JAMA Psychiatry | Original Investigation

Discriminating Heterogeneous Trajectories of Resilience and Depression After Major Life Stressors Using Polygenic Scores

Katharina Schultebraucks, PhD; Karmel W. Choi, PhD; Isaac R. Galatzer-Levy, PhD; George A. Bonanno, PhD

IMPORTANCE Major life stressors, such as loss and trauma, increase the risk of depression. It is known that individuals show heterogeneous trajectories of depressive symptoms following major life stressors, including chronic depression, recovery, and resilience. Although common genetic variation has been associated with depression risk, genomic factors that could help discriminate trajectories of risk vs resilience following adversity have not been identified.

OBJECTIVE To assess the discriminatory accuracy of a deep neural net combining joint information from 21 psychiatric and health-related multiple polygenic scores (PGSs) for discriminating resilience vs other longitudinal symptom trajectories with use of longitudinal, genetically informed data on adults exposed to major life stressors.

DESIGN, SETTING, AND PARTICIPANTS The Health and Retirement Study is a longitudinal panel cohort study in US citizens older than 50 years, with data being collected once every 2 years between 1992 and 2010. A total of 2071 participants who were of European ancestry with available depressive symptom trajectory information after experiencing an index depressogenic major life stressor were included. Latent growth mixture modeling identified heterogeneous trajectories of depressive symptoms before and after major life stressors, including stable low symptoms (ie, resilience), as well as improving, emergent, and preexisting/chronic symptom patterns. Twenty-one PGSs were examined as factors distinctively associated with these heterogeneous trajectories. Local interpretable model-agnostic explanations were applied to examine PGSs associated with each trajectory. Data were analyzed using the DNN model from June to July 2020.

EXPOSURES Development of depression and resilience were examined in older adults after a major life stressor, such as bereavement, divorce, and job loss, or major health events, such as myocardial infarction and cancer.

MAIN OUTCOMES AND MEASURES Discriminatory accuracy of a deep neural net model trained for the multinomial classification of 4 distinct trajectories of depressive symptoms (Center for Epidemiologic Studies–Depression scale) based on 21 PGSs using supervised machine learning.

RESULTS Of the 2071 participants, 1329 were women (64.2%); mean (SD) age was 55.96 (8.52) years. Of these, 1638 (79.1%) were classified as resilient, 160 (7.75) in recovery (improving), 159 (7.7%) with emerging depression, and 114 (5.5%) with preexisting/chronic depression symptoms. Deep neural nets distinguished these 4 trajectories with high discriminatory accuracy (multiclass micro-average area under the curve, 0.88; 95% CI, 0.87-0.89; multiclass macro-average area under the curve, 0.86; 95% CI, 0.85-0.87). Discriminatory accuracy was highest for preexisting/chronic depression (AUC 0.93), followed by emerging depression (AUC 0.88), recovery (AUC 0.87), resilience (AUC 0.75).

CONCLUSIONS AND RELEVANCE The results of the longitudinal cohort study suggest that multivariate PGS profiles provide information to accurately distinguish between heterogeneous stress-related risk and resilience phenotypes.

JAMA Psychiatry. doi:10.1001/jamapsychiatry.2021.0228
Published online March 31, 2021.

Supplemental content

Author Affiliations: Department of Emergency Medicine, Columbia University Irving Medical Center, New York, New York (Schultebraucks); Data Science Institute, Columbia University, New York, New York (Schultebraucks); Department of Psychiatry, NYU Grossman School of Medicine, New York, New York (Schultebraucks); Department of Psychiatry, Massachusetts General Hospital, Boston (Choi); Department of Psychiatry, NYU Grossman School of Medicine, New York, New York (Galatzer-Levy); Teachers College, Columbia University, New York, New York (Bonanno).

Corresponding Author: Katharina Schultebraucks, PhD, Department of Emergency Medicine, Columbia University Medical Center, 622 W 168th St, New York, NY 10032 (ks3796@cumc.columbia.edu).

Exposure to major life stressors, such as bereavement,[1] divorce,[2] and job loss,[3] can increase the risk of major depression.[4] Similarly, major health events, such as myocardial infarction[5] and cancer,[6] can increase the risk for psychiatric disorders. Nonetheless, it is well established that psychological responses to such events tend to follow heterogeneous symptom trajectories.[7,8] Some individuals exposed to major life stressors will exhibit persistent symptom elevations (chronic), while others will show initially high symptoms that decrease with time (recovery) or low-to-moderate symptoms that worsen with time (emergent). However, the most common trajectory is one of stable good mental health (resilience).[7]

Although a number of factors associated with these trajectory patterns have been identified—ranging from personality and behavioral factors to neurobiological markers[9-12]—their individual effects have been modest, suggesting that other key factors may be needed to increase explanatory power.[6,13,14] One potential yet unexplored strategy for discriminating resilience and risk trajectories is using genetic factors that may underlie these potential adaptive mechanisms. The regularity with which these trajectories have been observed across a wide range of adverse life events[7,13,14] is consistent with possible underlying genetic factors. Moreover, evidence suggests that resilience, although multifactorial, has a notable genetic component. Twin studies indicate that 31% to 52% of observed variance in resilience phenotypes (eg, positive psychological functioning despite life stressors) can be explained by genetic differences.[15-17] However, the extent to which genetic data can be used to accurately discriminate more precisely defined longitudinal trajectories of resilience to adversity is not well understood.

Genome-wide association studies (GWAS) have been used to estimate associations between millions of genetic variants and diverse phenotypes, ranging from health conditions to psychiatric disorders such as schizophrenia and major depression.[18-23] Polygenic scores (PGSs)[24] aggregate genome-wide contributions into an overall score reflecting an individual's overall genetic propensity for a given trait or disorder[25] and have been shown to explain the risk of various psychiatric disorders.[26-28] In addition, PGS have been used as genetic proxies of health-related, cognitive, and mental health traits to enable the identification of underlying factors that might contribute to psychosocial outcomes.[29] Although the explanatory power of individual PGSs is modest, combining multiple PGSs has been applied to increase the predictive power for psychosocial outcomes.[29,30] Despite the utility of combining multiple PGSs, to our knowledge, this approach has not yet been used to distinguish multinomial trajectory outcomes following adversity. Doing so can lead to several benefits, including a better understanding of the utility of genetic factors associated with resilience and enhancing the programmatic identification and allocation of clinical resources for those in need. However, to date, genomic studies of psychological resilience have relied on outcomes defined by cross-sectional designs or, at most, prospective studies with data from baseline and a single follow-up.[31]

To address this gap, powerful computational methods are required that combine multiple PGSs and can accurately discriminate between longitudinal trajectories of resilience and depression after major life stressors. Data-driven deep learning is particularly useful for handling the joint probabilistic information of multiple PGSs. This method allows for potential nonlinear and higher-order dependencies, handles multicollinearity, requires no assumptions about the association between different PGSs, and is well equipped to maximize the discriminatory accuracy required for the multinomial classification of heterogeneous trajectories.

This study used large-scale GWAS results for psychiatric and health phenotypes and a genetically informed, longitudinal cohort of adults exposed to major life stressors. Specifically, trajectories of depression symptoms were based on the results of 5 previous studies in a nationally representative cohort of adults[1-3,5,6] in which depression has been measured before and following major life stressors. Using these data, we tested the discriminatory accuracy of a deep learning model combining joint information from 21 psychiatric and health-related PGSs for discriminating resilience vs other longitudinal trajectory patterns.

Key Points

Question Is it possible to accurately discriminate longitudinal trajectories of depression and resilience by using multiple polygenic scores as psychiatric risk and health indicators?

Findings In this longitudinal cohort study including 2071 participants, resilience and symptomatic trajectories were accurately discriminated using 21 polygenic scores using deep neural nets. The resilience trajectory was associated with lower polygenic scores for several psychiatry disorders as well as metabolic risk.

Meaning The results of this study suggest that polygenic scores can be used to determine long-term risk for depression and resilience.

