	200.999	583.352	20.847
ENR	38.097	204.987	579.365	17.722
CART	36.080	365.869	418.483	19.739
RF	35.780	262.465	521.887	20.039
SVM	32.623	334.551	449.801	23.196

Note. CV, cross-validation; W1, wave 1 (2004–2006); W2, wave 2 (2013–2014); MDD, major depressive disorder; LASSO, least absolute shrinkage and selection operator; ENR, elastic net regularization; CART, classification and regression trees; RF, random forest; SVM, support vector machine. Inverse probability weights (IPWs) are applied in each multivariable ML model

Table 3 Calibration analysis for the model with 46 W1 validated composite predictors of emergent W2 MDD using random subsampling

Model	Metric	Estimate	LCI	UCI
LASSO	ICI	0.376	0.357	0.393
	ECE	0.374	0.355	0.392
	Brier Score	0.210	0.201	0.219
Ridge	ICI	0.364	0.346	0.382
	ECE	0.364	0.346	0.382
	Brier Score	0.198	0.189	0.206
ENR	ICI	0.366	0.347	0.383
	ECE	0.365	0.346	0.382
	Brier Score	0.199	0.189	0.207
CART	ICI	0.385	0.362	0.404
	ECE	0.384	0.362	0.404
	Brier Score	0.246	0.231	0.261
RF	ICI	0.392	0.371	0.410
	ECE	0.391	0.371	0.410
	Brier Score	0.224	0.214	0.232
SVM	ICI	0.414	0.393	0.433
	ECE	0.415	0.395	0.434
	Brier Score	0.245	0.235	0.256

Note. W1, wave 1 (2004–2006); W2, wave 2 (2013–2014); MDD, major depressive disorder; LCI, lower bound of the 95% confidence intervals (CIs); UCI, upper bound of the 95% CIs; LASSO, least absolute shrinkage and selection operator; ICI, integrated calibration index; ECE, expected calibration error; ENR, elastic net regularization; CART, classification and regression trees; RF, random forest; SVM, support vector machine. Bold values denote the algorithm with the best specific multivariable ML model performance metric. Inverse probability weights (IPWs) are applied in each multivariable ML model

CI = 0.346–0.382). This pattern of calibration metrics indicated moderate-to-good correspondence between predicted probabilities and observed prevalences. Finally, the calibration performance of the elastic net was comparable to that of the ridge algorithm and superior to CART, RF, and SVM (Fig. 1).

Magnitude and direction of each multivariable W1 predictor of W2 emergent MDD

Table 4 summarizes the SHAP scores, indicating the relative magnitude of the top 20 W1 multivariable-validated composite predictors of W2 emergent MDD among non-MDD cases at W1. Figure 2 shows both the magnitude and direction of the contribution of W1 predictors in the SHAP, after accounting for potential nonlinearities and higher-order interactions; the numbers reflect their estimated relative importance. Regarding psychosocial factors, higher perceived stress (#1), early-life minimization and neglect (#3), lower family support (#5), higher family strain (#11), and lower spousal support (#16) were associated with a stronger likelihood of W2 emergent MDD. With respect to demographic variables, younger

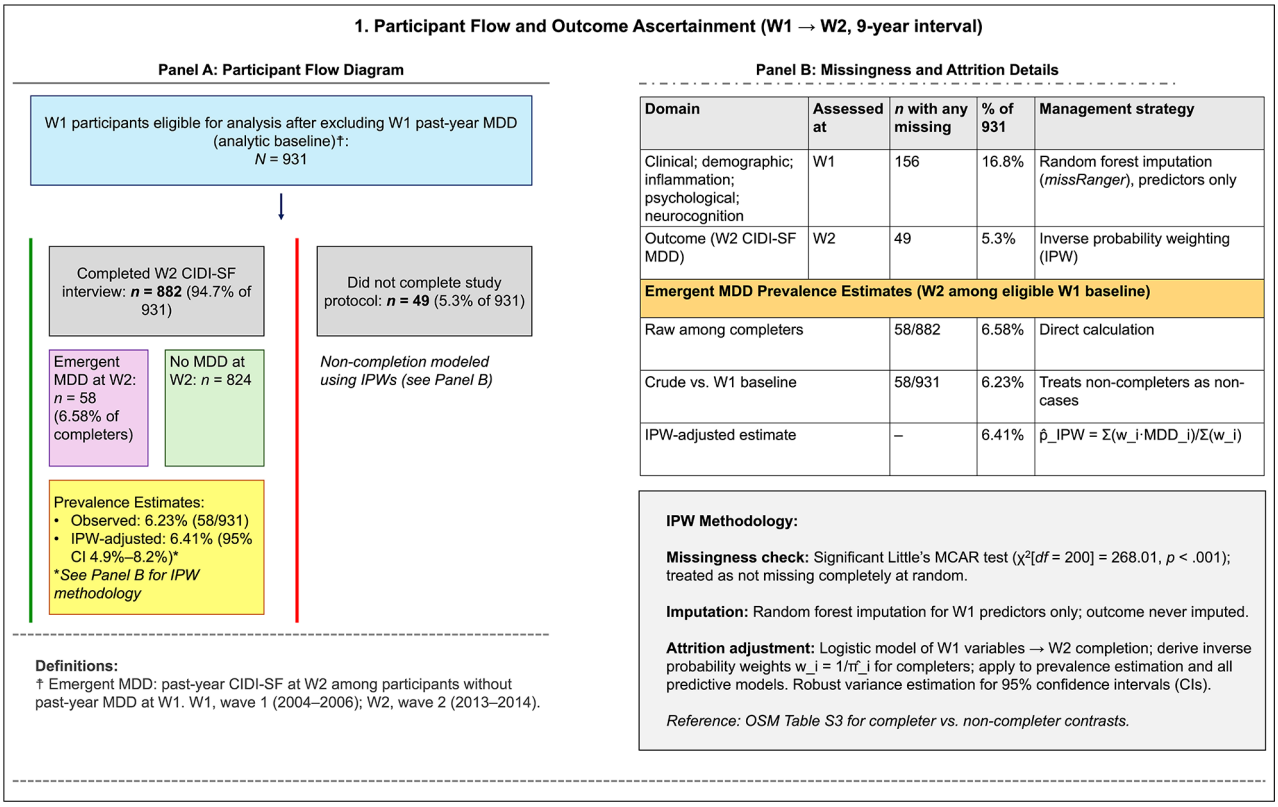

Fig. 1 Overview of study participant flow. W1, wave 1 (2004–2006); W2, wave 2 (2013–2015); MDD, major depressive disorder; CIFI-SF, Composite International Diagnostic Interview-Short Form; IPW, inverse probability weights; MCAR, missing completely at random; OSM, online supplemental materials

