

Data-Driven Insights into Mental Health: Mapping Multi-Level Predictors to Longitudinal Outcomes

Robert Jirsaraie

Jirsaraie@wustl.edu

Washington University in St. Louis

Deanna Barch

Washington University in St. Louis <https://orcid.org/0000-0003-1693-8506>

Ryan Bogdan

WUSTL

Scott Marek

Washington University School of Medicine

Janine Bijstervosch

Washington University, St. Louis <https://orcid.org/0000-0002-1385-9178>

Aristeidis Sotiras

Washington University in St Louis <https://orcid.org/0000-0003-0795-8820>

Nicole Karcher

Washington University School of Medicine <https://orcid.org/0000-0001-5887-5529>

Article

Keywords: Computational Models, Data Mining, Risk Factors, Biomarkers, Psychology

Posted Date: February 26th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6038031/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Mental Health on September 15th, 2025. See the published version at <https://doi.org/10.1038/s44220-025-00500-9>.

Abstract

A key challenge in predicting a person's internal state of mind is that there are a wide range of contributing factors that each have a subtle, yet significant, influence on mental health. To address this challenge, we trained machine learning algorithms on multiple sources of variation that could meaningfully contribute to psychological distress. Data mining techniques were used to identify key risk factors for predicting current symptoms and longitudinal outcomes from the Adolescent Brain Cognitive Developmental dataset (n = 11,552). Our results consistently revealed that social conflicts were the strongest indicators of psychopathology. Family fighting was the best predictor of current symptoms, whereas social victimization among peers was the top predictor of changes across time and future symptom severity. Sex-differences also emerged as a critical factor for predicting psychopathology, as females exhibited greater symptoms on average, and their symptoms became more severe over time. The long-term mental health of females was best predicted by social exclusion and reputational damage, whereas aggression and anti-social tendencies were the strongest predictors among males. While these findings provide novel insight into the developmental origins of psychopathology, our best performing models could only explain up to 40% of the variation between individuals. Future research is needed to obtain a more complete understanding of all the factors that meaningfully contribute to mental health.

INTRODUCTION

Approximately 15% of the population struggles with issues related to mental health,¹ and the number of affected individuals is trending higher in recent years.²⁻⁴ Despite the increasing prevalence of psychiatric disorders, progress has been limited in terms of developing predictive models of psychopathology that can be used to identify high risk individuals. A critical step towards this objective is to understand which risk factors that are the most predictive of psychological distress, as well as the protective mechanisms that are most essential for emotional well-being. However, a key challenge in predicting mental health is that there are a wide range of contributing factors that each have a subtle, yet meaningful, influence.⁵ Decades of research has made it clear that mental illness arises from a combination of genetic dispositions and life events, such as environmental stressors,⁶ family history,⁷ developmental processes,⁸ and psychosocial dynamics.⁹ Neuroimaging studies have also uncovered a multitude of clinical biomarkers, spanning across multiple metrics from magnetic resonance imaging (MRI), including neuroanatomical structure,¹⁰ functional activation,¹¹ and structural connectivity.¹² While each of these factors can only partially explain individual differences in symptoms of psychopathology, there is potential to obtain a more complete picture by integrating information across all levels of analyses.¹³ Here, we derived machine learning models that were based on a wide array of phenotypes and trained across a large number of individuals that were followed longitudinally as part of the Adolescent Brain Cognition Developmental (ABCD)¹⁴ study. We aimed to develop accurate models of general psychopathology and identify which features were the most predictive of individual differences and longitudinal changes.

Our study builds upon a growing body of research that used data-driven techniques to identify individuals who are at a heightened risk for a given psychiatric disorder. Much of the prior literature used classification algorithms to assign individuals into groups with varying degrees of symptom severity^{15–17} or predict the progression of symptoms over time.^{18–20} Broadly speaking, prior studies were moderately successful at separating individuals with clinically significant symptoms pertaining to psychosis-spectrum disorders (AUC > 0.70),¹⁶ bipolar disorder (AUC > 0.71),²¹ and attention-deficit hyperactivity disorder (AUC > 0.80).²² Similar accuracy levels were reported by longitudinal studies that identified which high-risk individuals would exhibit worse symptoms over time and eventually exceed clinical thresholds for a psychotic disorder (AUC > 0.73).²³ However, classification-based approaches might not be the most optimal when accounting for people who exhibit symptoms but do not meet clinical thresholds for a psychiatric disorder.²⁴ Indeed, it appears more challenging to achieve similar accuracy levels when detecting differences in temporary psychotic-like experiences (AUC > 0.60)¹⁶ or more common disorders including depression and anxiety (AUC > 0.66).¹⁷ To better account for individuals with subclinical symptoms, a regression-based algorithm may be most optimal, as it yields precise predictions about mental health along a continuous dimension.

We focused on predicting mental health differences by training regression-based models across a nationwide sample that included many individuals with subclinical and transdiagnostic symptoms (**Figure S1**). This approach aligns with the principles outlined by the Research Domain Criteria (RDoC),^{24–26} which emphasizes the need to understand dimensions of psychopathology that transcend categorical diagnostic boundaries along with their mechanisms across multiple areas of study. In further alignment with the RDoC initiative, we integrated features across nine distinct levels of analysis, including information about family history,⁷ environmental stressors,²⁷ maladaptive behaviors,²⁸ and neurobiology.¹² Our multimodal feature set consisted of 963 individual predictors, much larger than prior studies.^{15–21} Another unique element of this study is that we made a deliberate effort to avoid circularity, by excluding features pertaining to a person's thoughts and feelings, whereas prior models incorporated psychological features that were directly related to mental health.^{16–19} Consequently, prior models may not have been optimally designed to uncover novel risk factors, because they may have been biased towards clinical and psychological predictors. For example, prior literature found that the most important predictor of psychosis was experiencing bizarre or unusual thoughts,²³ and another study reported that positive feelings and being optimistic about the future was the most predictive of subjective well-being.²⁹ The focus of this research article is to identify which non-clinical factors may predispose individuals to acute psychological distress in the near and long term.

The predictive utility of our features was evaluated based on their ability to detect current symptoms and longitudinal changes throughout adolescence. Our longitudinal sample consisted of youths that were between the ages of 9 to 16 (**Figure S2**), which is a critical developmental period where there are rapid changes in terms of cognitive, affective, and social function.³⁰ This is also true for mental health, as it is estimated that 74% of young adults with a psychiatric disorder received their diagnosis before the age of

18.³¹ As such, longitudinal analyses are essential to account for emerging symptoms that were not present at baseline (**Figure S3**). We evaluated the developmental progression of symptoms by creating three sets of predictive models, with each set optimized to predict mental health at a certain point in time. First, we trained models across the entire dataset and predicted current symptoms of general psychopathology using features that were collected at the same timepoint as a given mental health assessment. Second, we fit a linear regression across all available timepoints for a given individual and used the individual's beta coefficient to quantify the rate of change in symptom severity across time. For these models, we used the same set of predictive features as before but only included the first available timepoint for each feature. Lastly, we again used the first available timepoint for each predictive feature and optimized a new set of models to predict future symptoms, as measured by the last mental health assessment, which was completed approximately four years after baseline. Each of these three mental health outcomes represented a unique perspective about the clinical characteristics of our sample.

RESULTS

Predicting Current Symptom Severity

Our first aim was to train gradient boosting (XGB) models that could predict individual differences in current symptoms of general psychopathology, as measured by the total score from the Brief Problem Monitoring self-report assessment.³² XGB models were chosen to account for non-linear relationships and complex interactions between our 963 features (see attached data dictionary), which were aggregated across nine broad categories: sMRI ($n = 170$), dMRI ($n = 160$), family ($n = 160$), fMRI ($n = 148$), environment ($n = 143$), adversity ($n = 80$), behavior ($n = 78$), demographics ($n = 32$), and cognition ($n = 14$). Predictive models that were trained across all multimodal features were the most accurate (**Figure 1**), as they yielded out-of-sample predictions that were able to explain 32% of the variation in current symptoms.

To understand which sets of features contained the most predictive utility, we trained nine unimodal models, with each model using a distinct subset of features. The accuracy of the unimodal models varied greatly in their ability to predict current symptoms with the most accurate models being derived from information about the family ($R^2 = 0.21$), which was followed by environment ($R^2 = 0.19$) and adversity ($R^2 = 0.13$). In contrast, the unimodal models that were the least predictive of current symptoms were all based on neurobiological features extracted from multimodal MRI scans ($R^2 < 0.02$). Steiger tests were used to compare differences in the coefficients of determination, which suggested that the accuracy levels of the unimodal models were each statistically different from each other ($p < 0.001$), except for the three brain-based models of psychopathology ($p > 0.056$).