Methods

We included data from 2071 participants from the US nationally representative cohort of older adults in the Health and Retirement Study (HRS).[32,33] The HRS is a longitudinal prospective study of US citizens born between 1931 and 1947 with data collected once every 2 years between 1992 and 2010.[32] Data were analyzed using the DNN model from June to July 2020. Participants were surveyed on mental and physical health-related aspects as described in detail elsewhere.[34] The HRS measures depressive symptom severity using the abbreviated 10-item version[35] of the Center for Epidemiologic Studies-Depression (CES-D) scale.[36] The optimal cutoff score of the 10-item CES-D is 4, with scores greater than or equal to 4 indicating depression with a sensitivity of 97%, a specificity of 84%, and a positive predictive value of 85%.[37] For the present study, we included all participants of European ancestry with available depressive symptom trajectory information who experienced 1 of the following index depressogenic major life stressors: bereavement,[1] myocardial infarction,[5] divorce,[2]

cancer,[6] or job loss.[3] Participants who experienced more than 1 of these major life stressors during this period were excluded from the analysis to define a cohort of patients with a single index event as a common reference point for the longitudinal assessment of pre- and post-event follow-up. eFigure 1 in the Supplement presents a flowchart describing the sample selection. The HRS data are deidentified and publicly available, and all participants provided written informed consent in the HRS study; participants received financial compensation in the HRS study. Our study was a secondary data analysis for which no reimbursement was paid.[32,33] This secondary data analysis was exempt from institutional review board approval in accordance with the policies of the New York University Institutional Review Board and the Teachers College Columbia University Institutional Review Board.

The outcome of interest was defined as the longitudinal trajectory course of depressive symptoms measured before and in the years following an index major life stressor. We combined data from 5 different samples from the HRS[1-3,5,6] that all previously identified the same 4 prospective trajectories of depressive symptoms: resilience, characterized as a stable trajectory of low symptom severity before and following the index event; recovery, characterized by initial clinically elevated symptoms that steadily decrease following the index event; emerging depression, characterized by low to moderate symptoms that increased above the clinically significant threshold following the index event; and preexisting and chronic depression marked by clinically elevated symptoms before and subsequent to the index event. All trajectories were identified using latent growth mixture modeling[38] with a floating baseline method[39] in which participants' data were centered on the year of the index event.[39] **Figure 1** shows depression symptom trajectories at 1 time point before and 2 time points following the index stressor. In the present study, we examined all participants previously assigned to 1 of these 4 latent growth mixture modeling trajectories with available PGSs. Further details are reported in the eMethods and eResults in the Supplement.

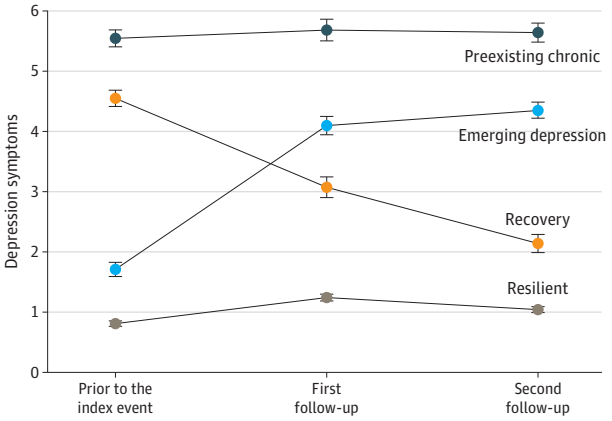
We included 21 different PGSs as candidate features (**Table 1**). These PGSs were selected for their potential relevance for discriminating heterogeneous stress responses based on previous analyses and publications[1-3,5,6] of the HRS data set.[40] Details on genomic data processing and PGS construction in the HRS are provided elsewhere.[56] A brief description is presented in the eMethods in the Supplement.

Statistical Analysis

We applied supervised learning for multinomial classification using a multilayer feedforward neural network to train a deep neural net (DNN). The primary outcome of this study was the multinomial classification task to differentiate the 4 symptom trajectories based on the 21 PGSs (Table 1), age, and the types of major life stressors. Benchmark models as well as further technical details about feature preprocessing, model development, and model validation are presented in the eMethods and eResults in the Supplement.

To examine why the model assigned each participant to a given latent trajectory of depression symptoms, we applied methods for explainable machine learning. Local interpretable model-agnostic explanations (LIME) via submodular optimization[57] were used for the explanations of classifications in human-interpretable form by approximating the output of the DNN model locally with penalized general linear models. The results were averaged per latent growth mixture modeling class to estimate which set of features, on average, influenced the classification of depressive symptom trajectories the most. The trajectories were built using MPlus, version 7.3,[58] the DNN was built using Keras Tensorflow, version 2.1.0 in Python 3.7,[59] and nested cross-validation was performed using Scikit-learn, version 0.23.[60]

Figure 1. Trajectories of Depression Using Combined Health and Retirement Study Data



The points represent the mean (SE) Center for Epidemiologic Studies-Depression (CES-D) score per trajectory class at each time point (a 10-item scale, with scores greater than or equal to 4 indicating depression). Depressive symptoms were measured every 2 years in the Health and Retirement Study. The index major life stressor occurred between the *Prior to the index event* and *First follow-up* points.

Results

We examined a set of 21 PGSs as candidate features (Table 1) in a sample of 2071 participants from the HRS cohort. The cohort included 1329 women (64.2%) and 739 men (35.7%); mean (SD) age was 55.96 (8.52) years. **Table 2** provides other descriptive statistics, including years of education, for each trajectory.

Neural networks achieved good discriminatory power to distinguish all 4 trajectories with a multiclass macro-average AUC of 0.86 (95% CI, 0.85-0.87) and a micro-average AUC of 0.88 (95% CI, 0.87-0.89) (**Figure 2**A; average precision, 0.79; average recall, 0.60; average specificity, 0.82; average F1, 0.64; average geometric mean, 0.70; and average index balanced accuracy, 0.48). All 4 trajectories were classified with high discriminatory accuracy (Figure 2B). A comparison of the nested vs nonnested cross-validation performance is shown in eFigure 2 on the Supplement.

The 15 most important features identified using LIME for each trajectory are shown in **Figure 3**. LIME feature impor-

Table 1. Overview About the PGS Candidate Features in the Data Set[a]

Source	PGS[a]	Phenotype
Wray et al,[41] 2018	EA_PGS3_MDD2_PGC18	Major depressive disorder
Schizophrenia Working Group of the Psychiatric Genomics Consortium,[18] 2014	PGS_SCZ_PGC14	Schizophrenia
Duncan et al,[42] 2018	EA_PGS3_PTSDEA_PGC18	Posttraumatic stress disorder
Demontis et al,[43] 2019	EA_PGS3_ADHD_PGC17	Attention-deficit/hyperactivity disorder
Arnold et al,[44] 2018	EA_PGS3_OCD_IOCDF17	Obsessive-compulsive disorder
Psychiatric GWAS Consortium Bipolar Disorder Working Group,[20] 2011	EA_PGS3_BIP_PGC11	Bipolar disorder
de Moor et al,[45] 2015	PGS_neuroticism_SSGAC16	Neuroticism
Ripke et al,[46] 2013	PGS_depsymp_SSGAC16	Depressive symptoms
Otowa et al,[47] 2016	EA_PGS3_ANXFS_ANGST16	Anxiety symptoms continuous
Okbay et al,[48] 2016	PGS_well-being_SSGAC16	Well-being
	EA_PGS3_EXTRAV_GPC17	Extraversion
Davies et al,[49] 2015	PGS_GenCog_CHARGE15	Cognitive function
Lee et al,[50] 2018	PGS_EDU3_SSGAC18	Educational attainment
Bolton et al,[51] 2014	EA_PGS3_CRTSL_CORNET14	Cortisol
Furberg et al,[52] 2010	PGS_EvrSmk_TAG10	Smoking behavior
Shungin et al,[53] 2015	PGS_WC_GIANT15	Waist circumference
	PGS_WHR_GIANT15	Body fat distribution
Locke et al,[54] 2015	PGS_BMI_GIANT15	Body mass index
Willer et al,[55] 2013	EA_PGS3_HDL_GLGC13	High-density lipoprotein cholesterol
	EA_PGS3_LDL_GLGC13	Low-density lipoprotein cholesterol
	EA_PGS3_TC_GLGC13	Total cholesterol

Abbreviations: GWAS, genome-wide association study; HRS, Health and Retirement Study; PGS, polygenic score.