Table 4 Mean absolute SHAP values for the top 20 W1 validated composite predictors of emergent W2 MDD diagnosis

Predictor of emergent W2 MDD diagnosis	|SHAP|
1. W1 Perceived stress (+)	0.031
2. W1 Age (years) (–)	0.028
3. W1 CTQ Minimization and neglect (+)	0.022
4. W1 GAD severity (+)	0.022
5. W1 Family support (–)	0.021
6. W1 Problem-focused coping (–)	0.017
7. W1 Self-acceptance (–)	0.017
8. W1 Health locus of control (–)	0.015
9. W1 Self-directedness (–)	0.012
10. W1 Self-esteem (–)	0.010
11. W1 Family strain (+)	0.010
12. W1 Insight into the past (–)	0.010
13. W1 White individuals (–)	0.010
14. W1 Psychological well-being (–)	0.010
15. W1 Panic disorder severity (+)	0.009
16. W1 Episodic memory (Immediate) (+)	0.009
17. W1 Spousal support (–)	0.009
18. W1 SUD severity (+)	0.009
19. W1 Behavioral disengagement (+)	0.009
20. W1 Perceived uncontrollability (+)	0.008

Note. SHAP, Shapley additive explanations; W1, wave 1 (2004–2006); W2, wave 2 (2013–2014); MDD, major depressive disorder; CTQ, childhood trauma questionnaire; GAD, generalized anxiety disorder; SUD, substance use disorder. |SHAP| denotes the mean absolute SHAP value for each predictor

age (#2) and non-White individuals (#12) were correlated with higher W2 emergent MDD risk. These coping strategies were also correlated with higher probability of W2 emergent MDD: lower problem-focused coping (#6), self-acceptance (#7), health locus of control (#8), self-directedness (#9), self-esteem (#10), insight into the past (#13), psychological well-being (#14), as well as higher behavioral disengagement (#19) and perceived uncontrollability (#20). For clinical variables, higher W1 GAD severity (#4), panic disorder severity (#15), and SUD severity (#18) were associated with greater W2 emergent MDD risk. For neurocognitive variables, higher total EM (#16) was correlated with greater likelihood of W2 emergent MDD. Note that, as we excluded W1 MDD cases, the significant W1 GAD, panic disorder, and SUD severity predictors suggested subthreshold symptom severity with restricted ranges.

Sensitivity analyses

Sensitivity analyses examined the pattern of predictive performance (Tables S11 and S12), model calibration strength (Table S13), and predictor-outcome associations (Figure S1) under complete-case analysis for the primary analyses. Model predictive performance under complete-case analysis, both in terms of discrimination and calibration, was not as strong as that obtained from the primary