[Insert Figure 1]

To further disentangle which individual features were the most predictive, we used the mean Gini impurity values from our multimodal models. This is a feature importance technique among tree-based

algorithms that quantifies how much an individual predictor was able to correctly classify or separate observations across a given dataset. Predictors frequently used to make important splits in an ensemble of decision trees are considered more important. This data-driven technique indicated that a summary score of family conflict was the single most important predictor of current symptoms, accounting for 3.72% of the predictive power across our multimodal algorithms (**Figure 2**). The second and fourth largest predictors were items from the same scale about family environment,³³ which asked adolescents about whether their family fights a lot (2.89%) or criticize each other frequently (2.00%). The most important predictor of the environmental features was related to social dynamics outside the family, namely a summary score of reputational victimization (1.76%). Similarly, losing a close friend was the most significant risk factor among the features related to adversity (1.04%). Partial dependence plots indicated that the three most important features significantly increased the predicted symptoms of generalized psychopathology (**Figure S4**).

[Insert Figure 2]

Predicting the Rate of Change in Symptoms Severity

Predictive accuracy was substantially lower when using baseline features to predict longitudinal changes in generalized psychopathology, as measured by fitting a linear regression across all mental health assessments and using the beta coefficient to quantify changes in symptom severity over time. The multimodal model, which used all 963 features, remained the most accurate and yielded predictions that explained 10% of the interindividual variability in the rate of change of symptoms across time (**Figure 3**). Comparatively, the most accurate unimodal models were those that were based on information about environment ($R^2 = 0.06$), demographics ($R^2 = 0.05$), and family ($R^2 = 0.03$). The features pertaining to adversity were much less predictive of longitudinal changes, as models trained exclusively on adversity features were the second worst overall ($R^2 < 0.01$). Neurobiological MRI features were also among the least accurate at predicting changes in symptoms ($R^2 < 0.01$). Significant testing revealed that environmental and demographics models were not statistically different from each other ($p = 0.43$), but they were significantly better than family, behavior, and cognition ($p < 0.001$), which in turn were significantly more accurate than adversity and all brain-based models ($p < 0.001$).

[Insert Figure 3]

As shown in Figure 4, Gini impurities from the multimodal models also indicated that environmental features were the most predictive, especially those pertaining to social dynamics between peers, as measured by the Peer Experiences Questionnaire (PEQ).³⁴ Specifically, a question asking adolescents whether they were left out of an activity was the most important predictor of changes in symptoms across time (1.56%). This item was used to derive a summary score of reputational victimization, which was the 3rd most important overall (0.66%). Similar items from the PEQ were also weighed heavily, including 'a kid tried to damage my social reputation by spreading rumors' (0.52%) and 'a kid gossiped about me so others would not like me' (0.45%). Aside from social conflicts between peers, biological sex

was another key factor our models used to predict changes in symptoms. Partial dependence analyses indicated that females were predicted to experience more symptoms over time ($\beta = 0.53$), whereas males were predicted to have slightly less symptoms with age ($\beta = -0.06$). This behavior by our model was substantiated by the actual changes in symptoms (**Figure S5**), which became worse among females ($\beta = 0.64$) and slightly better among males ($\beta = -0.16$). Partial dependence analyses further indicated that social exclusion and victimization increased the predicted change in symptoms, and the largest increase occurred when individuals had no conflicts vs. experiencing a couple of adverse interactions a week (**Figure S4**).

[Insert Figure 4]

Predicting Future Symptoms Severity

Compared to models predicting rate of change, we were able to derive more accurate models when using the first available timepoint to predict future symptoms, which were measured approximately four years after baseline. Multimodal models accounted for 18% of the individual differences in future symptoms (**Figure 5**), which was the most accurate. The rank ordering of the unimodal models was similar to results predicting longitudinal changes, as the most accurate feature sets still included environment ($R^2 = 0.10$), family ($R^2 = 0.06$), and demographics ($R^2 = 0.06$). However, adversity factors were more predictive of future symptoms relative to longitudinal changes, as they yielded predictions that were the 4th most accurate across all nine unimodal models ($R^2 = 0.04$). Feature sets that demonstrated the lowest accuracy were related to cognition and neurobiology ($R^2 < 0.01$). Environmental models significantly outperformed all other unimodal models ($p < 0.002$). The relative differences between the family, demographics, adversity, and behavior models were not significant ($p > 0.072$), but these models were statistically better than those that were trained using metrics of cognition or neurobiology ($p < 0.001$).

[Insert Figure 5]

As shown in Figure 6, we also observed converging results in the top predictors of future symptoms, as the most important features from prior analyses re-emerged, including reputational victimization (2.03%) and biological sex (0.86%). Novel predictors that were important to a secondary degree included a summary score of school involvement (0.65%) as well as a summary score regarding how closely parents monitored their children (0.50%). The third most important predictor was an item from the PEQ asking adolescents whether a kid tried to damage their social reputation by spreading rumors. However, partial dependence plots demonstrated that differences to this question did not have a substantial impact on predicted future symptoms (**Figure S4**), suggesting that the predictions from our model were largely based on reputational victimization and biological sex. Indeed, reputational victimization had a strong impact on the predictions of future symptoms, and females were also projected to have higher symptoms than males (**Figure S4**).

[Insert Figure 6]

Domain-Specific Models

Given that general psychopathology is a multifaceted construct made up of distinct symptom domains that are differentially associated with age,³⁵ we built domain-specific models to understand which risk factors were most predictive of inattention, externalizing, and internalizing symptoms. As performed for general psychopathology, we derived separate models to predict current symptoms, changes across time, and future symptom severity. Broadly speaking, the predictive accuracy of our domain-specific models was very heterogeneous across all sets of features and mental health outcomes (**Figure S6**). Nevertheless, a re-occurring pattern that emerged was that models predicting general psychopathology were more accurate than the domain-specific models. Family conflict and frequent fighting were still consistently the most important factors for predicting current symptoms across all dimensions of psychopathology (**Table S1**). Similarly, summary scores and items from the PEQ regarding social victimization were the most predictive of longitudinal mental health outcomes (**Table S2**). Biological sex was also ranked highly across all dimensions, especially for internalizing symptoms, as it was the most predictive of longitudinal changes and future symptoms (**Table S3**).

However, we observed a divergence among the secondary predictors for each distinct dimension of psychopathology, as operationalized by the features that were ranked as the 4th - 10th most important. For instance, models predicting internalizing symptoms were more biased towards features about family history of mental/emotional problems, parental issues with alcohol, positive views towards marijuana, and the neighborhood disadvantage score (**Tables S1-3**). In contrast, inattention models leveraged many factors related to the school, such as a summary score of school involvement, academic grades, boredom at school, relationship with teachers, and crystallized intelligence (**Tables S1-3**). Lastly, externalizing models appeared to heavily weight features about overt and reputational aggression, parent acceptance and monitoring, as well as having negative experiences with the police (**Tables S1-3**).

Sex-Specific Models

Biological sex was a key factor that our models used in their estimation of generalized psychopathology, especially regarding longitudinal outcomes. As such, we created predictive algorithms that were trained exclusively on males or females to further understand the apparent sex-differences. Nearly all ten feature sets were able to derive more accurate predictions when trained and tested on females relative to males (**Figure S7**). The largest difference in predictive accuracy was among the multimodal models predicting current symptoms, as models trained exclusively on females were the most accurate of all the algorithms that we developed in this study, explaining approximately 40% of the variability in current symptoms. Yet, the most important risk factors were similar for both males and females, which included questions about family fighting and conflict (**Table S4**). However, a notable divergence emerged in the risk factors for predicting the rate of change and future symptom severity. The risk factors for females were consistent with our main findings, as metrics about social environment were the most important, including reputational and relational victimization (**Table S4**). These summary scores were driven by item-level questions including “another kid gossiped about me so others would not like me” and “a kid

tried to damage my social reputation by spreading rumors.” In contrast, social victimization was a secondary risk factor for males, as their longitudinal mental health outcomes were best predicted by features of aggressive and hostile social tendencies (**Table S4**). For males, the following features were among the most predictive of within-subject changes and future symptom severity: “I threatened to hurt or beat up another kid” and “I left another kid out of what I was doing.”

Converging Results from Sensitivity Analyses

Our multimodal features were slightly more accurate when using an ensemble machine learning approach, which works by averaging predictions across multiple algorithms (**Figure S8**). In particular, our ability to explain differences in current symptoms of general psychopathology was 31.9% with gradient boosting and 33.4% with ensemble learning, which was a statistically significant improvement ($p = 0.04$). The relative differences between feature sets did not change, as the rank ordering of the most important feature sets matched the results produced by gradient boosting. Additional sensitivity analyses were conducted to validate the performance of our longitudinal models by controlling for baseline symptoms (**Figure S9**), evaluating bi-directional stepwise selection (**Figure S10**), and comparing pairwise-correlations (**Figure 7**). Taken together, these sensitivity analyses suggested that most of our models significantly enhanced the prediction of longitudinal outcomes, beyond simply using baseline symptoms to estimate changes over time or future symptoms. Social conflicts and biological sex also re-emerged as the key predictors (**Figure S11**).