[a] These PGSs were selected for their potential relevance for discriminating heterogeneous stress responses based on previous analyses of the HRS data set.[40]

Table 2. Participant Characteristics

	No. (%)				
Sample characteristic	Total	Resilient	Recovery	Emerging depression	Preexisting/ chronic
No. of samples	2071 (100)	1638 (79.1)	160 (7.7)	159 (7.7)	114 (5.5)
Age in 1992, mean (SD), y	55.96 (8.52)	56.24 (8.25)	55.35 (10.13)	55.64 (9.02)	53.67 (8.51)
Sex					
Women	1329 (64.2)	1011 (61.7)	110 (68.8)	117 (73.6)	91 (79.8)
Men	739 (35.7)	625 (38.2)	50 (31.3)	42 (26.4)	22 (19.3)
Missing	3 (0.1)	2 (0.1)	0	0	1 (0.9)
Years of education, mean (SD)	12.77 (2.51)	12.96 (2.44)	12.50 (2.30)	12.14 (2.41)	11.32 (3.24)

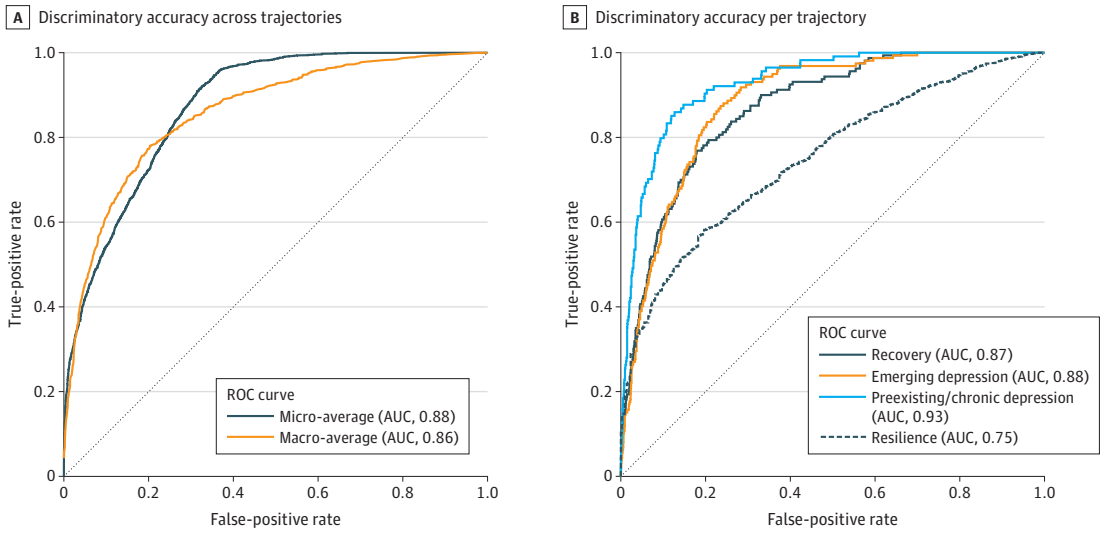
tance should not be considered a true explanation but as a heuristic approach that can lead to novel hypotheses about the input-output association.

The LIME results (Figure 3) indicated that each trajectory was characterized by a unique profile of PGSs. For example, membership in the resilience trajectory was positively associated (LIME >0) with lower PGSs for schizophrenia (≤0.39) and attention-deficit/hyperactivity disorder (≤0.46), lower-to-medium PGSs for posttraumatic stress disorder (>0.29 to ≤0.65), lower PGSs for neuroticism (>0.22 to ≤0.62), lower PGSs for waist circumference (≤0.36) and body mass index (>0.17 to ≤0.64), and high PGSs for educational attainment (>0.60). The resilience trajectory was negatively associated (LIME <0) with lower PGSs for well-being (≤0.66) and higher PGSs for total cholesterol level (>0.56) and extraversion (>0.50). By contrast, membership in the emerging depression trajectory was positively associated (LIME >0) with higher PGSs for schizophrenia (>0.57 to ≤0.80) and attention-deficit/hyperactivity disorder (>0.47 to ≤0.68) as well as body mass index (>0.64 to ≤0.89), and the recovery trajectory was associated with higher PGSs for schizophrenia (>0.57 to ≤0.80), and attention-deficit/hyperactivity disorder (>0.47 to ≤0.68), and lower PGSs for depressive symptoms (>0.30 to ≤0.57), and negatively associated (LIME <0) with higher major depressive disorder (>0.38 to ≤0.61) PGSs, and the preexisting/chronic depression trajectory was associated with higher PGSs for depressive symptoms (>0.61), and anxiety symptoms (>0.26 to ≤0.65), as well as lower educational level attainment PGSs (≤0.25).

Discussion

To our knowledge, this study is the first investigation of polygenic contributions to longitudinal trajectories of risk and

Figure 2. Receiver Operating Characteristic (ROC) Curves



Multinomial classification performance presenting the discriminatory accuracy over all trajectories (A) and the discriminatory accuracy for each individual trajectory (B). AUC indicates area under the curve.

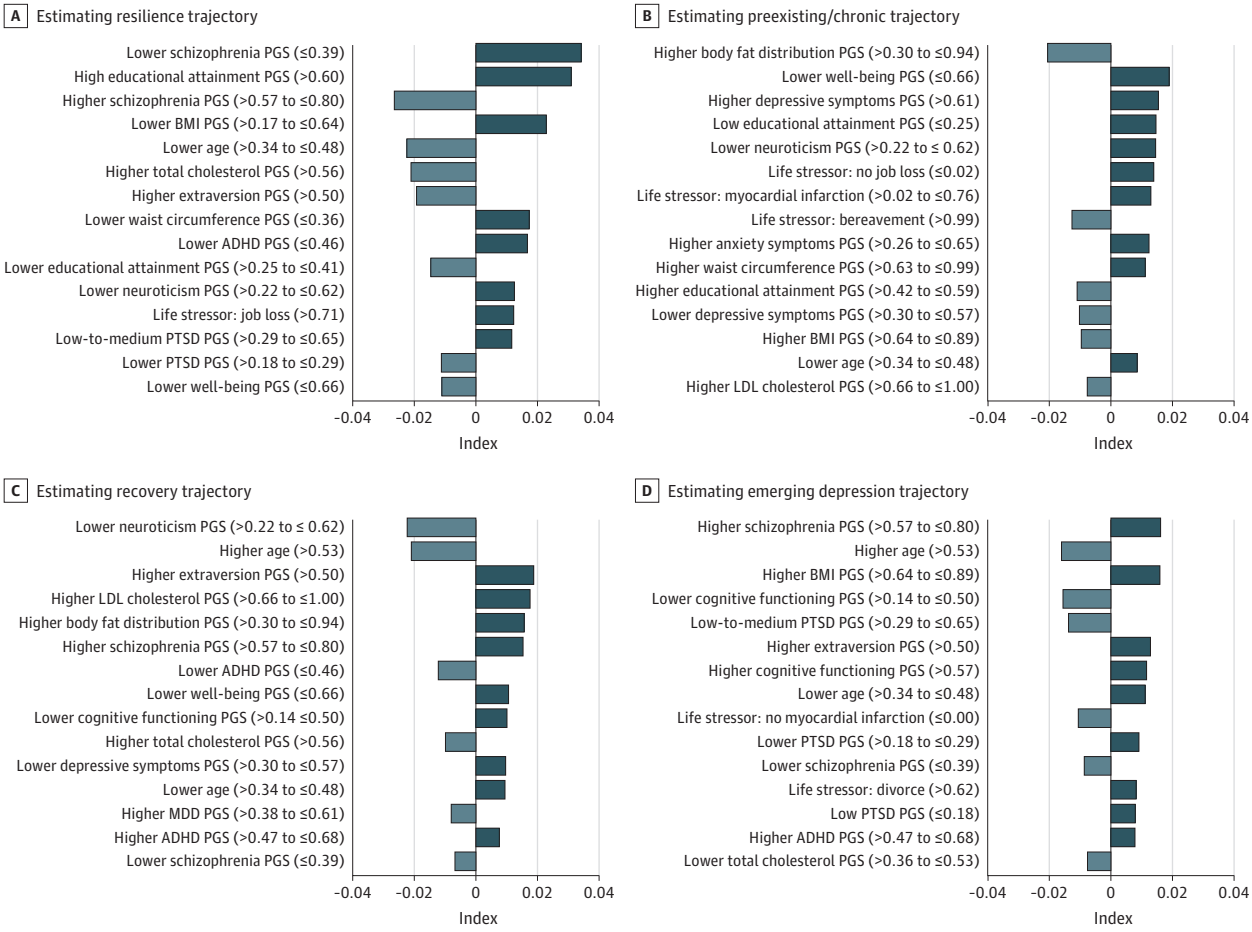
resilience following major life stressors. Previous work has identified a set of heterogeneous patterns of response to stressful life events, ranging from chronically elevated depressive symptoms to the stable absence of such symptoms (ie, resilience).[1-3,5,6] The consistency with which these trajectory patterns have been identified across diverse stressor events suggests a plausible genetic basis. Using DNNs to combine PGSs for a range of health and psychiatric traits, we were able to accurately classify longitudinal trajectories of depression-related risk and resilience following a major life stressor. The prognosis of depressive symptoms in the aftermath of major life stressors is clinically important and the accurate discrimination between distinct trajectories such as resilience and emerging depressive symptoms potentially opens new windows for targeted interventions across time. Our results show that individual PGSs alone were not able to discriminate the 4 trajectories in this sample (eResults in the Supplement). By combining different PGSs for a range of psychiatric and health-related characteristics, the computational power of a DNN model increased discriminatory accuracy and allowed us to distinguish these heterogeneous outcome patterns. Our results further show that the classification of the preexisting chronic depression was the most accurate while the classification of resilience was most difficult. This finding suggests that resilience is a more complex construct that is influenced by many risk and protective factors and it seems that the genetic component can explain only part of it, whereas in chronic depression the contribution of genetic factors for the model's performance seems to be more pronounced.

