Using gastrointestinal distress reports to predict youth anxiety risk: Implications for mental health literacy and community care

Paul Alexander Bloom¹ | Ian Douglas² | Michelle VanTieghem¹ | Nim Tottenham¹ | Bridget Callaghan³

¹ Psychology Department, Columbia University, New York, New York, USA
² Department of Psychology, University of Texas at Austin, Austin, Texas, USA
³ Psychology Department, University of California Los Angeles, Los Angeles, California, USA

Correspondence
Paul Alexander Bloom, Psychology Department, Columbia University
Schermerhorn Hall 409A, 1180 Amsterdam Ave, New York, NY 10029, USA.
Email: paul.bloom@columbia.edu

Funding information
National Science Foundation, Grant/Award Number: Graduate Research Fellowship DGE 1644869; National Institute of Mental Health, Grant/Award Numbers: 4R00MH113821, R01MH091864; Brain and Behavior Research Foundation, Grant/Award Number: 24739

Abstract
This study investigates the generalizability and predictive validity of associations between gastrointestinal (GI) symptoms and youth anxiety to establish their utility in community mental health decision-making. We analyzed data from youth ages 3 to 21 years in volunteer cohorts collected in Los Angeles (N = 327) and New York City (N = 102), as well as the Healthy Brain Network cohort (N = 1957). Youth GI distress was measured through items taken from the parent-reported Child Behavior Checklist (CBCL). We examined generalizability of GI–anxiety associations across cohorts and anxiety reporters, then evaluated the performance of these models in predicting youth anxiety in holdout data. Consistent with previous work, higher levels of gastrointestinal distress were associated with more parent-reported youth anxiety behaviors in all three cohorts. Models trained on data from the Healthy Brain Network cohort predicted parent-reported and child-reported anxiety behaviors, as well as clinician-evaluated anxiety diagnoses, at above chance levels in holdout data. Models which included GI symptoms often, but not always, outperformed models based on age and sex alone in predicting youth anxiety. Based on the generalizability and predictive validity of GI–anxiety associations investigated here, GI symptoms may be an effective tool for child-facing professionals for identifying children at risk for anxiety (Preprint: https://psyarxiv.com/zgavu/).

KEYWORDS
anxiety, gastrointestinal, prediction, replication, youth

1 | INTRODUCTION

Psychiatric distress is increasingly common, especially among youth (Sakolsky & Birmaher, 2008; Twenge et al., 2019; Yatham et al., 2018). For example, the median age of onset for anxiety disorders is 11 years (Kessler et al., 2005), and early onset mental health problems are associated with more chronic courses of illness (Beesdo-Baum & Knappe, 2012; Godoy et al., 2019). As such, early intervention for mental health issues is a public health concern, prompting several professional bodies to prioritize recognition, referral, and treatment competencies in the domain of youth mental health amongst their members (Bell & McKay, 2004).

Mental health care for youth is dispersed across multiple systems, including family members, primary care providers (PCPs), schools (Luthar, Ebbert, et al., 2020; Luthar, Kumar, et al., 2020; Moon et al., 2017), juvenile justice (Duong et al., 2020; Loyd et al., 2019; Yoder et al., 2017), and child welfare (Gudiño et al., 2012; Shin, 2005). As PCPs are often the initial access point for mental health services,
particularly within economically disadvantaged communities, primary care is typically the first line of defense for managing young people’s mental health (Heneghan et al., 2008; O’Brien et al., 2016). Over 25% of presentations to pediatric primary care are for psychosocial distress (Cooper et al., 2006), and more presentations occur for unexplained physical and somatic complaints that might be psychogenic (Rushton et al., 2002). Not surprisingly, universal mental health screening has been recommended for PCPs, as well as teachers, school nurses, and other child-facing professionals (Dowdy et al., 2015; Dworsky et al., 2014). Nonetheless, uptake of such screening is varied across contexts and is far from universally adopted, even within pediatric primary care. Common concerns from teachers and primary care pediatrics include a lack of training and confidence in providing treatments or making psychiatric referrals based on observed or reported psychological states (Green et al., 2019; Reinke et al., 2011). Moreover, as many children may lack the knowledge/insight to effectively report their mental health symptoms (Muir et al., 2010), screening which relies only on self-reported psychological states may miss youth who present with primarily physical manifestations of distress. Given such constraints, tools that go beyond direct observation and understanding of psychological states are needed to enhance the detection of youth psychosocial distress.