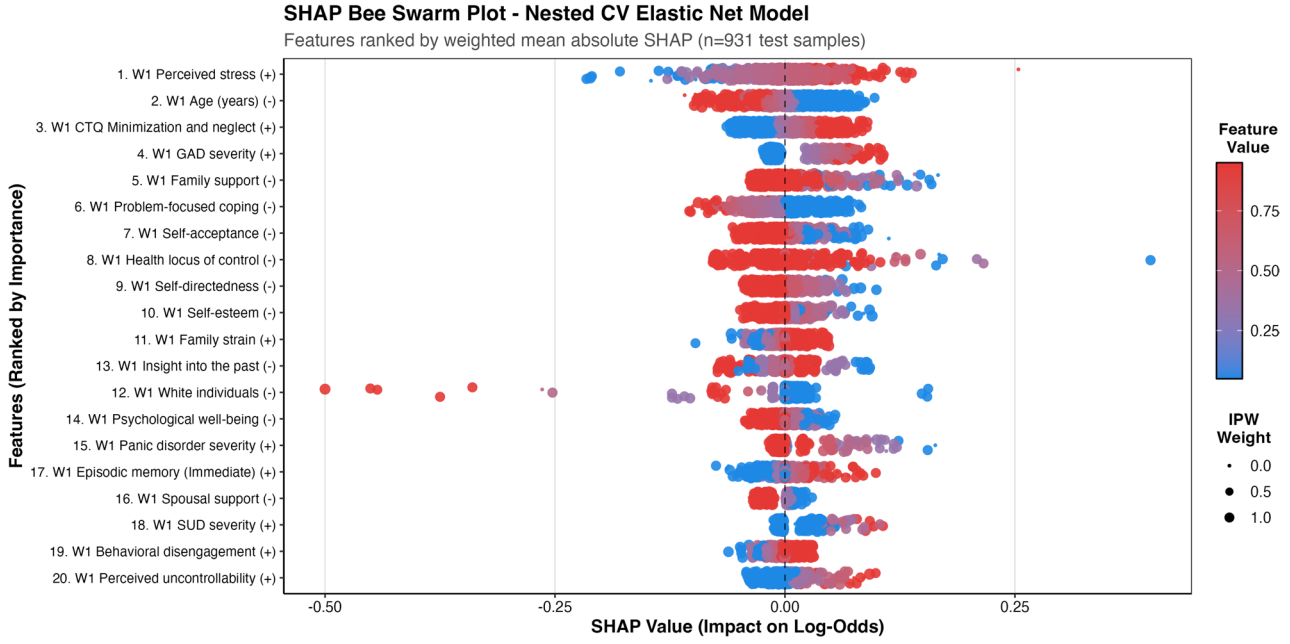

Fig. 2 SHAP bee swarm plot of the elastic net classification model of validated composite W1 variables predicting emergent W2 MDD diagnosis. SHAP, Shapley additive explanations; W1, wave 1 (2004–2006); W2, wave 2 (2013–2014); MDD, major depressive disorder; CV, cross-validation; CTQ, childhood trauma questionnaire; GAD, generalized anxiety disorder; SUD, substance use disorder. The SHAP bee swarm plot from the nested CV elastic net classification illustrates the relative importance of each validated composite W1 variable in predicting the emergent W2 MDD among non-MDD cases at W1. W1 variables are organized by the weighted average absolute SHAP values, such that positive signs (+) indicate predictors linked to a higher likelihood of emergent MDD. Conversely, lower signs (–) denote predictors associated with a lower likelihood of emergent MDD. Each point reflects participant-level data. The horizontal (x-axis) shows the sign and strength of the effect on the log-odds of W2 emergent MDD. Point color reflects the normalized predictor value (red is higher; blue is lower). The point size denotes the inverse probability weighting (IPW) derived from completer vs. non-completer status. Higher SHAP values suggest stronger contribution of the W1 predictor to the model outcome

analyses using RF imputation with IPW. However, many of the observed predictor-outcome correlation patterns remained in the sensitivity analyses.

Another set of sensitivity analyses tested predictive performance with two unique predictor sets. One predictor set included seven distal risk factors: age, gender, race, education level, household income, CTQ minimization and denial, and the CTQ total scale. This distal risk factor set had poor predictive performance metrics across all ML models (Table S14). The other predictor set comprised 39 mechanism-focused W1 variables, excluding the seven distal risk factors. The mechanisms-focused predictor set showed acceptable predictive performance, with elastic net classification again performing best (Table S15).

Discussion

The present study tested the predictive accuracy of ML models for identifying emergent MDD 9 years later, using 46 theory-informed, psychometrically validated composite biopsychosocial predictors. Regularized linear models, particularly elastic net, performed best in terms of discrimination and calibration compared to tree- and kernel-based algorithms. Pending external validation studies [88], the results highlight the viability of using explainable, parsimonious ML algorithms to detect distal risk factors for long-term emergent MDD. The SHAP analyses further identified the direction and magnitude of critical theory-based predictors of emergent MDD. Plausible accounts are provided to encourage greater precision medicine research to predict long-term, emergent MDD.

Why did regularized linear models, particularly the elastic net, outperform their ensemble and nonlinear counterparts in predicting nine-year emergent MDD? This pattern indicated that the validated composite predictors at baseline had an additive effect, rather than a multiplicative impact, on nine-year emergent MDD, given the observed prevalences and sample size of our study. These results might be explained by the effectiveness of regularization methods in handling high-dimensional biopsychosocial data, managing multicollinearity, and distilling the essential predictors of future psychiatric outcomes [89, 90]. Our results also concurred with prior evidence that regularized elastic net excels at predicting depression outcomes better than other complex ML algorithms that consider nonlinear and moderation effects for datasets with limited sample sizes [91]. Findings also reinforced the 'no free lunch' theorem in ML,

which stipulates that any ML algorithm's relative benefits are context-dependent instead of universal [92].

Relatedly, PPV values were consistently low across all models, whereas NPV outcomes were consistently high, which is a pattern typical for community samples with low MDD base rates. Stated differently, PPV is correlated with the prevalence rates of the outcome [93–95]. These patterns highlight greater pragmatic value in ruling out future emergent MDD cases and in building population-level risk-stratification models [96, 97]. Multivariable models with better PPV performance would be needed for prognostication, early identification, and intervention purposes [98]. Longitudinal studies employing dynamically measured, multimodal predictors and recruiting samples with a higher, more balanced prevalence of emergent MDD could enhance all aspects of multivariable predictive performance metrics.

Concordant with discrimination performance, the elastic net model showed moderate-to-good calibration performance among the examined algorithms, indicating an acceptable alignment between predicted probabilities and observed prevalences. Dual strength in both calibration and discrimination is crucial for real-world clinical utility, as both aspects enhance clinical decision-making, interpretability, and risk prediction and stratification [78, 99]. Collectively, future psychiatric studies identifying distal risk factors should continue testing calibration and discrimination to build clinically actionable predictive models.

The SHAP analysis identified an intricate set of multivariable predictors of emergent MDD. Readers should interpret these multivariable patterns as indicating how SHAP computes each baseline variable's marginal effect on the model output. Despite baseline variables being associated or possibly connected via mediational pathways, the ML method treats them as simultaneous risk factors rather than causal chains, thereby reducing concerns about confounding effects in this predictive setting. Higher perceived stress, early adversity characterized by minimization and neglect, and poorer social support dimensions were clear risk factors. These patterns might indicate the stress-sensitization effects of early adversity and its long-term repercussions on interpersonal dynamics [100, 101], plausibly mediated through immune regulatory processes [102], that heighten emergent MDD risk.

Demographically, younger adults and non-White persons were at risk of emergent MDD. These gradients might reflect developmental and structural factors. Perhaps younger adults experience more stressful life transitions across career, finances, marriage, and other essential life domains than their older counterparts [103]. For non-White participants, lived experiences, such as chronic discrimination, fewer accrued coping resources, and less access to culturally adapted prevention programs and treatments [104], might explain these patterns.

Several psychological coping factors were also implicated in predicting emergent MDD. The constellation of variables suggested that less usage of approach-focused coping and emotion regulation strategies, coupled with limiting beliefs about perceived control, were notable risk factors. Avoidance patterns and learned helplessness mindsets, as reflected in higher behavioral disengagement [105], blunted self-construals (e.g., self-acceptance) [25, 42, 106], and weaker insights [107], may also adversely impact reward learning and problem-solving. Together, these factors might jointly increase the probability of emergent MDD nine years later among non-MDD cases at baseline.