[Insert Figure 7]

DISCUSSION

We uncovered three key risk factors that consistently emerged as the top predictors across all dimensions of psychopathology. Family conflict was most predictive of psychopathology in the near term, whereas long-term mental health outcomes were best predicted by social victimization among peers. Biological sex was another key risk factor, as females exhibited greater symptoms on average, and their symptoms became more severe with age. Despite these sex-differences, models trained exclusively on males or females both suggested that family conflict and social factors among peers were still the most predictive of psychopathology. The main difference between males and females was the specific social interactions with peers that were the most predictive of mental health challenges. Specifically, the mental health of females was best predicted by social exclusion and reputational damage, such as being the victim of unwanted gossiping and rumors. In contrast, the mental health of males was most impacted by aggression and antisocial behaviors, such as deliberately excluding others from activities and making harmful threats. These findings demonstrate that mental health is very sensitive to social environmental factors, but the specific behaviors and subjective experiences that are the most predictive largely differ between certain demographics or subpopulations.

We also found evidence of secondary risk factors that were specific to certain dimensions of psychopathology and complemented prior research findings. Positive attitudes towards marijuana were

a key feature used in the prediction of future internalizing symptoms. This result build on prior findings that indicated prenatal marijuana exposure broadly increased risk of psychopathology during middle childhood³⁶ and early adolescence.³⁷ Similarly, many clinical studies have focused on the genetic predispositions and the heritability patterns of mental illness.⁷ Factors relating to parental emotional issues were also among the most heavily weighted in the predictions of future internalizing symptoms, provide further evidence that family history may play a role for certain psychiatric disorders. Another recurring finding is that mental health issues tend to coincide with cognitive deficits.³⁸ Collectively, our cognitive features did not derive accurate models of psychopathology, but crystallized intelligence was a specific feature that had an outsized contribution to our multimodal predictions. Sensitivity analyses further uncovered that crystallized intelligence was a key factor for males and even more relevant in the predictions of inattention and internalizing symptoms.

An equally important part of our study is to acknowledge which factors did not meaningfully contribute to the predictions of psychopathology. Our results consistently conveyed that neurobiological properties were the least predictive of psychopathology, which aligns with our previous study that was conducted on adolescents and young adults.¹² Despite neurodevelopmental research uncovering many biomarkers that can significantly explain individual differences, it appears that these differences are not substantial enough for machine learning algorithms to yield accurate and generalizable predictions about mental health in early life.³⁹ Similarly, age was not a significant predictor across our models, despite its robust influences on mental health.⁵ We suspect that age-related differences would have been more important if our models were developed across a larger developmental period or during late adolescence when clinical severe symptoms are most likely to manifest.⁴⁰

The predictive capabilities of our models differed greatly in their ability to detect individual differences in current symptoms vs longitudinal outcomes. All models were more accurate at predicting current psychopathology relative to changes across time or future symptom severity. These results suggest that the early markers of psychopathology are relatively limited and not very deterministic of longitudinal outcomes. We also observed that symptom severity changed drastically throughout this developmental period with most people exhibiting less symptoms over time, which was especially true for males. Our models performed the worst at predicting longitudinal changes but were better at estimating future symptom severity at the final timepoint. The best way to estimate the trajectory of someone's symptoms over the next four years was to use their baseline symptoms. However, our multimodal model yielded predictions of future symptoms that were more accurate than relying on baseline symptom levels. This finding underscores that our multimodal models captured a unique source of variability that meaningfully contributes to psychopathology and would otherwise be missed if such risk factors were not taken into consideration. Broadly speaking, models predicting general psychopathology were more accurate than the domain-specific models, suggesting that future predictive models may benefit from targeting broader measures of mental health that reflect comorbidities across multiple dimensions.

Our best performing models could only explain up to 40% of the individual differences in psychopathology. Therefore, it is very likely that there are certain individuals with mental health issues that cannot be attributed to the risk factors that we uncovered using data driven techniques. For example, there could be important risk factors that our models did not utilize because they were uncommon among the individuals in our sample. To further explore this idea, we examined the prediction errors of our models. The two observations with the largest errors came from the same person, whereby this individual reported much higher levels of general psychopathology than predicted by our models. This person was a male from a relatively affluent family without any reported social conflicts. In fact, they had a very supportive relationship with their parent(s) and enjoyed school activities, while also scoring above average on crystallized intelligence. Unfortunately, at least one of their parents had been hospitalized and suffered from mental health issues related to anxiety and suicide. They also reported a death in the family, and at least one of their parents was away from home frequently. It is possible that this individual was grieving the loss of someone close to them, which could explain why they have some of the highest levels of psychopathology across the entire ABCD dataset. However, this unique combination of life events was relatively rare, resulting in our models not detecting that this person was experiencing high levels of psychological distress.

Our study contained several limitations. First, we grouped features into nine categories based on how closely a questionnaire fit into a given domain. Yet, there was an inevitable overlap, as certain scales contained item-level questions belonging to multiple groups or certain features reflected information that spanned across multiple phenotypes. For example, the peer experiences questionnaire contains many items about the social environment and interactions with peers, but a subset of questions is about social behaviors from the individual, rather than a reflection of their environment. Additionally, the life events scale was designed to measure adversity, but many of these questions asked about stressful experiences related to the family. As such, our unimodal models trained on subsets of features might not be the most precise, which is why we implemented additional feature importance techniques at the scale of the individual features. Second, all features were not collected at every timepoint, which likely led to some bias in our results and placed downward pressure on the predictive performance of our models. Missing data was most common at the first timepoint, especially for the most important features, which may partly explain why predictive accuracy was also the lowest at baseline (**Figure S3**). Further, our results are based on a narrow developmental stage, and it is not clear how relevant these risk factors would be during late adolescence or early adulthood. Lastly, each feature importance method has a unique set of biases and confounds, which is why we implemented three complementary techniques.

Our study uncovered key factors that were predictive of psychopathology, but future research is needed to derive models that are more accurate. This appears to be a critical challenge, as even the models from prior studies which included psychological and clinically relevant features did not yield predictions that were strongly correlated with symptoms.^{16–19} Improving model accuracy is essential if machine learning is to provide more insightful predictions that extend beyond traditional diagnostic criteria. Feature importance techniques would have greater potential to decode psychopathology and identify key

mechanisms if they are applied to highly predictive algorithms. One way to potentially develop more accurate models is by incorporating additional relevant predictors, especially those that describe a person's social environment both within the family and between peers. Another promising approach may be to derive models that are specific to certain subpopulations, because the most important risk factors for psychopathology may differ between certain demographics, as evident by the pronounced sex-differences that we delineated. We made considerable progress in understanding the developmental origins of psychopathology, but we do not yet have the full picture in terms of identifying all the factors that meaningfully contribute to mental health.

Methods

Data

We used data from the ABCD 4.0 public release (<https://dx.doi.org/10.15154/1523041>), which consists of 11,551 individuals that were repeatedly sampled between the ages of 9 to 16 years (**Figure S2**). Data was collected across nine waves, but we primarily used the following three timepoints: baseline (n = 10,977), year 2 (n = 9,711), and year 4 (n = 4,150). Data preprocessing was completed the goal of reducing missingness and maximizing the amount of data available at our three timepoints. First, we propagated data across all timepoints if a feature was non-time varying or stable across development, such as biological sex and racial identity. Second, we transposed the time-varying features by using data from the excluded timepoints to replace the missing values at our timepoints of interest. Specifically, data was either pulled forward by up to one year ("prospectively") or back by up to one year ("retroactively"). Lastly, we removed duplicated features that contained redundant information, as determined by whether they had pairwise correlation above 0.90 with any other selected variables. We report which variables underwent these preprocessing steps using data dictionaries, which can be used to identify the 148 non-time varying features, 86 retroactively transposed features, 61 prospectively transposed features, and the 286 redundant features that were excluded. Our full preprocessed dataset consisted of 24,838 timepoints from which we aggregated 963 unique features that were used to predict mental health outcomes. These preprocessing steps resulted in a more complete dataset, as on average 81% of the values were recorded across the entire sample and all features were collected for at least two of the three timepoints. This was an important detail, because we did not impute any of the missing values in our dataset, as it would be difficult to account for the longitudinal nature in which the data was collected. The demographic and clinical characteristics of this sample is reported in Table S5.

Targets

Mental health was measured dimensionally using a self-report scale known as the Youth Brief Problem Monitor,³² which was collected at all timepoints. This questionnaire contains 19-items that ask clinically relevant questions about thoughts and behaviors, such as "I am unhappy, sad, or depressed." Answers are recorded based on three ordinal responses regarding how true a given statement is, ranging from not true (0), somewhat true (1), or very true (2). The sum across all items is used to quantify general

psychopathology, reflecting the overall level of behavioral and emotional challenges. Specific dimension were also quantified by using the sum across a subset of items. The Brief Problem Monitoring contains three subscales that each reflect a unique dimension of psychopathology: inattention, externalizing, and internalizing symptoms.