Further research to elucidate the association between PGSs and resilience is necessary to probe the mechanistic underpinnings that underlie these findings. Although such explanation is beyond the methodologic approach of this study, our results nonetheless provide a demonstration of value of multiple PGSs in accurately discriminating depression and resilience. Moreover, although PGSs do not provide specific mechanistic information, they provide accessible proxy information that may help researchers better target systems for further investigation. For example, a PGS of educational level reflects several single-nucleotide variants that represent systems involved in cognitive function and susceptibility to environmental stress. Thus, the PGS approach offers 2 advantages: the ability to compare genetic neurobiological systems in a multivariate manner and the ability to reduce the dimensionality of genetic information so that it can be used for a workable clinical risk profile.

Previous work has shown that polygenic risk for depression is associated with depression symptom trajectories across time[61] and also separately with the risk of depression after stressful life events.[62] However, to our knowledge, no published research has examined polygenic influences on depression symptom trajectories following exposure to stressful life events. Resilience to adversity is best understood as a longitudinal process rather than discrete points in time,[8,13,14] but few genetic studies have been able to examine resilience in this way.[63] By evaluating polygenic contributions to symptom trajectories in the aftermath of stressful life events, we can assess the relevance of genetic factors for their ability to discriminate longitudinal patterns of psychological response to stress, including resilience. We found that multinomial logistic regression models based only on PGSs for major depression and depressive symptoms were unable to discriminate between the 4 identified risk and resilience trajectories. Rather, as our results suggest, a broader range of PGSs and a computationally more complex model, such as deep learning, are better suited to account for such nonlinear dependencies.

The main advantage of our DNN approach is the ability to identify and use information in the data that is a priori un-

Figure 3. Variable Importance Based on Local Interpretable Model-Agnostic Explanations for Discriminating Each of the 4 Trajectories



Trajectories shown for estimating resilience trajectory (A), preexisting/chronic trajectory (B), recovery trajectory (C), and emerging depression trajectory (D). Values of a given feature are indexed between 0 and 1, with lower feature values closer to 0 and higher values closer to 1. The vertical line at the 0 mark of the x-axis represents no association, and positive values on the x-axis (ie, blue bars) represent a positive association of feature values and outcome. Negative values on the x-axis (ie, gray bars) represent a negative association of feature values and outcome. The y-axis shows the top 15 features. Because the features are used as continuous variables in the model, different value ranges can be associated both positively and negatively. For instance, in panel D, the most important feature positively associated with membership in the resilient latent growth mixture modeling class are values of the schizophrenia polygenic score (PGS) of less than or equal to 0.39, while higher schizophrenia PGSs (ie, >0.57) are negatively associated with resilience. ADHD indicates attention-deficit/hyperactivity disorder; BMI, body mass index; LDL, low-density lipoprotein; MDD, major depressive disorder; and PTSD, posttraumatic stress disorder.

known. Deep neural nets produce nonlinear mappings between the values of the candidate features and the outcome of interest.[64] Currently, genotype-phenotype association studies often use linear additive models to assess polygenic influences without accounting for potentially more complex interactions among variables, as is the case in DNNs.[64] Using flexible DNNs, we explored the ability for multiple PGSs to discriminate different trajectories of depressive symptoms and resilience. Our results suggest that, although individual PGSs show limited utility for discriminating longitudinal trajectories of risk vs resilience, combining multiple PGSs may yield informative probabilistic information for this task.

As a trade-off, a limitation of the computationally more powerful DNN approach is that results cannot be easily interpreted as linear associations between an outcome and a limited set of PGSs. Potential higher-order interactions and nonlinear association complicate the interpretations, and we did not test such higher-order interactions and nonlinear associations, nor do we see a feasible and theoretically sound possibility of modeling such associations in a cogent framework. Although deep learning results tend to be computationally demanding and difficult to interpret, promising methods have emerged to enhance interpretability of DNNs, including the strategy we used (LIME via submodular optimization) to approximate the deep learning model and estimate by rank order which variable—at a specific range of values—is most important for the classification task.

Although the identified features should not be interpreted as etiologic factors, several associations are noteworthy. For instance, lower polygenic risk for schizophrenia and attention-deficit/hyperactivity disorder were relevant for discriminating the resilience trajectory from other trajectories,

and higher polygenic risk for depression was associated with long-lasting depressive symptoms (ie, the chronic symptom trajectory). Other relevant factors associated with a resilience trajectory included higher PGSs for well-being and educational attainment, lower neuroticism PGSs, lower body mass index and waist circumference PGSs, and lower total cholesterol level PGSs. Taken together, the associations suggest that lower vulnerability for psychiatric risk and greater likelihood for well-being as indicated by PGSs increase the chances of enduring psychological health after major life stressors. Overall, these findings align with previous research on psychiatric, cognitive, and biological factors that may be relevant for stress responses. However, owing to the exploratory nature of our DNN analysis, further studies using a null-hypothesis significance testing design are needed to establish such associations with scientific confidence. The goal of this study was to assess for what we believe to be the first time whether a combination of PGSs could show utility for discriminating longitudinal trajectories of risk vs resilience. The findings suggest that polygenic information for a range of psychiatric and health traits can reveal probabilistic information to help identify subgroups of individuals who could benefit from the knowledge of their most likely response to major life stressors, including resilience, that may guide the targeting of preventive strategies.[65]

Strengths and Limitations

There are several limitations of this study. First, DNNs are less transparent compared with linear statistical models.[66] As mentioned, we used the most up-to-date methods to provide interpretable estimates. Second, we used the most recent PGSs derived and made available for research by the HRS study team.[56] For some traits, there may be even more updated GWAS based on larger samples. However, we selected the most recent available PGSs in the database where possible and provide these citations in Table 1. Although the HRS study yields comprehensive information about PGSs that have been used to build the DNN model, there are additional potentially relevant PGSs, such as more recently published major depressive disorder and other psychiatric GWAS[67] that could further increase the discriminatory accuracy of this approach. Despite these limitations, this study was conducted using a relatively large sample of adults from a well-reputed longitudinal cohort with extensive genomic and phenotypic data, allowing us to test polygenic influences on outcome trajectories following major life stressors.