Despite their restricted ranges, higher baseline GAD, panic disorder, and SUD symptom severity were correlated with higher probability of emergent MDD among those without past-year MDD at baseline. These clinical variables likely signaled transdiagnostic indicators of either MDD incidence or recurrence via social-cognitive processes, such as interpersonal issues [108] and threat hypervigilance [109]. Collectively, these comorbid symptoms might raise the likelihood of emergent MDD through various stress generation processes.

Intriguingly, EF variables were not predictive of emergent MDD; however, stronger total EM was counterintuitively associated with a higher risk of emergent MDD. Plausibly, stronger EM might be intimately linked to the tendency to recall stressful events in ways that perpetuate non-constructive ruminative brooding. Relatedly, collider bias might account for this pattern [110], as people without MDD at baseline were followed up, potentially creating hidden confounding that could lead to misleading relationships.

Upon conducting a complete case analysis, the predictive performance metrics of the multivariable ML models were compromised. Perhaps deleting participants with incomplete responses resulted in the loss of valuable information and skewed the sampling distribution [111]. Comparatively, our primary analyses imputed plausible values using the RF approach to leverage all available data, and random subsampling addressed class imbalance in the emergent MDD outcome variable. However, many similar inferences would have been drawn, suggesting that the primary outcomes were stable and not due to missing-data handling strategies.

The present study should be interpreted within the context of its limitations. First, because our study measured past-year MDD at each time point, it was not possible to distinguish between recurrent and first-onset episodes over the 9 years. Outcomes should thus be construed as distal risk for past-year MDD status at W2 among non-cases at W1, instead of the incidence of first-onset MDD.

Relatedly, this design shortcoming could have underestimated the actual incidence rates of emergent MDD. Second, as we focused on internal validation with the NCV approach, future studies should conduct external validation to enhance the model's clinical utility [88]. Third, future attempts to develop a multivariable predictive model that excels in both discrimination and calibration should expand the predictor set, for example, by including electronic health records (EHR) data [112]. Fourth, given the low rate of emergent MDD (6.23%), replication efforts should recruit larger and more balanced samples [113]. Fifth, baseline self-reported predictors may shift over time and be liable to recall biases, underscoring the importance of exploring time-varying models. Sixth, no causal conclusions can be drawn from this multivariable prediction model. Seventh, given the well-established inequities in access to and care [114], future research should recruit more diverse samples by race, gender identity, and socioeconomic status. Finally, the exclusion criterion of removing W1 MDD cases, though important for examining emergent cases, did not capture the full range of symptom severity, a caveat for readers as they construe these findings.

However, the study's strengths include a 9-year follow-up, a high-dimensional predictor set comprising 46 theory-informed and psychometrically validated composite biopsychosocial variables, robust tests of ML models, and the NCV method, which minimized overfitting. In addition, the selected variables in the predictor set were accessible and inexpensive to assess. Finally, discovering that simpler linear models, such as elastic net, were better at predicting long-term emergent MDD than more complex models accounting for nonlinear relationships could guide multivariable predictive modeling of psychiatric outcomes with low observed prevalence.

If externally validated in larger, more balanced samples, the clinical implications, including early identification of long-term emergent MDD risk and individualized prevention, merit attention. First, the capacity of regularized linear models, especially the elastic net, to identify risk factors of emergent MDD nine years later with accessible biopsychosocial self-report data offers practicality and portability to community mental health and primary care settings. Second, the high NPV values imply the elastic net model's strength to accurately rule out future risk among low-risk participants, optimizing the allocation of prevention resources. Third, the SHAP analysis could enable personalized prevention by modifying key factors, such as alleviating perceived stress, optimizing BMI, and increasing self-acceptance and social support in adults flagged as high risk for emergent MDD. Fourth, interpretable and transparent algorithms, such as elastic net, might enhance clinicians' trust in prognostic models and ease shared decision-making between patients and providers [115, 116]. Collectively, these clinical implications highlight the potential for scalable, actionable, and cost-effective prognostic instruments to be used to prevent the emergence of MDD over extended time horizons.

Abbreviations

AUC	Area under the receiver operating characteristic curve
AUD	Alcohol use disorder
BMI	Body mass index
BTACT	Brief Test of Adult Cognition by Telephone
CART	Classification and regression trees
CI	Confidence intervals
CIDI-SF	Composite International Diagnostic Interview Short Form
CRP	C-reactive protein
CTQ	Childhood Trauma Questionnaire
df	Degrees of freedom
DSM-III-R	Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised
ECE	Expected calibration error
EF	Executive functioning
EHR	Electronic health records
EM	Episodic memory
GAD	Generalized anxiety disorder
GCRC	General Clinical Research Center
ICAM-1	Intracellular adhesion molecule-1
ICI	Integrated calibration index
IL-6	Interleukin-6
IL-6R	Interleukin-6 receptor
IL-8	Interleukin-8
IL-10	Interleukin-10
IPW	Inverse probability weights
IRB	Institutional review board
LASSO	Least absolute shrinkage and selection operator
MDD	Major depressive disorder
MDE	Major depressive episode
MIDUS	Midlife Development in the United States
ML	Machine learning
MSD	Meso scale diagnostics
NCV	Nested cross-validation
NPV	Negative predictive value
NUS	National University of Singapore
OLS	Ordinary least squares
OSM	Online supplemental materials
PCA	Principal components analysis
PD	Panic disorder
PPV	Positive predictive value
PYP	Presidential Young Professorship
RF	Random forest
SD	Standard deviation
SHAP	Shapley additive explanations
SMOTE	Synthetic minority oversampling technique
SUD	Substance use disorder
SVM	Support vector machine
TNF-α	Tumor necrosis factor-alpha
TRIPOD	Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis
W1	Wave 1
W2	Wave 2
XAI	Explainable artificial intelligence

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-026-07902-8.

Supplementary Material 1

Acknowledgements
The authors thank the MIDUS study participants and the research teams at Harvard University, Georgetown University, the University of California, Los Angeles, and the University of Wisconsin-Madison for their contributions to data collection. The authors also acknowledge support from the National University of Singapore.