Features

We aimed to incorporate as many clinically relevant features as possible into our predictive algorithms, which are described in the accompanying data dictionary released with this article. All features were categories into nine sets of phenotypes which are listed as follows in order of largest to the smallest number of features. *Structural MRI* included 170 features of cortical thickness and volumetric metrics from subcortical, cortical, and global parcels. *Diffusion MRI* contained 160 features that were properties of fractional anisotropy, mean diffusivity, and volume of white matter fiber tracts. *Family* consisted of 150 variables that included details about quality time together, home environment, relations between family members, attitudes towards family obligations, family history of psychiatric disorders, and potential family issues with substance abuse. *Functional MRI* included 148 features of the average network connectivity during resting-state scans and the beta weights of the n-back task that consisted of negative vs. neutral faces. *Environment* consisted of 142 features across school involvement, experiences with police, neighborhood safety, social dynamics between peers, and the area of deprivation indices. *Adversity* consisted of 80 features about financial hardship, witnessed violence, physical injuries, major life changes, and loss of a close relationship. *Behavior and biological information* were represented across 78 features on school attendance, physical activity, head injuries, drug usage, attitudes towards drugs, hormones, prosocial behaviors, and the parent report of screen time and social media usage. *Demographics* contained 32 predictors about age, biological sex, racial identity, sexual orientation, religious affiliations, parental education, parental marital status, and family income. *Cognition* and decision making were measured across 14 features based on performance during the delayed discounting task, cognition dice task, NIH toolbox cognitive battery, and the fifth edition of the Wechsler intelligence scale for children. Age was included as an additional predictor in each MRI feature set to account for its robust influence on neurobiology.⁴¹ In aggregate, we evaluated the predictive power of 963 unique features.

Predictive Algorithms

Extreme Gradient Boosting (XGB) was used to optimize psychopathology models, which works by constructing an ensemble of decision trees that iteratively split the data, with each successive tree aiming to reduce prediction errors from the previous one. Early stoppage was implemented so that the number of trees would dynamically increase as long as validation accuracy improved. We used shallow decision trees (max depth of 3) and a low learning rate (0.05) for increased reliability. Additional parameters were tuned to mitigate issues of overfitting by limiting model complexity through gamma, min child weight, and lasso regularization. Additional sensitivity analyses were conducted using an ensemble machine learning framework as implemented through the AutoGluon software package.⁴² A

diverse set of tree-based and deep learning algorithms were optimized using this approach, and the final predictions were calculated using weighted averaging based on each model's validation accuracy.

Cross Validation

5-fold cross validations were used to generate out-of-sample predictions across every individual in our dataset. To limit the potential of test-train leakage,⁴⁴ we grouped all timepoints belonging to an individual, such that a given participant's data was never separated between any of the train, test, or validation sets. Additionally, stratification was implemented to ensure that each split of the data maintained a similar distribution of psychopathology, which was essential to account for the imbalances across our sample. Lastly, the training data was also split so that 20% could be used as a validation set to perform hyperparameter tuning.

Accuracy Metric

Predictive accuracy of our regression-based algorithms was quantified using the coefficient of determination, because it is the only standardized performance metric that can be easily interpreted across targets with different ranges or scales. These values represent the proportion of the variance in symptoms that can be explained by predictions from a given model. If predictions perfectly matched symptoms the coefficient of determination would be 1, whereas a value of 0 would indicate that the model did not explain any variation in symptoms of psychopathology. It is possible the coefficient of determination can be negative if the model performs worse than simply using the mean of the test sample, which can also be an indicator that the model is performing worse than random chance.¹²

Feature Importance

Three techniques were used to evaluate the predictive power of our features. First, we trained unimodal models across a subset of features and used the accuracy of the predictions to quantify which features were most important at a group-level. To further understand the individual features that were most predictive, we used Gini impurities from our multimodal algorithms, trained across all features simultaneously. These values reflect how much a feature was used to make important splits across all decision trees in an XGB model. Since we derived five models to obtain out-of-sample predictions from all participants in our sample (see cross validation), the Gini impurities across all five models were averaged and subsequently multiplied by 100. As a result, the impurity of a given feature could be interpreted as its percent contribution to the predictive power across a set of models. Lastly, we used partial dependence analyses to understand how the most important features influence the predictions of psychopathology, while holding all other features constant. By examining these dependencies, we were able to gain a deeper insight into how a feature contributed to our model in terms of its magnitude and direction on predicted mental health outcomes. Altogether, these techniques provided a comprehensive assessment of the most important features in our dataset.

Statistical Analysis

We analyzed results across three sets of predictive algorithms with each set predicting a unique mental health outcome using a distinct sample composition. First, we used all 24,838 observations to train models to predict current symptoms of psychopathology based on features that were collected at the same timepoint. Second, we trained models on the first available timepoint for each feature with the goal of predicting the rate of change in symptoms across time. Linear regressions were fitted on each subject's data to quantify how much their symptoms varied by age, as measured by the beta coefficient. For these analyses, participants with one timepoint were removed ($n = 1,778$), resulting in our models being trained on baseline data from 9,599 individuals. Lastly, we again used the first available timepoint for each feature and optimized a new set of algorithms to predict future symptoms, as measured by the last mental health assessment, which was completed approximately four years after baseline (**Figure S2**). Our third set of models was trained across 3,933 individuals, as we excluded participants who did not complete a mental health assessment at the last timepoint ($n = 5,666$). To evaluate the comparative differences in model accuracy, we used Steiger tests to determine which variations in model performance were statistically significant. Qualitative insights into feature importance were provided by evaluating the features with the largest Gini impurities and performing partial dependence analyses. The sensitivity of our main findings was assessed by repeating these statistical procedures to account for individual differences at baseline, sex-differences, specific dimensions of psychopathology, and using an ensemble machine learning framework.

Declarations

Data availability

The Adolescent Brain Cognitive Developmental study is a publicly available dataset that can be accessed from the National Data Archive repository: <https://dx.doi.org/10.15154/1523041>.

Code availability

All code pertaining to this study was written in R version 4.3.1 or Python version 3.10.4 and has been made publicly available: <https://doi.org/10.5281/zenodo.14728772>.

References

1. Wright, E. *et al.* Prevalence of and Inequities in Poor Mental Health Across 3 US Surveys, 2011 to 2022. *JAMA Network Open* **8**, e2454718–e2454718 (2025).
2. Panchal, N., Rudowitz, R. & Cox, C. Recent trends in mental health and substance use concerns among adolescents. *KFF*. Accessed: May **14**, 2023 (2022).
3. Thom, J. *et al.* Trends in the Diagnostic Prevalence of Mental Disorders, 2012–2022: Using Nationwide Outpatient Claims Data for Mental Health Surveillance. *Deutsches Ärzteblatt International* **121**, 355 (2024).

4. Xiong, J. *et al.* Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of affective disorders* **277**, 55–64 (2020).
5. Casey, B. J., Oliveri, M. E. & Insel, T. A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biological psychiatry* **76**, 350–353 (2014).
6. Grant, K. E. *et al.* The influence of stressors on the development of psychopathology. in *Handbook of developmental psychopathology* 205–223 (Springer, 2014).
7. Bogdan, R., Baranger, D. A. & Agrawal, A. Polygenic risk scores in clinical psychology: bridging genomic risk to individual differences. *Annual review of clinical psychology* **14**, 119–157 (2018).
8. Pollak, S. D. Developmental psychopathology: Recent advances and future challenges. *World psychiatry* **14**, 262–269 (2015).
9. Kessler, R. C., Price, R. H. & Wortman, C. B. Social factors in psychopathology: Stress, social support, and coping processes. *Annual review of psychology* **36**, 531–572 (1985).
10. Jirsaraie, R. J., Sheffield, J. M. & Barch, D. M. Neural correlates of global and specific cognitive deficits in schizophrenia. *Schizophrenia Research* **201**, 237–242 (2018).
11. Jirsaraie, R. J. *et al.* Prenatal Exposure to Maternal Mood Entropy is Associated with a Weakened and Inflexible Salience Network in Adolescence. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2023) doi:10.1016/j.bpsc.2023.08.002.
12. Jirsaraie, R. J. *et al.* Mapping the neurodevelopmental predictors of psychopathology. *Molecular Psychiatry* (2024) doi:10.1038/s41380-024-02682-7.
13. Barch, D. M. The Power of Collaboration and Data Aggregation. *Biological Psychiatry* **84**, 626–628 (2018).
14. Casey, B. J. *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience* **32**, 43–54 (2018).
15. Moore, T. M. *et al.* Development of a probability calculator for psychosis risk in children, adolescents, and young adults. *Psychological Medicine* **52**, 3159–3167 (2022).
16. Karcher, N. R. *et al.* Examining the Most Important Risk Factors Predicting Persistent and Distressing Psychotic-like Experiences in Youth. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2024).
17. Meehan, A. J. *et al.* Developing an individualized risk calculator for psychopathology among young people victimized during childhood: A population-representative cohort study. *Journal of Affective Disorders* **262**, 90–98 (2020).
18. Cannon, T. D. *et al.* An individualized risk calculator for research in prodromal psychosis. *American Journal of Psychiatry* **173**, 980–988 (2016).
19. Cannon, T. D. *et al.* Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Archives of general psychiatry* **65**, 28–37 (2008).
20. Osborne, K. J. & Mittal, V. A. External validation and extension of the NAPLS-2 and SIPS-RC personalized risk calculators in an independent clinical high-risk sample. *Psychiatry Research* **279**,