Conclusions

Prospective longitudinal investigations that capture changes over time are ideal for examining the genetic influences on resilience. Drawing on a prospective population-based sample of older adults exposed to major life stressors, our results suggest how genetic information may be used to identify protective genomic factors of resilience. The algorithm can be used to discriminate distinct trajectories of depressive symptoms in response to major life stressors as diverse as bereavement, job loss, divorce, myocardial infarction, or cancer. A focus on resilience is important as it helps to identify individuals who have a lower propensity to experience stress-related psychiatric morbidity across time. This information is useful because it might lead to retargeting individuals who may benefit more from intervention and may help to prevent overtreatment or less-efficient enrollment in clinical research. To our knowledge, this represents the first investigation of the discriminatory ability of PGSs for heterogeneous trajectories following major life stressors and would benefit from replication efforts by future studies that combine genomic and longitudinal outcome data following major life stressors. Because this data-driven study is inherently exploratory, external validation of the findings is an important next step and a prerequisite before the clinical use of the model is justified. Owing to the importance of accurately distinguishing between resilience and risk for emergent depressive symptoms following major life stressors, the presented approach to combine multiple PGSs using computational methods may be a useful approach for developing prognostic models that have potential to provide new areas for targeted interventions over time.

ARTICLE INFORMATION

Accepted for Publication: February 4, 2021.

Published Online: March 31, 2021. doi:10.1001/jamapsychiatry.2021.0228

Author Contributions: Drs Schultebraucks and Bonanno had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: All authors.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Schultebraucks, Galatzer-Levy.
Obtained funding: Bonanno.
Administrative, technical, or material support: Galatzer-Levy, Bonanno.
Supervision: Bonanno.

Conflict of Interest Disclosures: Dr Galatzer-Levy has received a grant from the National Institute of Mental Health (K01MH102415). No other disclosures were reported.

REFERENCES

1. Maccallum F, Galatzer-Levy IR, Bonanno GA. Trajectories of depression following spousal and child bereavement: a comparison of the heterogeneity in outcomes. *J Psychiatr Res*. 2015;69:72-79. doi:10.1016/j.jpsychires.2015.07.017

2. Malgaroli M, Galatzer-Levy IR, Bonanno GA. Heterogeneity in trajectories of depression in response to divorce is associated with differential risk for mortality. *Clin Psychol Sci*. 2017;5(5):843-850. doi:10.1177/2167702617705951

3. Stolove CA, Galatzer-Levy IR, Bonanno GA. Emergence of depression following job loss prospectively predicts lower rates of reemployment. *Psychiatry Res*. 2017;253:79-83. doi:10.1016/j.psychres.2017.03.036

4. Breslau N. Epidemiologic studies of trauma, posttraumatic stress disorder, and other psychiatric disorders. *Can J Psychiatry*. 2002;47(10):923-929. doi:10.1177/070674370204701003

5. Galatzer-Levy IR, Bonanno GA. Optimism and death: predicting the course and consequences of depression trajectories in response to heart attack. *Psychol Sci*. 2014;25(12):2177-2188. doi:10.1177/0956797614551750

6. Burton CL, Galatzer-Levy IR, Bonanno GA. Treatment type and demographic characteristics as predictors for cancer adjustment: prospective trajectories of depressive symptoms in a population sample. *Health Psychol*. 2015;34(6):602-609. doi:10.1037/hea0000145

7. Galatzer-Levy IR, Huang SH, Bonanno GA. Trajectories of resilience and dysfunction following potential trauma: a review and statistical evaluation. *Clin Psychol Rev*. 2018;63:41-55. doi:10.1016/j.cpr.2018.05.008

8. Bonanno GA. Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? *Am Psychol*. 2004;59(1):20-28. doi:10.1037/0003-066X.59.1.20

9. Schultebraucks K, Shalev AY, Michopoulos V, et al. A validated predictive algorithm of post-traumatic stress course following emergency department admission after a traumatic stressor. *Nat Med*. 2020;26(7):1084-1088. doi:10.1038/s41591-020-0951-z

10. Schultebraucks K, Qian M, Abu-Amara D, et al. Pre-deployment risk factors for PTSD in active-duty personnel deployed to Afghanistan: a machine-learning approach for analyzing multivariate predictors. *Mol Psychiatry*. 2020;1-12.

11. Hinrichs R, van Rooij SJ, Michopoulos V, et al. Increased skin conductance response in the immediate aftermath of trauma predicts PTSD risk. *Chronic Stress (Thousand Oaks)*. 2019;3:2470547019844441. doi:10.1177/2470547019844441

12. Michopoulos V, Beurel E, Gould F, et al. Association of prospective risk for chronic PTSD symptoms with low TNFα and IFNγ concentrations in the immediate aftermath of trauma exposure. *Am J Psychiatry*. 2020;177(1):58-65. doi:10.1176/appi.ajp.2019.19010039

13. Bonanno GA, Romero SA, Klein SI. The temporal elements of psychological resilience: an integrative framework for the study of individuals, families, and communities. *Psychol Inquir*. 2015;26(2):139-169. doi:10.1080/1047840X.2015.992677

14. Bonanno GA, Westphal M, Mancini AD. Resilience to loss and potential trauma. *Annu Rev Clin Psychol*. 2011;7:511-535. doi:10.1146/annurev-clinpsy-032210-104526

15. Boardman JD, Blalock CL, Button TM. Sex differences in the heritability of resilience. *Twin Res Hum Genet*. 2008;11(1):12-27. doi:10.1375/twin.11.1.12

16. Amstadter AB, Myers JM, Kendler KS. Psychiatric resilience: longitudinal twin study. *Br J Psychiatry*. 2014;205(4):275-280. doi:10.1192/bjp.bp.113.130906

17. Waaktaar T, Torgersen S. Genetic and environmental causes of variation in trait resilience in young people. *Behav Genet*. 2012;42(3):366-377. doi:10.1007/s10519-011-9519-5

18. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

19. Hyde CL, Nagle MW, Tian C, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat Genet*. 2016;48(9):1031-1036. doi:10.1038/ng.3623

20. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet*. 2011;43(10):977-983. doi:10.1038/ng.943

21. Robinson EB, St Pourcain B, Anttila V, et al; iPSYCH-SSI-Broad Autism Group. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet*. 2016;48(5):552-555. doi:10.1038/ng.3529

22. Nagel M, Jansen PR, Stringer S, et al; 23andMe Research Team. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet*. 2018;50(7):920-927. doi:10.1038/s41588-018-0151-7

23. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5

24. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res*. 2007;17(10):1520-1528. doi:10.1101/gr.6665407

25. Sugrue LP, Desikan RS. What are polygenic scores and why are they important? *JAMA*. 2019;321(18):1820-1821. doi:10.1001/jama.2019.3893

26. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet*. 2013;9(3):e1003348. doi:10.1371/journal.pgen.1003348

27. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z

28. Wray NR, Lee SH, Mehta D, Vinkhuyzen AA, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*. 2014;55(10):1068-1087. doi:10.1111/jcpp.12295

29. Schoeler T, Choi SW, Dudbridge F, et al. Multi-polygenic score approach to identifying individual vulnerabilities associated with the risk of exposure to bullying. *JAMA Psychiatry*. 2019;76(7):730-738. doi:10.1001/jamapsychiatry.2019.0310

30. Krapohl E, Patel H, Newhouse S, et al. Multi-polygenic score approach to trait prediction. *Mol Psychiatry*. 2018;23(5):1368-1374. doi:10.1038/mp.2017.163

31. Stein MB, Choi KW, Jain S, et al. Genome-wide analyses of psychological resilience in U.S. Army soldiers. *Am J Med Genet B Neuropsychiatr Genet*. 2019;180(5):310-319. doi:10.1002/ajmg.b.32730

32. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol*. 2014;43(2):576-585. doi:10.1093/ije/dyu067

33. Juster FT, Suzman R. An overview of the Health and Retirement Study. *J Hum Resour*. 1995;30:S7-S56. doi:10.2307/146277

34. Steffick DE. Documentation of affective functioning measures in the Health and Retirement Study: HRS/AHEAD Documentation Report 2000. Accessed February 16, 2021. https://hrs.isr.umich.edu/publications/biblio/5411