Author contributions
NHZ was responsible for conceptualization, design, data analysis and interpretation, original manuscript writing, visualization, methodology, and project supervision. ATP, NCJ, and KJH each contributed to project supervision and critical revision of the manuscript. All authors read and approved the final manuscript.

Funding
Since 1995, the MIDUS study has been supported by the John D. and Catherine T. MacArthur Foundation Research Network and the National Institute on Aging (P01-AG020166; U19-AG051426). This secondary analysis was additionally supported by the National University of Singapore (NUS) Presidential Young Professorship (PYP) Start-Up Grant and the White Space Fund, both awarded to Professor Zainal (NHZ). The funders and original MIDUS investigators had no role in the design, analysis, interpretation, or writing of this manuscript.

Data availability
The datasets analyzed during the current study are publicly available through the Inter-university Consortium for Political and Social Research (ICPSR) repository. MIDUS 2 (2004-2006) data can be accessed at https://doi.org/10.3886/ICPSR04652, and MIDUS 3 (2013-2014) data can be accessed at https://doi.org/10.3886/ICPSR36346. The data analytic scripts used in this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The MIDUS study received ethics approval from institutional review boards (IRBs) at Harvard University, Georgetown University, the University of California, Los Angeles, and the University of Wisconsin-Madison. All participants provided written informed consent in accordance with the IRB requirements at these institutions and the Declaration of Helsinki. The current secondary analysis utilized publicly available, de-identified data and was therefore exempt from additional IRB review.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, Block AS4, Level 2, 9 Arts Link, Singapore 117570, Singapore
[2]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA
[3]Department of Psychiatry, Harvard Medical School, Boston, MA, USA
[4]Center for Technology and Behavioral Health, Dartmouth College, Geisel School of Medicine, Lebanon, NH, USA

Received: 2 July 2025 / Accepted: 9 February 2026
Published online: 24 February 2026

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5-TR. 5th ed. Washington, DC: American Psychiatric Association Publishing; 2022.
2. Abdoli N, Salari N, Darvishi N, Jafarpour S, Solaymani M, Mohammadi M, Shohaimi S. The global prevalence of major depressive disorder (MDD) among the elderly: A systematic review and meta-analysis. Neurosci Biobehav Rev. 2022;132:1067–73.
3. Shorey S, Ng ED, Wong CHJ. Global prevalence of depression and elevated depressive symptoms among adolescents: A systematic review and meta-analysis. Br J Clin Psychol. 2022;61(2):287–305.
4. Moreno-Agostino D, Wu YT, Daskalopoulou C, Hasan MT, Huisman M, Prina M. Global trends in the prevalence and incidence of depression: a systematic review and meta-analysis. J Affect Disord. 2021;281:235–43.
5. Cambridge OR, Knight MJ, Mills N, Baune BT. The clinical relationship between cognitive impairment and psychosocial functioning in major depressive disorder: A systematic review. Psychiatry Res. 2018;269:157–71.
6. Bird T, Tarsia M, Schwannauer M. Interpersonal styles in major and chronic depression: A systematic review and meta-analysis. J Affect Disord. 2018;239:93–101.
7. Berk M, Kohler-Forsberg O, Turner M, Penninx B, Wrobel A, Firth J, Loughman A, Reavley NJ, McGrath JJ, Momen NC, et al. Comorbidity between major depressive disorder and physical diseases: a comprehensive review of epidemiology, mechanisms and management. World Psychiatry. 2023;22(3):366–87.
8. Zainal NH, Newman MG. Executive function and other cognitive deficits are distal risk factors of generalized anxiety disorder 9 years later. Psychol Med. 2018;48(12):2045–53.
9. Turkheimer FE, Veronese M, Mondelli V, Cash D, Pariante CM. Sickness behaviour and depression: an updated model of peripheral-central immunity interactions. Brain Behav Immun. 2023;111:202–10.
10. Dantzer R, O'Connor JC, Freund GG, Johnson RW, Kelley KW. From inflammation to sickness and depression: when the immune system subjugates the brain. Nat Rev Neurosci. 2008;9(1):46–56.
11. Zainal NH, Newman MG. Increased inflammation predicts nine-year change in major depressive disorder diagnostic status. J Abnorm Psychol. 2021;130(8):829–40.
12. Mooney SJ, El-Sayed AM. Stigma and the etiology of depression among the obese: an agent-based exploration. Soc Sci Med. 2016;148:1–7.
13. Smith NB, Monteith LL, Rozek DC, Meuret AE. Childhood abuse, the interpersonal-psychological theory of suicide, and the mediating role of depression. Suicide Life-Threatening Behav. 2018;48(5):559–69.
14. Bolger KE, Patterson CJ. Pathways from child maltreatment to internalizing problems: perceptions of control as mediators and moderators. Dev Psychopathol. 2001;13(4):913–40.
15. Pak G, Bae SM. Influences of loneliness, life purpose, and aging satisfaction on depression in older US adults: analysis of 12-year longitudinal data. Int J Ment Health Addict. 2025;23(2):1566–1578.
16. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull. 1985;98(2):310–57.
17. Liu RT, Kleiman EM, Nestor BA, Cheek SM. The hopelessness theory of depression: A quarter century in review. Clin Psychol (New York). 2015;22(4):345–65.
18. Aneshensel CS, Stone JD. Stress and depression: a test of the buffering model of social support. Arch Gen Psychiatry. 1982;39(12):1392–6.
19. Eberhart NK, Hammen CL. Interpersonal style, stress, and depression: an examination of transactional and diathesis-stress models. J Soc Clin Psychol. 2010;29(1):23–38.
20. Bentley KH, Gallagher MW, Boswell JF, Gorman JM, Shear MK, Woods SW, Barlow DH. The interactive contributions of perceived control and anxiety sensitivity in panic disorder: A triple vulnerabilities perspective. J Psychopathol Behav Assess. 2012;35(1):57–64.
21. Mac Giollabhui N, Ng TH, Ellman LM, Alloy LB. The longitudinal associations of inflammatory biomarkers and depression revisited: systematic review, meta-analysis, and meta-regression. Mol Psychiatry. 2021;26(7):3302–14.
22. Mannan M, Mamun A, Doi S, Clavarino A. Is there a bi-directional relationship between depression and obesity among adult men and women? Systematic review and bias-adjusted meta analysis. Asian J Psychiatry. 2016;21:51–66.
23. McKay MT, Cannon M, Chambers D, Conroy RM, Coughlan H, Dodd P, Healy C, O'Donnell L, Clarke MC. Childhood trauma and adult mental disorder: A systematic review and meta-analysis of longitudinal cohort studies. Acta Psychiatrica Scandinavica. 2021;143(3):189–205.
24. Liu Y, Zhang N, Bao G, Huang Y, Ji B, Wu Y, Liu C, Li G. Predictors of depressive symptoms in college students: A systematic review and meta-analysis of cohort studies. J Affect Disord. 2019;244:196–208.
25. Sowislo JF, Orth U. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. Psychol Bull. 2013;139(1):213–40.