- 9–14 (2019).
21. Birmaher, B. *et al.* A risk calculator to predict the individual risk of conversion from subthreshold bipolar symptoms to bipolar disorder I or II in youth. *Journal of the American Academy of Child & Adolescent Psychiatry* **57**, 755–763 (2018).
 22. Doom, J. R., Cicchetti, D., Harkness, K. & Hayden, E. The developmental psychopathology of stress exposure in childhood. *The Oxford handbook of stress and mental health* 265–285 (2018).
 23. Fusar-Poli, P. *et al.* Transdiagnostic risk calculator for the automatic detection of individuals at risk and the prediction of psychosis: second replication in an independent national health service trust. *Schizophrenia Bulletin* **45**, 562–570 (2019).
 24. Wright, A. G. *et al.* The structure of psychopathology: toward an expanded quantitative empirical model. *Journal of abnormal psychology* **122**, 281 (2013).
 25. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC medicine* **11**, 1–8 (2013).
 26. Cuthbert, B. N. The RDoC framework: facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World psychiatry* **13**, 28–35 (2014).
 27. Gur, R. E. *et al.* Burden of Environmental Adversity Associated With Psychopathology, Maturation, and Brain Behavior Parameters in Youths. *JAMA Psychiatry* **76**, 966–975 (2019).
 28. Aldao, A. & Nolen-Hoeksema, S. When are adaptive strategies most predictive of psychopathology? *Journal of abnormal psychology* **121**, 276 (2012).
 29. Pelt, D. H. M. *et al.* Building machine learning prediction models for well-being using predictors from the exposome and genome in a population cohort. *Nature Mental Health* **2**, 1217–1230 (2024).
 30. Lee, T. Y. *et al.* Prediction of psychosis: model development and internal validation of a personalized risk calculator. *Psychological medicine* **52**, 2632–2640 (2022).
 31. Jones, P. B. Adult mental health disorders and their age at onset. *British Journal of Psychiatry* **202**, s5–s10 (2013).
 32. Achenbach, T. M., McConaughy, S., Ivanova, M. & Rescorla, L. Manual for the ASEBA brief problem monitor (BPM). *Burlington, VT: ASEBA* **33**, (2011).
 33. Moos, R. H. & Moos, B. S. Family environment scale. in *Handbook of measurements for marriage and family therapy* 82–86 (Routledge, 2013).
 34. Prinstein, M. J., Boergers, J. & Vernberg, E. M. Overt and Relational Aggression in Adolescents: Social-Psychological Adjustment of Aggressors and Victims. *Journal of Clinical Child & Adolescent Psychology* **30**, 479–491 (2001).
 35. Caspi, A., Houts, R. M., Fisher, H. L., Danese, A. & Moffitt, T. E. The general factor of psychopathology (p): Choosing among competing models and interpreting p. *Clinical Psychological Science* **12**, 53–82 (2024).
 36. Paul, S. E. *et al.* Associations between prenatal cannabis exposure and childhood outcomes: results from the ABCD study. *JAMA psychiatry* **78**, 64–76 (2021).

37. Baranger, D. A. A. *et al.* Prenatal cannabis exposure, the brain, and psychopathology during early adolescence. *Nature Mental Health* **2**, 975–986 (2024).
38. Barch, D. M. & Sheffield, J. M. Cognitive impairments in psychotic disorders: common mechanisms and measurement. *World Psychiatry* **13**, 224–232 (2014).
39. Jirsaraie, R. J. *et al.* Benchmarking the generalizability of brain age models: Challenges posed by scanner variance and prediction bias. *Human Brain Mapping* **44**, 1118–1128 (2023).
40. Paus, T., Keshavan, M. & Giedd, J. N. Why do many psychiatric disorders emerge during adolescence? *Nature reviews. Neuroscience* **9**, 947–957 (2008).
41. Jirsaraie, R. J. *et al.* A systematic review of multimodal brain age studies: Uncovering a divergence between model accuracy and utility. *Patterns* **4**, (2023).
42. Erickson, N. *et al.* Autogluon-tabular: Robust and accurate automl for structured data. *arXiv preprint arXiv:2003.06505* (2020).

Figures

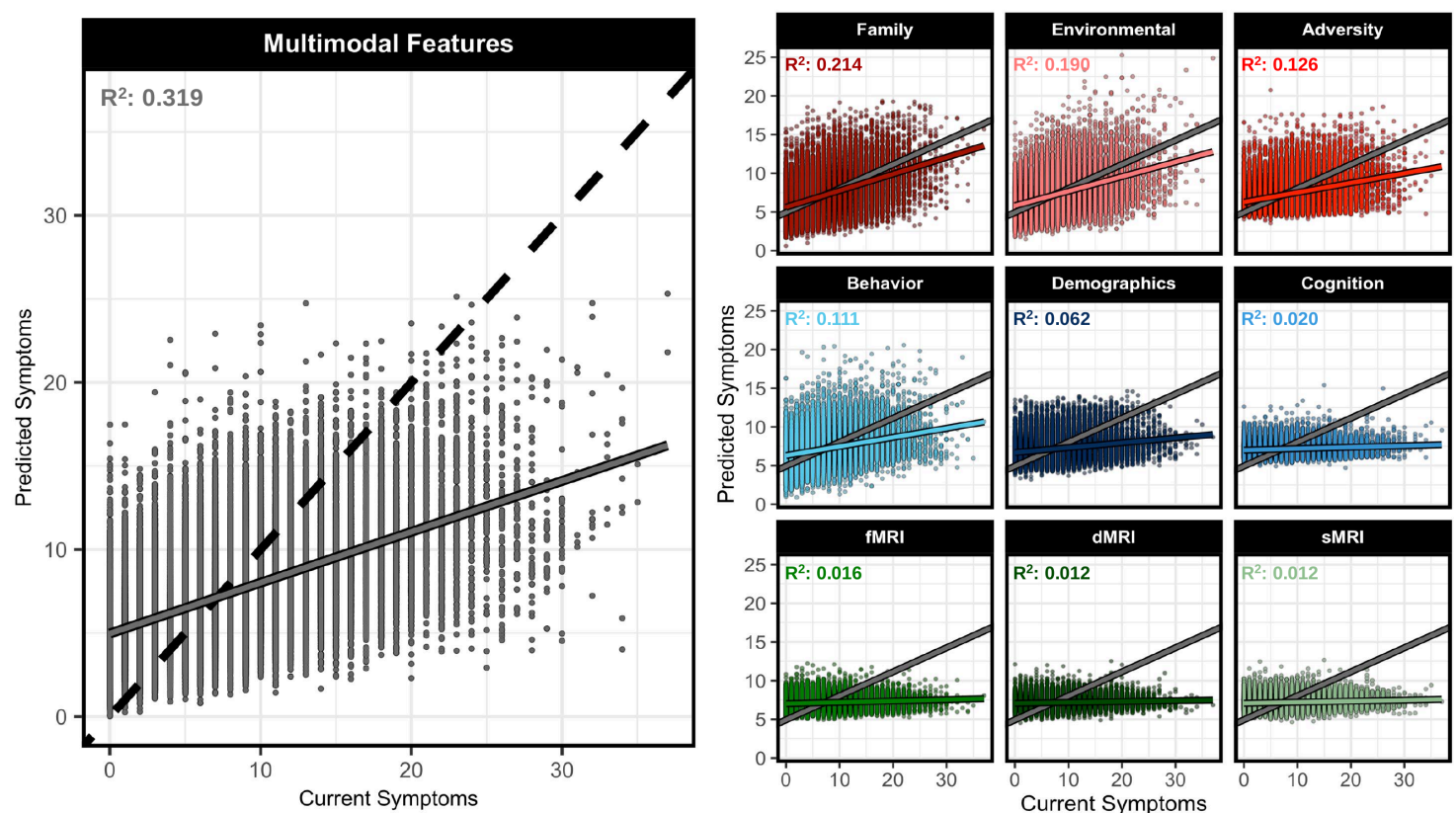


Figure 1

Multimodal predictions of current symptoms accounted for 32% of the variation in general psychopathology, largely driven by information about the family, environment, and adversity. Each scatterplot displays the accuracy between the out-of-sample predictions and current symptoms of general psychopathology. The out-of-sample predictions were aggregated across all five cross validated

models. For perspective, the dashed black line represents a perfect correlation between predictions and symptoms. Multimodal features derived the most accurate prediction as displayed by the solid gray line on the left. Each of the unimodal predictions are color-coded and ordered from most to least accurate on the right. To compare the relative differences between the multimodal and unimodal models, the slope of the multimodal model is also displayed on the right, as shown by the solid gray lines.