One strategy to address limitations in professional expertise and child self-reports of mental states is to use comorbid or prodromal physical health symptoms in treatment and referral decisions. Mental health disorders are highly comorbid with somatic symptoms (Bodas et al., 2008; Cohen et al., 1998; Koloski et al., 2012; Waters et al., 2013), and gastrointestinal (GI) complaints (e.g., unexplained abdominal pain) in particular are associated with concurrent and future anxiety in youth (Callaghan et al., 2020; Campo et al., 2004; Kim et al., 2019; Saps et al., 2009) and are common (8%–10%) in primary health settings (Starfield et al., 1980). GI distress symptoms may thus be helpful in psychotherapeutic referral processes for youth. Nonetheless, in primary care settings, psychiatric referrals occur in less than 20% of GI distress cases even when a psychogenic root is suspected (Rushton et al., 2002), suggesting there is room to improve the use of physical symptoms for psychiatric referral. Considering that GI symptoms are common and might be more easily reported by youth, as well as observable to parents/PCPs/teachers/other child-facing professionals, routine screening of GI symptoms within primary care and community settings may assist individuals in making accurate treatment or referral decisions for youth for mental health services, bolstering early intervention and prevention efforts.

Previously (Callaghan et al., 2020), we demonstrated an association between parent-reported child GI distress and child anxiety using five questions assessing GI symptoms taken from the Child Behavior Checklist (CBCL; Achenbach & Edelbrock, 1983). In the current study, rather than investigating causal mechanisms behind this relationship, we seek to generalize GI–anxiety associations across samples and anxiety outcomes and evaluate their potential for clinical use. If responses on those items could be used in a predictive framework to identify youth at risk for anxiety, this would highlight an opportunity for primary care and community settings to incorporate reports of GI symptoms into mental health screening and referral decisions. Beyond the knowledge of correlations between mental illness and GI distress (which may already be used in many settings), predictive validity of these associations may enable more efficient and accurate use of this knowledge for decision-making in individual cases. In the present study, we work to establish the generalizability and predictive validity of such associations prerequisite for their use in serving youth mental health needs.

2 METHODS

2.1 Participants

We analyzed data from three cohorts in this study (see Tables 1–3). Two of these cohorts were collected by our lab in two cities: Los Angeles (LA, N = 327) and New York City (NYC, N = 102). The third was a publicly available dataset based on New York City: Healthy Brain Network (HBN, N = 1957; Alexander et al., 2017; Das et al., 2012). Each cohort included children and adolescents between ages 3 and 21 years. In the LA and NYC cohorts, some youth had experienced caregiving adversity (institutional or foster care). In the HBN cohort, recruitment strategies (community-referred recruitment to increase participation of families who were concerned about the psychiatric health of their children) resulted in a higher proportion of participants meeting clinical diagnoses for mental illness than expected in the general population.

Associations between GI symptoms and anxiety were previously published in the LA cohort (Callaghan et al., 2020). Here, we analyzed data from the LA and NYC cohorts before preregistering replication and predictive modeling analyses for the HBN cohort (https://osf.io/687ky). We completed most preregistered analyses as planned and acknowledge several differences in the supplement. Within the HBN dataset, we analyzed both complete-case data (reported in analyses in the main manuscript) and all data using imputation strategies (supplemental) in parallel. Before conducting analyses, we split the HBN dataset into a training set (75% of the participants) and a holdout set (75% of the participants). Henceforth, we refer to “HBN training data” to mean the training set, “HBN holdout data” to mean the holdout set, and “HBN full dataset” to mean all of the available HBN data (both training and holdout).

2.2 Child GI symptoms

In all cohorts, we assessed child GI symptomatology using parent reports from the CBCL for ages 1.5–5 and 4–18 years (Achenbach, 1991). We included four items present in both versions so that symptoms could be compared across ages. Using a three-point scale (0 = not true, 1 = somewhat true, 2 = very true) to indicate the degree to which their children were experiencing “physical problems without known medical cause,” parents reported incidence of nausea, stomachaches/cramps, vomiting, and constipation (Supporting Information Methods). The internal consistency of these four items, as measured by Cronbach’s alpha, was 0.63 (95% CI [0.61, 0.66]) in the full HBN
TABLE 1  Sample information on sex, age, GI symptoms, anxiety symptoms, and median annual household income for all cohorts