35. Kohout FJ, Berkman LF, Evans DA, Cornoni-Huntley J. Two shorter forms of the CES-D (Center for Epidemiological Studies Depression) depression symptoms index. *J Aging Health*. 1993;5(2):179-193. doi:10.1177/089826439300500202

36. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. *J Youth Adolesc*. 1991;20(2):149-166. doi:10.1007/BF01537606

37. Irwin M, Artin KH, Oxman MN. Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Arch Intern Med*. 1999;159(15):1701-1704. doi:10.1001/archinte.159.15.1701

38. Muthén L, Muthén B. Mplus. Statistical analysis with latent variables: user's guide. Accessed February 16, 2021. http://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r3_web.pdf

39. Galatzer-Levy IR, Bonanno GA, Mancini AD. From marianthal to latent growth mixture modeling: a return to the exploration of individual differences in response to unemployment. *J Neuroscience Psychology Econ*. 2010;3(2):116. doi:10.1037/a0020077

40. Ware E, Schmitz L, Faul JD. HRS polygenic scores 2006-2010: genetic data. HRS documentation report. Accessed February 16, 2021. https://hrs.isr.umich.edu/news/hrs-polygenic-scores-2006-2010-genetic-data

41. Wray NR, Ripke S, Mattheisen M, et al; eQTLGen; 23andMe; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3

42. Duncan LE, Ratanatharathorn A, Aiello AE, et al. Largest GWAS of PTSD (N=20 070) yields genetic overlap with schizophrenia and sex differences in heritability. *Mol Psychiatry*. 2018;23(3):666-673. doi:10.1038/mp.2017.77

43. Demontis D, Walters RK, Martin J, et al; ADHD Working Group of the Psychiatric Genomics Consortium (PGC); Early Lifecourse & Genetic Epidemiology (EAGLE) Consortium; 23andMe Research Team. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019;51(1):63-75. doi:10.1038/s41588-018-0269-7

44. Arnold PD, Askland KD, Barlassina C, et al; International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OCGAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry*. 2018;23(5):1181-1188. doi:10.1038/mp.2017.154

45. de Moor MH, van den Berg SM, Verweij KJ, et al; Genetics of Personality Consortium. Meta-analysis of genome-wide association studies for neuroticism, and the polygenic association with major depressive disorder. *JAMA Psychiatry*. 2015;72(7):642-650. doi:10.1001/jamapsychiatry.2015.0554

46. Ripke S, Wray NR, Lewis CM, et al; Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*. 2013;18(4):497-511. doi:10.1038/mp.2012.21

47. Otowa T, Hek K, Lee M, et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol Psychiatry*. 2016;21(10):1391-1399. doi:10.1038/mp.2015.197

48. Okbay A, Baselmans BM, De Neve J-E, et al; LifeLines Cohort Study. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016;48(6):624-633. doi:10.1038/ng.3552

49. Davies G, Armstrong N, Bis JC, et al; Generation Scotland. Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53949).

Mol Psychiatry. 2015;20(2):183-192. doi:10.1038/mp.2014.188

50. Lee JJ, Wedow R, Okbay A, et al; 23andMe Research Team; COGENT (Cognitive Genomics Consortium); Social Science Genetic Association Consortium. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50(8):1112-1121. doi:10.1038/s41588-018-0147-3

51. Bolton JL, Hayward C, Direk N, et al; CORtisol NETwork (CORNET) Consortium. Genome wide association identifies common variants at the SERPINA6/SERPINA1 locus influencing plasma cortisol and corticosteroid binding globulin. *PLoS Genet*. 2014;10(7):e1004474. doi:10.1371/journal.pgen.1004474

52. Furberg H, Kim Y, Dackor J, et al; Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42(5):441-447. doi:10.1038/ng.571

53. Shungin D, Winkler TW, Croteau-Chonka DC, et al; ADIPOGen Consortium; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GEFOS Consortium; GENIE Consortium; GLGC; ICBP; International Endogene Consortium; LifeLines Cohort Study; MAGIC Investigators; MuTHER Consortium; PAGE Consortium; ReproGen Consortium. New genetic loci link adipose and insulin biology to body fat distribution. *Nature*. 2015;518(7538):187-196. doi:10.1038/nature14132

54. Locke AE, Kahali B, Berndt SI, et al; LifeLines Cohort Study; ADIPOGen Consortium; AGEN-BMI Working Group; CARDIOGRAMplusC4D Consortium; CKDGen Consortium; GLGC; ICBP; MAGIC Investigators; MuTHER Consortium; MIGen Consortium; PAGE Consortium; ReproGen Consortium; GENIE Consortium; International Endogene Consortium. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1038/nature14177

55. Willer CJ, Schmidt EM, Sengupta S, et al; Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45(11):1274-1283. doi:10.1038/ng.2797

56. Ware E, Schmitz L, Gard A, Faul JD. HRS polygenic scores—release 3, 2006-2012 genetic data. HRS documentation report. Published October 2018. Accessed January 7, 2020. https://hrsonline.isr.umich.edu/modules/meta/xyear/pgs/desc/PGENSCORES3DD.pdf

57. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. Presented at: 22nd Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining International Conference on Knowledge Discovery and Data Mining; August 2016.

58. Muthén LK, Muthén BO. *Mplus (version 7.3)*. Muthén & Muthén; 2014.

59. Keras CF. *The Python Deep Learning Library*. Astrophysics Source Code Library; 2018.

60. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.

61. Kwong ASF, López-López JA, Hammerton G, et al. Genetic and environmental risk factors associated with trajectories of depression symptoms from adolescence to young adulthood. *JAMA Netw Open*. 2019;2(6):e196587. doi:10.1001/jamanetworkopen.2019.6587

62. Choi KW, Chen C-Y, Ursano RJ, et al Prospective study of polygenic risk, protective factors, and incident depression following combat deployment in US Army soldiers. *bioRxiv*. 2018:361725. doi:10.1101/361725

63. Choi KW, Stein MB, Dunn EC, Koenen KC, Smoller JW. Genomics and psychological resilience: a research agenda. *Mol Psychiatry*. 2019;24(12):1770-1778. doi:10.1038/s41380-019-0457-6

64. Wainberg M, Merico D, Delong A, Frey BJ. Deep learning in biomedicine. *Nat Biotechnol*. 2018;36(9):829-838. doi:10.1038/nbt.4233

65. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018;19(9):581-590. doi:10.1038/s41576-018-0018-x