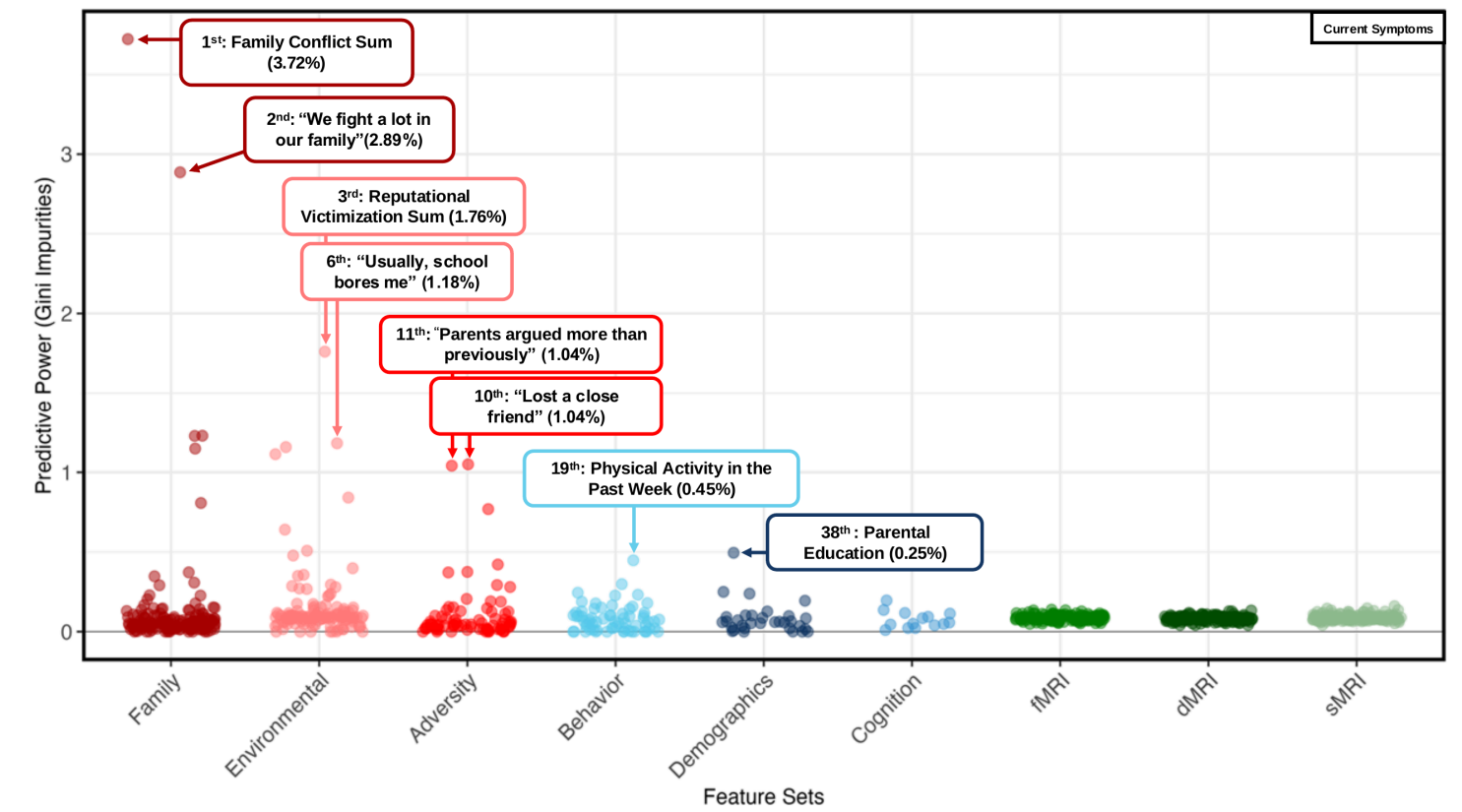


Figure 2

Family fighting was the best predictors of current symptoms in general psychopathology. Models trained on information about the family yielded the most accurate predictions of psychopathology, and these results were reflected by the Gini impurities from our multimodal models. Specifically, the top predictors of current symptoms were related to fighting, arguing, and verbal abuse between family members. Similarly, the second set of key predictors were related to the social environment, including victimization among peers.

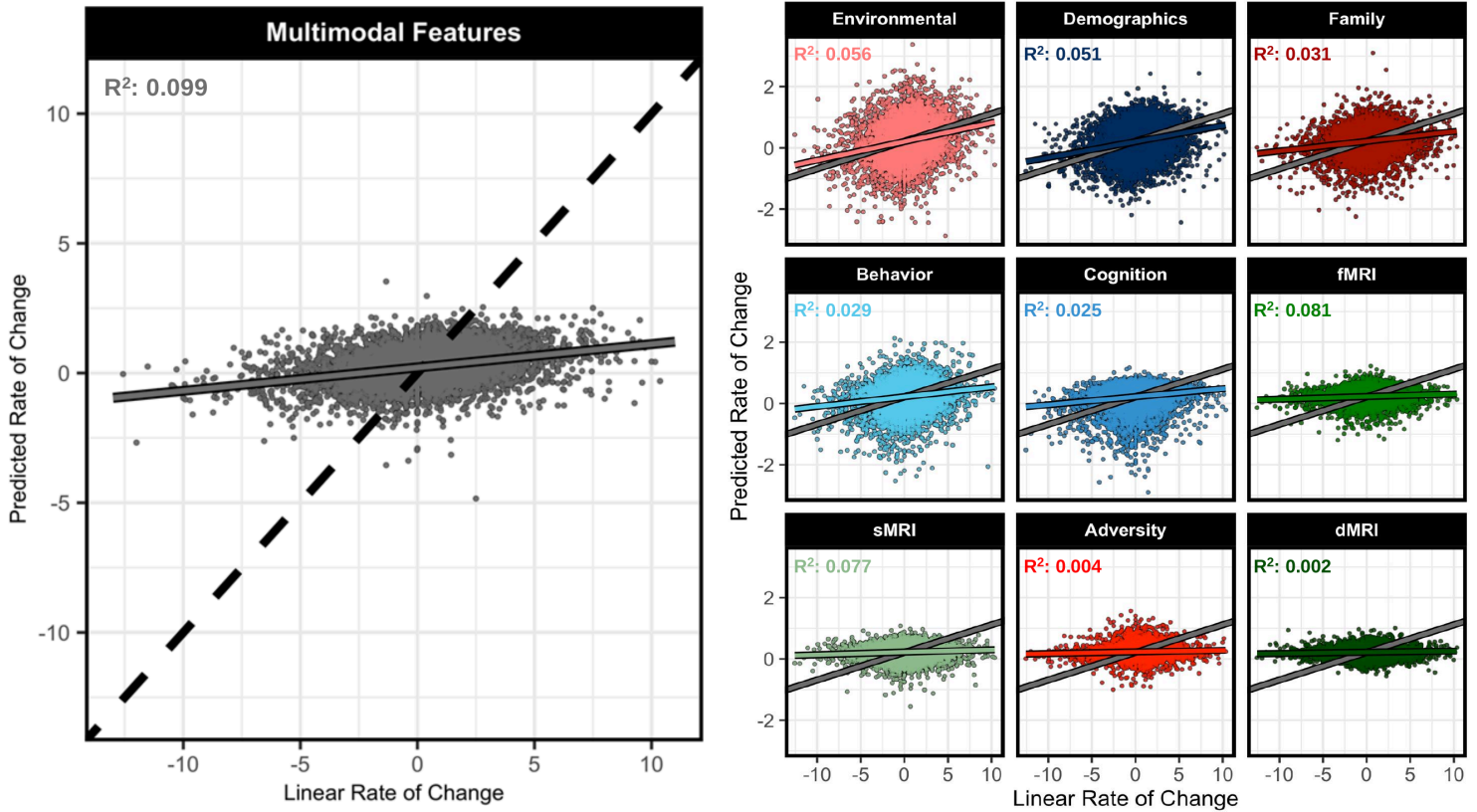


Figure 3

Multimodal predictions of the rate of change in general psychopathology explained up to 10% of the individual differences, with the most accurate features pertaining to environment, demographics, and family. Each scatterplot displays the accuracy between the out-of-sample predictions and changes in general psychopathology. Multimodal features derived the most accurate prediction as displayed by the solid gray line on the left. For perspective, the dashed black line represents a perfect correlation between predictions and symptoms. Each of the unimodal predictions are color-coded and ordered from most to least accurate on the right. To compare the relative differences between the multimodal and unimodal models, the slope of the multimodal model is also displayed on the right, as shown by the solid gray lines.