26. Msetfi RM, Kornbrot DE, Halbrook YJ. The association between the sense of control and depression during the COVID-19 pandemic: a systematic review and meta-analysis. Front Psychiatry. 2024;15:1323306.
27. Jacobson NC, Newman MG. Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. Psychol Bull. 2017;143(11):1155–200.
28. Wang Y, Chung MC, Wang N, Yu X, Kenardy J. Social support and posttraumatic stress disorder: A meta-analysis of longitudinal studies. Clin Psychol Rev. 2021;85:101998.
29. Rueger SY, Malecki CK, Pyun Y, Aycock C, Coyle S. A meta-analytic review of the association between perceived social support and depression in childhood and adolescence. Psychol Bull. 2016;142(10):1017–67.
30. Belvederi Murri M, Cattelani L, Chesani F, Palumbo P, Triolo F, Alexopoulos GS. Risk prediction models for depression in community-dwelling older adults. Am J Geriatric Psychiatry. 2022;30(9):949–60.
31. Giacalone M, Panarello D, Mattera R. Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators. Qual Quant. 2017;52(4):1831–59.
32. Chan JY-L, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong Z-W, Chen Y-L. Mitigating the multicollinearity problem and its machine learning approach: a review. Mathematics. 2022;10(8):1308.
33. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Ann Rev Clin Psychol. 2018;14:91–118.
34. Lewis MJ, Spiliopoulou A, Goldmann K, Pitzalis C, McKeigue P, Barnes MR. Nestedcv: an R package for fast implementation of nested cross-validation with embedded feature selection designed for transcriptomics and high-dimensional data. Bioinf Adv. 2023;3(1):vbad048.
35. Lee Y, Ragguett R-M, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, et al. Corrigendum to "Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review." J Affect Disord. 2020;274:1211–1215.
36. Cellini P, Pigoni A, Delvecchio G, Moltrasio C, Brambilla P. Machine learning in the prediction of postpartum depression: A review. J Affect Disord. 2022;309:350–7.
37. Sajjadian M, Lam RW, Milev R, Rotzinger S, Frey BN, Soares CN, Parikh SV, Foster JA, Turecki G, Muller DJ, et al. Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. Psychol Med. 2021;51(16):2742–51.
38. Fava GA, Sonino N, Lucente M, Guidi J. Allostatic load in clinical practice. Clin Psychol Sci. 2023;11(2):345–56.
39. Hammen C. Stress and depression. Ann Rev Clin Psychol. 2005;1:293–319.
40. Cavanagh A, Wilson CJ, Kavanagh DJ, Caputi P. Differences in the expression of symptoms in men versus women with depression: A systematic review and meta-analysis. Harv Rev Psychiatry. 2017;25(1):29–38.
41. Yim IS, Tanner Stapleton LR, Guardino CM, Hahn-Holbrook J, Dunkel Schetter C. Biological and psychosocial predictors of postpartum depression: systematic review and call for integration. Annu Rev Clin Psychol. 2015;11:99–137.
42. Popov S. When is unconditional self-acceptance a better predictor of mental health than self-esteem? J Rational-Emot Cognitive-Behav Ther. 2018;37(3):251–61.
43. Hakulinen C, Elovainio M, Pulkki-Råback L, Virtanen M, Kivimäki M, Jokela M. Personality and depressive symptoms: individual participant meta-analysis of 10 cohort studies. Depress Anxiety. 2015;32(7):461–70.
44. Zainal NH, Newman MG. Relation between cognitive and behavioral strategies and future change in common mental health problems across 18 years. J Abnorm Psychol. 2019;128(4):295–304.
45. Bain C, Shi D, Banad Y, Ethridge L, Norris J, Loeffelman J. A tutorial on supervised machine learning variable selection methods in classification for the social and health sciences in R. J Behav Data Sci. 2025;5(1):103–47.
46. Robila M, Robila SA. Applications of artificial intelligence methodologies to behavioral and social sciences. J Child Fam Stud. 2019;29(10):2954–66.
47. Ta V, Carrico L, Bousquet A. Binary classification: an introductory machine learning tutorial for social scientists. J Methods Meas Social Sci. 2021;12:1–39.
48. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Volume 103. New York, NY: Springer New York; 2013.
49. Ryff C, Almeida D, Ayanian J, Binkley N, Carr DS, Coe C, et al. Midlife in the United States (MIDUS 3), 2013–2014 [dataset]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR); 2019.
50. Ryff C, Almeida DM, Ayanian J, Carr DS, Cleary PD, Coe C, et al. Midlife in the United States (MIDUS 2), 2004–2006 [dataset]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR); 2017.