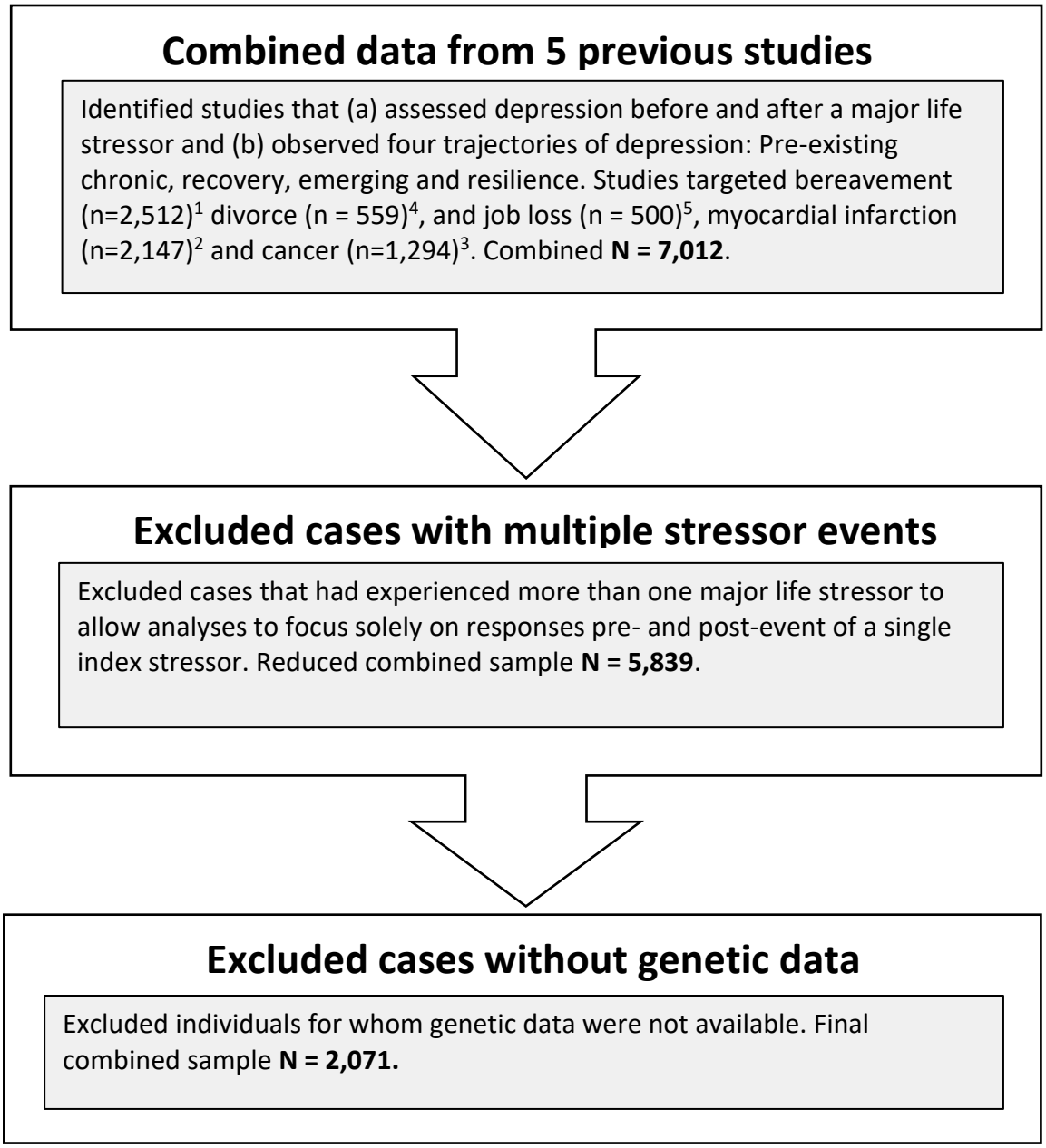
66. Lipton ZC. The mythos of model interpretability. *Queue*. 2018;16(3):31-57. doi:10.1145/3236386.3241340

67. Howard DM, Adams MJ, Clarke T-K, et al; 23andMe Research Team; Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-352. doi:10.1038/s41593-018-0326-7

Supplementary Online Content

Schultebraucks K, Choi KW, Galatzer-Levy IR, Bonanno GA. Discriminating heterogeneous trajectories of resilience and depression after major life stressors using polygenic scores. *JAMA Psychiatry*. Published online March 31, 2021. doi:10.1001/jamapsychiatry.2021.0228

This supplementary material has been provided by the authors to give readers additional information about their work.



eFigure 1. Flow chart describing the sample selection.

eMethods. Detailed Methods

Outcome used for supervised learning: trajectories of depressive symptom severity

The depression trajectories used for these analyses were created by combining results from five previous trajectory studies[1-5]. Each study used the HRS data base and identified the same four trajectory patterns (pre-existing/chronic, recovery, emerging, and resilience) using depression scores at least one time point prior to and at least two time points following a major life stressor (bereavement, divorce, job loss, myocardial infarction, and cancer diagnosis). The trajectory patterns observed in each study were highly similar to each other and to prototypical trajectory patterns repeatedly observed in previous prospective trajectory studies of major life stressors[6,7]. Each study identified trajectories using Latent Growth Mixture Modeling (LGMM) [8] and aligned data for cross-study comparison using the floating baseline methodology were time points are centered on the index stressor event[9]. Progressive model solutions in each study were determined using standard indices to assess model fit: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Sample Size Adjusted BIC (SSA-BIC), entropy values, and bootstrapped likelihood-ratio tests (B-LRT). Each study except one)[4] also utilized the Lo-Mendell-Rubin likelihood ratio test (LMR-LRT) and also reported considering interpretability and theoretical coherence in selecting final model solutions[8,8]. In considering model solutions, each study allowed the intercept parameter to freely vary and, with one exception[5], also included a quadradic parameter with fixed variance. The slope parameter was allowed to freely vary in two studies[1,2] and was fixed in three of the studies[5,3].

PGS for psychiatric and health traits

DNA samples for each HRS individual were genotyped using Illumina HumanOmni2.5 BeadChips. Quality control procedures were conducted by the HRS team to filter out individuals with problematic missingness (>2%) or first-degree relatedness in the HRS, as well as SNPs with lower call rates (<98%), violations of Hardy-Weinberg equilibrium ($p > 0.0001$), or chromosomal abnormalities. A two-step phasing and imputation process yielded roughly 21 million SNPs. Population ancestry was inferred through principal component (PC) analysis. Given that reference GWAS for scoring were developed on samples of European ancestry, only individuals of genetically determined European ancestry were retained, and top 10 ancestry-specific PCs were subsequently generated to account for residual stratification. PGS for each trait of interest were calculated as a weighted sum of alleles at a given SNP, multiplied by the estimated effect sizes for that SNP from corresponding large-scale GWAS summary statistics, including all available SNPs in common. Where HRS data had been included in a GWAS meta-analysis, summary statistics without HRS were used in order to yield independent weights. Standardized residuals of each PGS after adjustment for the top 10 ancestry-specific PCs were extracted for use as subsequent predictors.

Benchmark models

We compared model performance to benchmark models to evaluate how well the deep neural net approach can classify these trajectories compared to the performance of computationally simpler models that would be easier to interpret. Specifically, we used multinomial logistic regression to assess predictive power with only sociodemographic

variables (Table 1) or when using two candidate PGS for depressive symptoms and major depressive disorder.

Feature pre-processing, model development, and model validation

Categorical variables were dummy coded into binary values. Continuous variables were normalized to the range of [0;1]. For both, variables with near-zero variance across samples and variables with more than 35% missingness were removed. We used 10 times repeated nested cross-validation with a 3-fold inner loop and 3-fold outer loop to guard the adjustment of model weights against overfitting ("bias", i.e., a low validation error but high generalization error).

Primary outcome: multinomial classification of LGMM classes

Supervised learning adjusts the weight coefficients in such a way that the calculated estimates for the outcome is as concordant as possible[10]. In mathematical formalism, this is an optimization problem of minimizing a loss function for the mapping from inputs – the values of the candidate predictor variables – and the output of exactly one of four trajectories of depressive symptoms. Backward propagation of errors is used to solve this problem algorithmically[11].

The DNN consists of an input layer of 20 neurons, one fully connected dense hidden layer of 20 neurons with Rectified Linear Unit ('relu') activation function[12], uniform layer initialization[13], a 20%-dropout layer (i.e. randomly selecting neurons to be dropped-out with a 20% probability each epoch), and one four-neuron 'softmax' output layer to convert the multi-label classification into probability scores[14]. The DNN learned to classify

four classes and learning was performed with Keras Tensorflow 2.1.0 in Python 3.7[15]. Batch size was set to 32 and the number of epochs until convergence was limited to 200. An “epoch” is an iteration over adjusting all weights for all samples in the training dataset. Optimal weights were determined using ‘adam’ optimization using binary cross-entropy as loss function[16], which is the average of the individual cross entropies across the four outcome classes.

Nested-cross validation was implemented using scikit-learn 0.23[17] and applying the SMOTE algorithm before training the model[18]. Quite generally, cross-validation is a well-established method to train a DNN model. For cross-validation, the data is split into a subset of data to discover the optimal model parameters and a separate dataset for validating the trained model [19]. To obtain robust estimates, the process is repeated, and the average performance is reported. Since DNN models are complex with many parameters to be determined empirically during cross-validation, there are two tasks that need to be accomplished. (1) Choosing the correct model parameter (model selection) and (2) estimating the model performance accurately (performance validation). To obtain unbiased results, both tasks need to be performed separately. The nested cross-validation approach[20] accomplishes this, by splitting the data into separate data for model training and performance testing (validation) in a first step followed by another round of splitting the training data for model parameter selection in a second step[20]. This yields an inner cross-validation loop nested inside an outer loop of cross-validation. Using this nested loop structure, the model is iteratively fit in the outer loop by searching for optimal hyper-parameters (model selection) based on a given model selection criterion (e.g., “average precision") and by evaluating the model performance at each iteration on

separate data in the inner loop (performance validation). This procedure is computationally very expensive and laborious but well-suited to accomplish the two required tasks (1) model selection and (2) performance evaluation with comparatively small datasets.