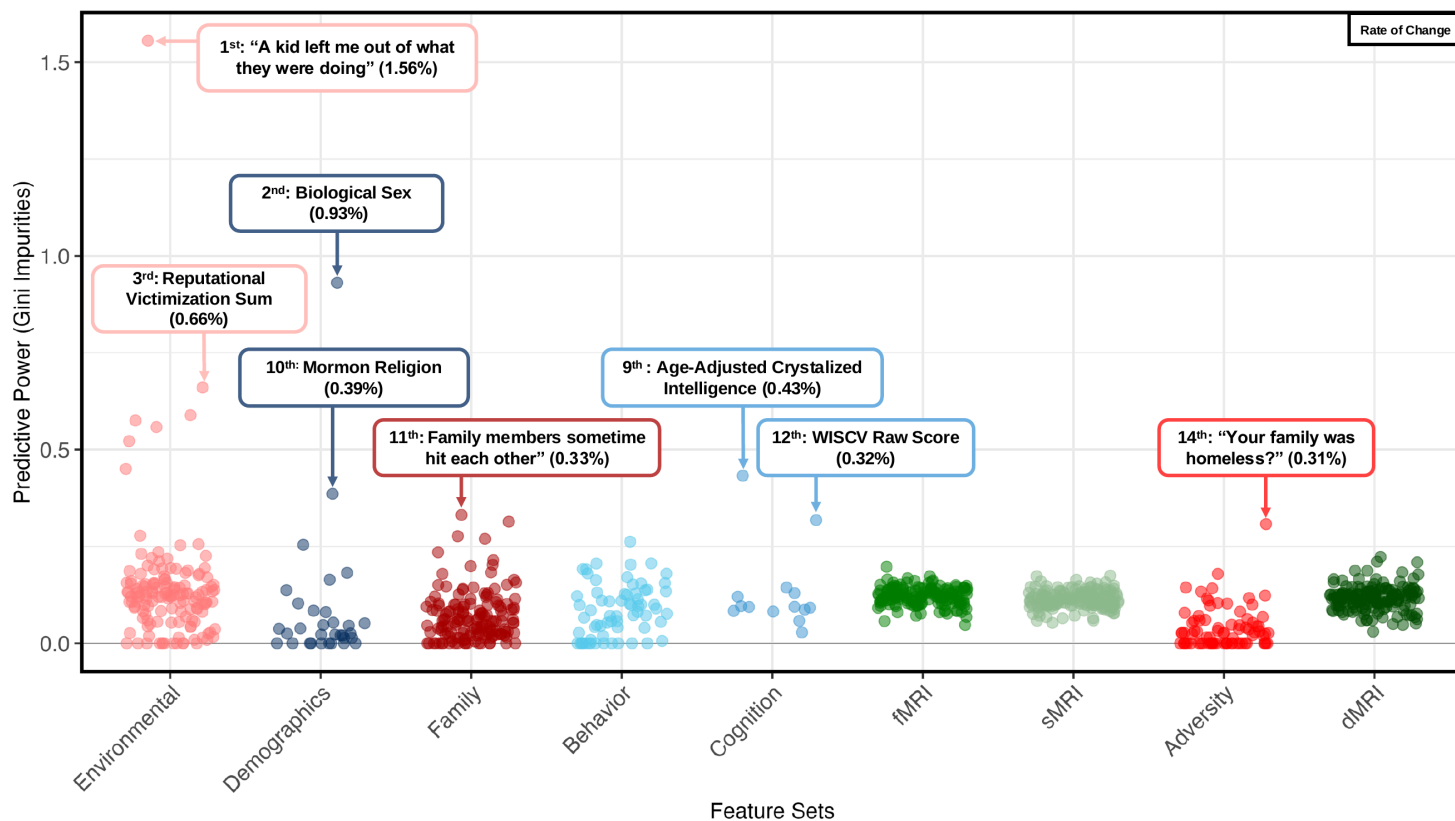


Figure 4

Social exclusion and reputational damage were significant predictors of changes in symptom severity over time, with apparent of sex-differences. Environmental features contained the most predictive power of changes in symptoms, which were largely driven by social conflicts between peers, especially being excluded from activities or experiencing a degradation of one's social image/standing. Biological sex was the only other key predictor that emerged, indicating a potential divergence between males and females.

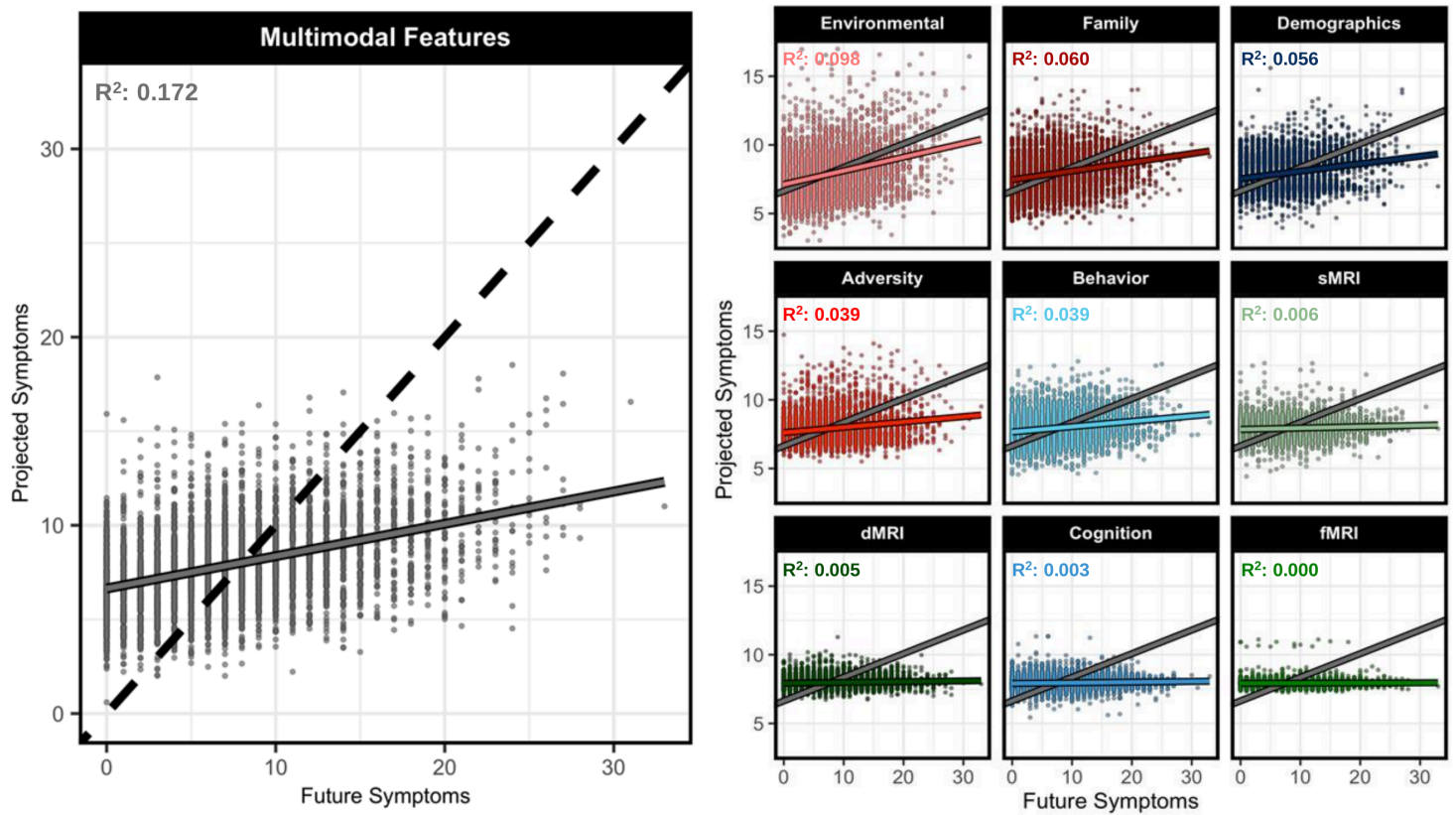


Figure 5

Multimodal predictions of future symptom severity captured as much as 17% of inter-individual variation, subsets of features that were the most important included environment, family, and demographics. Each scatterplot displays the accuracy between the out-of-sample predictions and future symptoms of general psychopathology. Multimodal features derived the most accurate prediction as displayed by the solid gray line on the left. For perspective, the dashed black line represents a perfect correlation between predictions and symptoms. Each of the unimodal predictions are color-coded and ordered from most to least accurate on the right. To compare the relative differences between the multimodal and unimodal models, the slope of the multimodal model is also displayed on the right, as shown by the solid gray lines.