51. Dienberg Love G, Seeman TE, Weinstein M, Ryff CD. Bioindicators in the MIDUS National study: protocol, measures, sample, and comparative context. J Aging Health. 2010;22(8):1059–80.
52. Kurth ML, Witzel DD, Cerino ES, Almeida DM. Longitudinal changes in coping strategies across midlife and older adulthood: findings from the midlife in the united States study. Aging Ment Health. 2025;29(3):423–34.
53. Friedman EM, Christ SL, Mroczek DK. Inflammation partially mediates the association of Multimorbidity and functional limitations in a National sample of middle-aged and older adults: the MIDUS study. J Aging Health. 2015;27(5):843–63.
54. Kessler RC, Andrews G, Mroczek D, Ustun B, Wittchen H-U. The world health organization composite international diagnostic interview Short-Form (CIDI-SF). Int J Methods Psychiatr Res. 1998;7(4):171–85.
55. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Edition (DSM-III-R). 3 ed. Washington, DC: American Psychiatric Association; 1987.
56. Gigantesco A, Morosini P. Development, reliability and factor analysis of a self-administered questionnaire which originates from the world health organization's composite international diagnostic Interview - Short form (CIDI-SF) for assessing mental disorders. Clin Pract Epidemiol Mental Health. 2008;4(1):8.
57. Ospina LH, Beck-Felts K, Ifrah C, Lister A, Messer S, Russo SJ, Gross JJ, Kimhy D. Inflammation and emotion regulation: findings from the MIDUS II study. Brain Behav Immun - Health. 2022;26:100536.
58. Aw TBH, Zainal NH. Interleukin-6 moderates the relationship between social support, strain, and future depressive symptoms. Brain Behav Immun Health. 2025;49:101122.
59. Strohacker K, Wing RR, McCaffery JM. Contributions of body mass index and exercise habits on inflammatory markers: a cohort study of middle-aged adults living in the USA. BMJ Open. 2013;3(5):e002623.
60. Cohen S. Perceived stress in a probability sample of the United States. The social psychology of health. edn. Thousand Oaks, CA, US: Sage Publications, Inc; 1988. pp. 31–67.
61. Bernstein DP, Fink L. Childhood trauma questionnaire: A retrospective self-report manual. San Antonio, TX: The Psychological Corporation; 1998.
62. Finley AJ, Baldwin CL, Hebbring TM, van Reekum CM, Thayer JF, Davidson RJ, Schaefer SM. Differences in emotion expression, suppression, and cardiovascular consequences between black and white Americans in the midlife in the united States (MIDUS) study. Psychosom Med. 2024;86(9):748–57.
63. Brim OG, Ryff CD, Kessler RC. How healthy are we? Chapter one. The MIDUS National Survey: An overview. In: Brim OG, Ryff CD, Kessler RC, editors. A national study of well-being at midlife. Chicago: University of Chicago Press; 2003. p. 1–34.
64. Lachman ME, Agrigoroaei S, Tun PA, Weaver SL. Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone (BTACT). Assessment. 2014;21(4):404–17.
65. Maki KG. Social support, strain, and glycemic control: A path analysis. Personal Relationships. 2020;27(3):592–612.
66. Wolk CB, Carper MM, Kendall PC, Olino TM, Marcus SC, Beidas RS. Pathways to anxiety-depression comorbidity: A longitudinal examination of childhood anxiety disorders. Depress Anxiety. 2016;33(10):978–86.
67. Prince EJ, Siegel DJ, Carroll CP, Sher KJ, Bienvenu OJ. A longitudinal study of personality traits, anxiety, and depressive disorders in young adults. Anxiety Stress Coping. 2021;34(3):299–307.
68. McHugh RK, Weiss RD. Alcohol use disorder and depressive disorders. Alcohol Res. 2019;40(1):1–15.
69. Selzer ML. The Michigan alcoholism screening test: the quest for a new diagnostic instrument. Am J Psychiatry. 1971;127(12):1653–8.
70. Turiano NA, Whiteman SD, Hampson SE, Roberts BW, Mroczek DK. Personality and substance use in midlife: conscientiousness as a moderator and the effects of trait change. J Res Pers. 2012;46(3):295–305.
71. R Core Team. R: A language and environment for statistical computing [Software]. Version 4.4.1. Vienna, Austria: R Foundation for Statistical Computing; 2025.
72. Mayer M. missRanger: Fast imputation of missing values [R package]. Version 2.5.0. 2024.
73. Pitt TM, Hetherington E, Adhikari K, Premji S, Racine N, Tough SC, McDonald S. Developing non-response weights to account for attrition-related bias in a longitudinal pregnancy cohort. BMC Med Res Methodol. 2023;23(1):295.
74. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC Med. 2015;13:1.

75. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–73.
76. Chan JY-L, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong Z-W, Chen Y-L. Mitigating the multicollinearity problem and its machine learning approach: A review. Mathematics. 2022;10(8):1283.
77. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77.
78. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA. 2017;318(14):1377–84.
79. Koul A, Becchio C, Cavallo A. PredPsych: A toolbox for predictive machine learning-based approach in experimental psychology research. Behav Res Methods. 2018;50(4):1657–72.
80. Dorai-Raj S. Binom: binomial confidence intervals for several parameterizations [R package]. Version 1.1–1.1. In: Comprehensive R Archive Network (CRAN), editor. 2006.
81. Canty A, Ripley BD. boot: Bootstrap R (S-Plus) Functions [R package]. Version 1.3–31. 2024.
82. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc. 2020;27(4):621–33.
83. Austin PC, Steyerberg EW. The integrated calibration index (ICI) and related metrics for quantifying the calibration of logistic regression models. Stat Med. 2019;38(21):4051–65.
84. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;30:4768–77.
85. Greenwell B. fastshap: Fast approximate Shapley values [R package]. Version 0.0.7. 2022. 2021. Available from: https://cran.r-project.org/web/packages/fastshap/index.html.
86. Henninger M, Strobl C. Interpreting machine learning predictions with LIME and Shapley values: theoretical insights, challenges, and meaningful interpretations. Behaviormetrika. 2024;52(1):45–75.
87. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global Understanding with explainable AI for trees. Nat Mach Intell. 2020;2(1):56–67.
88. Collins GS, Dhiman P, Ma J, Schlussel MM, Archer L, Van Calster B, Harrell FE Jr., Martin GP, Moons KGM, van Smeden M, et al. Evaluation of clinical prediction models (part 1): from development to external validation. BMJ. 2024;384:e074819.
89. Haynos AF, Wang SB, Lipson S, Peterson CB, Mitchell JE, Halmi KA, Agras WS, Crow SJ. Machine learning enhances prediction of illness course: a longitudinal study in eating disorders. Psychol Med. 2021;51(8):1392–402.
90. Bennemann B, Schwartz B, Giesemann J, Lutz W. Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. Br J Psychiatry. 2022;220(4):192–201.
91. Buckman JEJ, Cohen ZD, O'Driscoll C, Fried EI, Saunders R, Ambler G, DeRubeis RJ, Gilbody S, Hollon SD, Kendrick T, et al. Predicting prognosis for adults with depression using individual symptom data: a comparison of modelling approaches. Psychol Med. 2023;53(2):408–18.
92. Sterkenburg TF, Grünwald PD. The no-free-lunch theorems of supervised learning. Synthese. 2021;199(3–4):9979–10015.
93. Andersson S, Bathula DR, Iliadis SI, Walter M, Skalkidou A. Predicting women with depressive symptoms postpartum with machine learning methods. Sci Rep. 2021;11(1):7877.
94. Clapp MA, Castro VM, Verhaak P, McCoy TH, Shook LL, Edlow AG, Perlis RH. Stratifying risk for postpartum depression at time of hospital discharge. Am J Psychiatry. 2025;182(6):551–9.
95. Hochman E, Feldman B, Weizman A, Krivoy A, Gur S, Barzilay E, Gabay H, Levy J, Levinkron O, Lawrence G. Development and validation of a machine learning-based postpartum depression prediction model: A nationwide cohort study. Depress Anxiety. 2021;38(4):400–11.
96. Evans L, Wu Y, Xi W, Ghosh AK, Kim MH, Alexopoulos GS, Pathak J, Banerjee S. Risk stratification models for predicting preventable hospitalization in commercially insured late middle-aged adults with depression. BMC Health Serv Res. 2023;23(1):621.
97. Perlis RH. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. Biol Psychiatry. 2013;74(1):7–14.
98. Wang YF, Li XH, Zhou XY, Ke QQ, Ma HL, Li ZH, Zhuo YS, Liu JY, Liu XL, Yang QH. Development and validation of machine learning models based on stacked generalization to predict psychosocial maladjustment in patients with acute myocardial infarction. BMC Psychiatry. 2025;25(1):152.
99. Walsh CG, Sharman K, Hripcsak G. Beyond discrimination: A comparison of calibration methods and clinical usefulness of predictive models of readmission risk. J Biomed Inform. 2017;76:9–18.
100. Harkness KL, Hayden EP, Lopez-Duran NL. Stress sensitivity and stress sensitization in psychopathology: an introduction to the special section. J Abnorm Psychol. 2015;124(1):1–3.
101. Stroud CB, Davila J, Hammen C, Vrshek-Schallhorn S. Severe and nonsevere events in first onsets versus recurrences of depression: evidence for stress sensitization. J Abnorm Psychol. 2011;120(1):142–54.
102. Biltz RG, Sawicki CM, Sheridan JF, Godbout JP. The neuroimmunology of social-stress-induced sensitization. Nat Immunol. 2022;23(11):1527–35.
103. Jorgensen M, Smith ORF, Wold B, Haug E. Social inequality in the association between life transitions into adulthood and depressed mood: a 27-year longitudinal study. Front Public Health. 2024;12:1286554.
104. Bailey RK, Mokonogho J, Kumar A. Racial and ethnic differences in depression: current perspectives. Neuropsychiatr Dis Treat. 2019;15:603–9.
105. Compas BE, Jaser SS, Bettis AH, Watson KH, Gruhn MA, Dunbar JP, Williams E, Thigpen JC. Coping, emotion regulation, and psychopathology in childhood and adolescence: A meta-analysis and narrative review. Psychol Bull. 2017;143(9):939–91.
106. Garramone F, Baiano C, Russo A, D'Iorio A, Tedeschi G, Trojano L, Santangelo G. Personality profile and depression in migraine: a meta-analysis. Neurol Sci. 2020;41(3):543–54.
107. Yapko MD. Disturbances of temporal orientation as a feature of depression. In: Yapko MD, editor. Brief therapy approaches to treating anxiety and depression. 1st ed. Routledge; 1990. p. 13.
108. Iovoli F, Rubel JA, Steinbrenner T, Lauterbach R. Interpersonal problems and their mental health correlates: a meta-analytic review. J Clin Psychol. 2025;81(11):1046–1056.
109. Siddaway AP, Taylor PJ, Wood AM, Schulz J. A meta-analysis of perceptions of defeat and entrapment in depression, anxiety problems, posttraumatic stress disorder, and suicidality. J Affect Disord. 2015;184:149–59.
110. Lamp SJ, MacKinnon DP. Correcting for collider effects and sample selection bias in psychological research. Psychol Methods. 2025;30(5):1169–84.
111. Cornish RP, Bartlett JW, Macleod J, Tilling K. Complete case logistic regression with a dichotomised continuous outcome led to biased estimates. J Clin Epidemiol. 2023;154:33–41.
112. Nickson D, Meyer C, Walasek L, Toro C. Prediction and diagnosis of depression using machine learning with electronic health records data: a systematic review. BMC Med Inf Decis Mak. 2023;23(1):271.
113. Luedtke A, Sadikova E, Kessler RC. Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. Clin Psychol Sci. 2019;7(3):445–61.
114. Alves-Bradford JM, Trinh NH, Bath E, Coombs A, Mangurian C. Mental health equity in the twenty-first century: setting the stage. Psychiatr Clin North Am. 2020;43(3):415–28.
115. Monteith S, Glenn T, Geddes J, Whybrow PC, Achtyes E, Bauer M. Expectations for artificial intelligence (AI) in psychiatry. Curr Psychiatry Rep. 2022;24(11):709–21.
116. Starke G, Schmidt B, De Clercq E, Elger BS. Explainability as Fig leaf? An exploration of experts' ethical expectations towards machine learning in psychiatry. AI Ethics. 2022;3(1):303–14.

Publisher's note