In the present study, the training performance was evaluated using average precision comparing the nested cross-validation scores (inner loop) with the non-nested cross-validation scores (outer loop). The final metrics were calculated on the original data (without application of the SMOTE algorithm). The basic python code for nested CV is described in the Scikit-learn manual (last seen January 12, 2020: https://scikit-learn.org/stable/auto_examples/model_selection/plot_nested_cross_validation_iris.html)

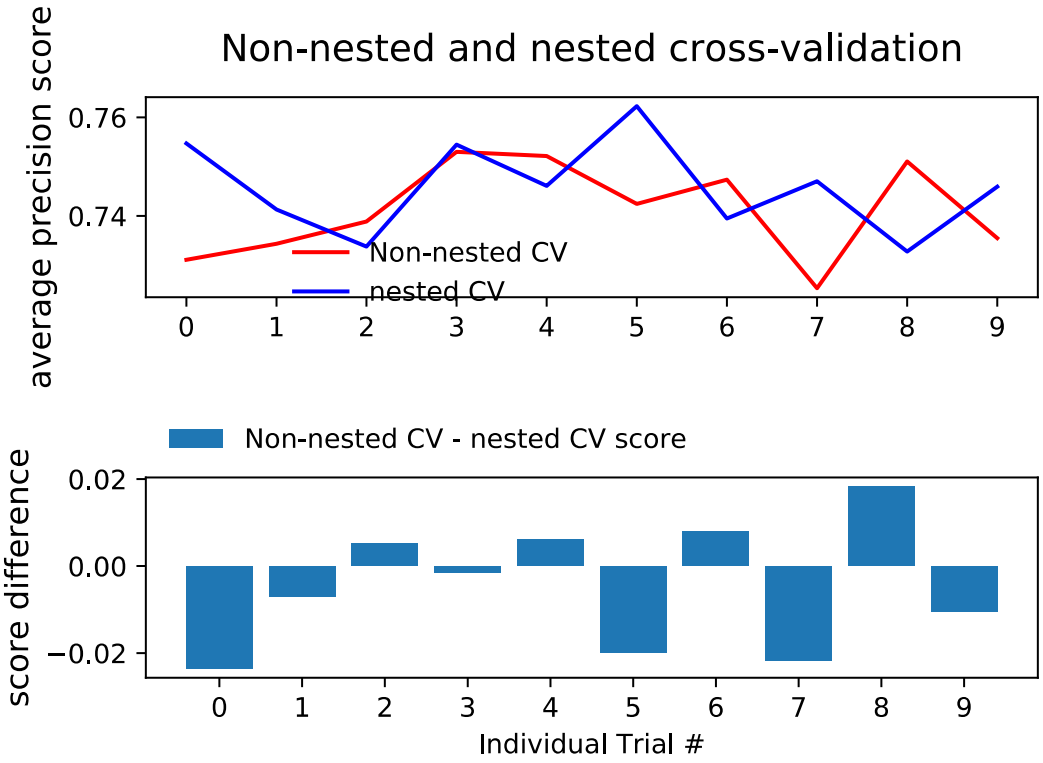
Variable importance

LIME provides “local interpretable model-agnostic” explanations by approximating the prediction of the DNN model locally. Although results cannot be interpreted causally and utmost caution should be exercised when interpreting the underlying genetic factors mechanistically, the variable importance is highly informative for gauging which variables are most important for the DNN model in making predictions. Since even small DNN models are computationally powerful enough to approximate complex non-linear functions between input and output, it is not feasible to explain how exactly the DNN works as a predictive model without making simplifications that reflect a trade-off between a better human understandable explanation, on the one hand, and a loss of formal accuracy regarding the model’s actual input and output relation on the other hand. LIME represents an established framework to explain the model’s prediction and increase

interpretability of the relationships between inputs and outputs ("model-agnostic approach")[21,22].

eResults. Detailed Results

None of the benchmark models using only PGS for depression (PGS for depressive symptoms and for major depressive disorder) or using the sociodemographic information alone (age and major life stressors) was able to discriminate between the four classes of the outcome over and beyond what would be expected by chance. The multinomial logistic regression using only the two depression PGS to predict the four trajectories yielded a chance-level multiclass AUC of 0.54 (sensitivity=0.29, specificity=0.77). The model using only sociodemographic information showed similar results (multiclass AUC=0.52, sensitivity=0.28, specificity=0.75). This indicates that the combination of multiple PGS was more informative for discriminating risk versus resilience trajectories than individual depression PGS or sociodemographic factors.



eFigure 2. Nested vs. non-nested cross-validation results showing that the model's predictions are stable for different splits of the data for training and evaluation (+/- 0.025 difference in precision). Shown are 10 comparison between 3-fold cross-validation and nested cross-validation.

eReferences:

1. Maccallum F, Galatzer-Levy IR, Bonanno GA. Trajectories of depression following spousal and child bereavement: A comparison of the heterogeneity in outcomes. *Journal of Psychiatric Research.* 2015;69:72-79.
2. Galatzer-Levy IR, Bonanno GA. Optimism and death: Predicting the course and consequences of depression trajectories in response to heart attack. *Psychological Science.* 2014;25(12):2177-2188.
3. Burton CL, Galatzer-Levy IR, Bonanno GA. Treatment type and demographic characteristics as predictors for cancer adjustment: Prospective trajectories of depressive symptoms in a population sample. *Health Psychology.* 2015;34(6):602.
4. Malgaroli M, Galatzer-Levy IR, Bonanno GA. Heterogeneity in trajectories of depression in response to divorce is associated with differential risk for mortality. *Clinical Psychological Science.* 2017;5(5):843-850.
5. Stolove CA, Galatzer-Levy IR, Bonanno GA. Emergence of depression following job loss prospectively predicts lower rates of reemployment. *Psychiatry research.* 2017;253:79-83.
6. Galatzer-Levy IR, Huang SH, Bonanno GA. Trajectories of resilience and dysfunction following potential trauma: A review and statistical evaluation. *Clin Psychol Rev.* 2018;63:41-55.
7. Bonanno GA. Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American psychologist.* 2004;59(1):20.

8. Muthén L, Muthén B. Mplus. *The comprehensive modelling program for applied researchers: user's guide.* 2020;5.
9. Galatzer-Levy IR, Bonanno GA, Mancini AD. From marianthal to latent growth mixture modeling: A return to the exploration of individual differences in response to unemployment. *Journal of Neuroscience, Psychology, and Economics.* 2010;3(2):116.
10. Svozil D, Kvasnicka V, Pospichal J. Introduction to multi-layer feed-forward neural networks. *Chemometrics and intelligent laboratory systems.* 1997;39(1):43-62.
11. Werbos P. Beyond Regression:" New Tools for Prediction and Analysis in the Behavioral Sciences. *Ph D dissertation, Harvard University.* 1974.
12. Hahnloser RH, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature.* 2000;405(6789):947.
13. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Paper presented at: Proceedings of the IEEE international conference on computer vision2015.
14. Bridle JS. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In: *Neurocomputing.* Springer; 1990:227-236.
15. Chollet F. Keras: The python deep learning library. *Astrophysics Source Code Library.* 2018.

16. De Boer P-T, Kroese DP, Mannor S, Rubinstein RY. A tutorial on the cross-entropy method. *Annals of operations research.* 2005;134(1):19-67.
17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. 2011;12:2825-2830.
18. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WPJJoair. SMOTE: synthetic minority over-sampling technique. 2002;16:321-357.
19. Kuhn M, Johnson K. *Applied predictive modeling.* Vol 26: Springer; 2013.
20. Cawley GC, Talbot NL. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research.* 2010;11(Jul):2079-2107.
21. Ribeiro MT, Singh S, Guestrin C. " Why should I trust you?" Explaining the predictions of any classifier. Paper presented at: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining2016.
22. Zhang Z, Beck MW, Winkler DA, Huang B, Sibanda W, Goyal H. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Annals of translational medicine.* 2018;6(11).