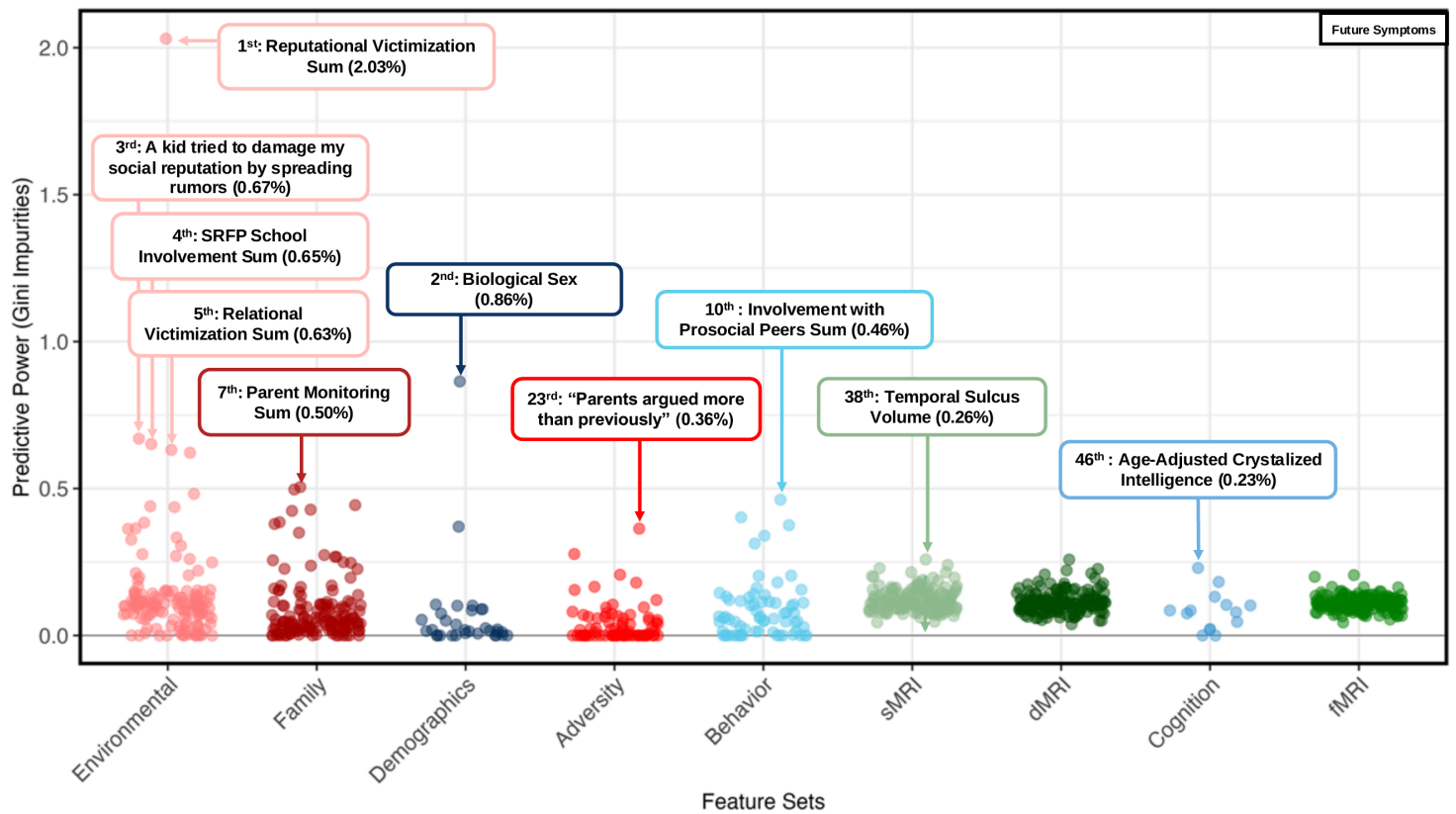


Figure 6

Reputational victimization and biological sex provided the most predictive utility of future symptom severity. Converging with our results predicting longitudinal changes, social conflicts between peers re-emerged as the top predictors, which included novel items about experiencing unwanted rumors and gossiping. Similar results were also found in terms of biological sex remaining as the second most important predictor, further indicating a potential divergence between males and females.

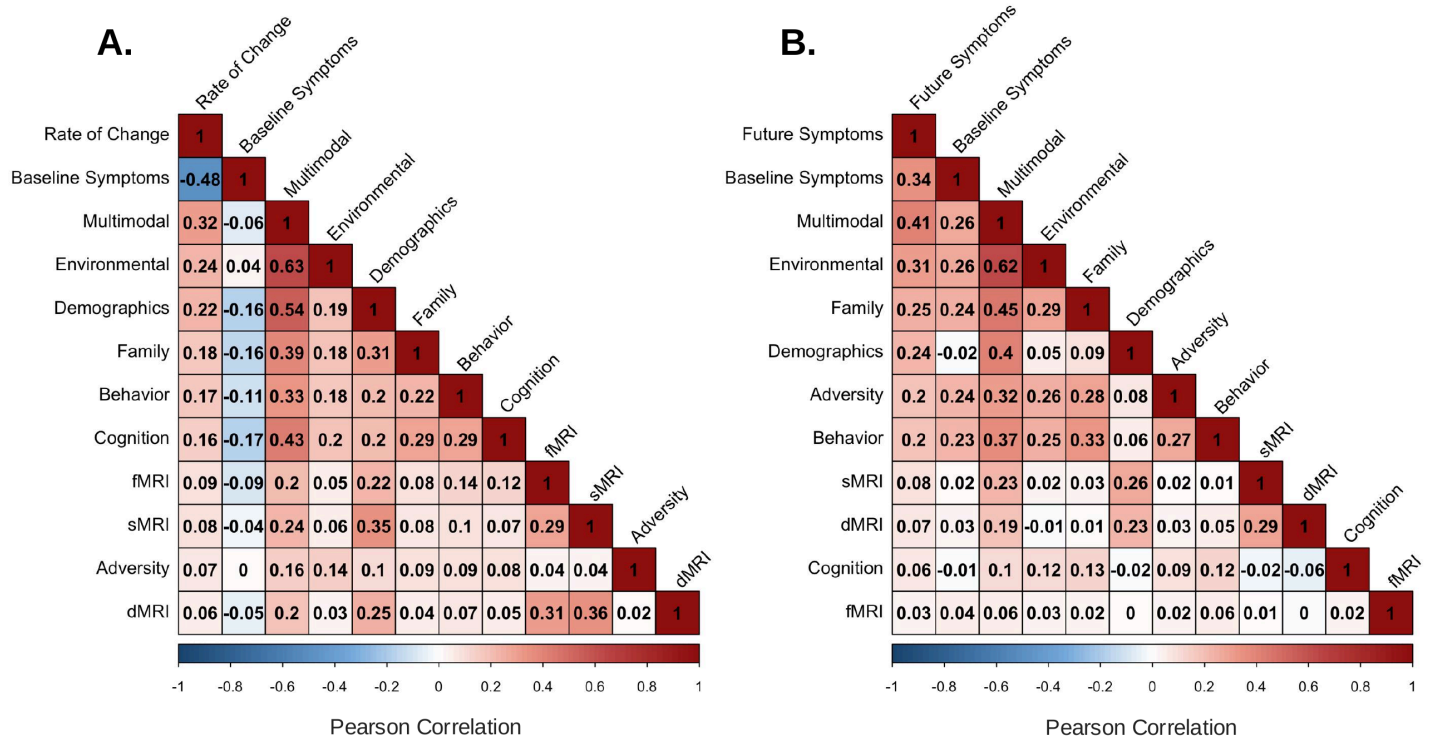


Figure 7

Pairwise-correlations between model predictions, baseline symptoms, and longitudinal outcomes. To better understand the accuracy of our longitudinal models, we compared their performance relative to simply using baseline symptoms. **A.** Changes in symptoms were negatively correlated with baseline symptoms ($r = -0.48$), which was stronger than any set of predictions we generated ($r < 0.32$). **B.** Future symptom severity was most strongly correlated with the multimodal predictions ($r = 0.41$), even more so than using baseline symptoms ($r = 0.34$). Predictions between each model were weakly to moderately correlated ($r < 0.63$), indicating that each model represents a unique source of variation. Combining predictions from models and baseline symptoms could improve estimates of longitudinal outcomes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalMaterial.docx](#)