<table>
<thead>
<tr>
<th>Row</th>
<th>HBN</th>
<th>LA</th>
<th>NYC</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1957</td>
<td>327</td>
<td>102</td>
</tr>
<tr>
<td>Sex (M/F)</td>
<td>1238/719</td>
<td>139/188</td>
<td>44/58</td>
</tr>
<tr>
<td>Mean age in years (SD, [range])</td>
<td>10.5 (3.7, [5, 21.9])</td>
<td>8.4 (4.1, [2.8, 18.6])</td>
<td>10.8 (3.5, [4.4, 17.6])</td>
</tr>
<tr>
<td>Mean GI sum score (SD, [range])</td>
<td>0.866 (1.31, [0, 8])</td>
<td>0.361 (0.77, [0, 4])</td>
<td>0.363 (0.77, [0, 4])</td>
</tr>
<tr>
<td>Nausea</td>
<td>0.83/0.15/0.03</td>
<td>0.94/0.06/0.01</td>
<td>0.93/0.07/0</td>
</tr>
<tr>
<td>Vomiting</td>
<td>0.93/0.06/0.01</td>
<td>0.99/0.01/0</td>
<td>0.99/0.01/0</td>
</tr>
<tr>
<td>Stomachache/cramps</td>
<td>0.71/0.25/0.04</td>
<td>0.85/0.13/0.02</td>
<td>0.85/0.15/0</td>
</tr>
<tr>
<td>Constipated</td>
<td>0.8/0.15/0.05</td>
<td>0.9/0.09/0.01</td>
<td>0.89/0.08/0.03</td>
</tr>
<tr>
<td>Mean SCARED-P (SD, [range])</td>
<td>14.73 (12.2, [0, 72])</td>
<td>13.71 (11.08, [0, 67])</td>
<td>12.93 (10.98, [0, 56])</td>
</tr>
<tr>
<td>Proportion meeting SCARED-P cutoff</td>
<td>0.19</td>
<td>0.15</td>
<td>0.13</td>
</tr>
<tr>
<td>Mean SCARED-C (SD, [range])</td>
<td>23.21 (16.13, [0, 80])</td>
<td>23.21 (16.13, [0, 80])</td>
<td>23.21 (16.13, [0, 80])</td>
</tr>
<tr>
<td>Proportion meeting SCARED-C cutoff</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Proportion w/KSADS anxiety diagnosis</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Median annual household income</td>
<td>$90,000–$99,999</td>
<td>$100,001–$150,000</td>
<td>$150,001–$200,000</td>
</tr>
</tbody>
</table>

Abbreviations: GI, gastrointestinal; HBN, Healthy Brain Network; KSADS, Kiddie Schedule for Affective Disorders and Schizophrenia; LA, Los Angeles; NYC, New York City; SCARED, Screen for Child Anxiety Related Disorders.

cohort, 0.45 (95% CI [0.37, 0.54]) in the LA cohort, and 0.47 (95% CI [0.33, 0.62]) in the NYC cohort. We used sum scores of these four items to calculate each participant’s GI symptom score. While previous work (Callaghan et al., 2020) within the LA cohort included one item from the Revised Child Anxiety and Depression Scale (RCADS; Chorpita et al., 2000) within total GI scores, RCADS data were not available for the HBN and NYC cohorts, so we only used CBCL items for the present study.

2.3  Child anxiety symptoms

2.3.1  Parent report

We measured anxiety-related behaviors in all cohorts using the Screen for Child Anxiety Related Disorders–Parent version (SCARED-P; Birmaher et al., 1997). Parents reported frequencies of each of 41 behaviors on this measure using a three-point scale (0 = never, 1 = sometimes, 2 = often). We removed item #11 (“my child gets stomachaches at school”) to avoid including explicit GI symptoms in our anxiety measure. Scores on this measure were summed (see Table 1). Cronbach’s alpha for the SCARED-P was 0.93 (95% CI [0.92, 0.94]) for the full HBN cohort, 0.92 (95% CI [0.91, 0.94]) for the LA cohort, and 0.92 (95% CI [0.90, 0.94]) for the NYC cohort.

2.3.2  Child report

In the HBN cohort only, youth reported on their own anxiety-related behaviors using a child version of the SCARED scale (SCARED-C; Birmaher et al., 1997). We summed anxiety symptom scores on SCARED-C using the same process as with the parent-reported version. Cronbach’s alpha for the SCARED-C was 0.94 (95% CI [0.94, 0.95]).

2.3.3  Consensus clinical diagnoses

In the HBN cohort, youth completed a computerized web-based version of the Kiddie Schedule for Affective Disorders and Schizophrenia–Children’s version (KSADS; Kaufman et al., 1997), and both youth and parents completed semi-structured DSM-5-based psychiatric interviews with a licensed clinician (KSADS-COMP). A clinical team reviewed these materials, as well as relevant materials collected during study participation to arrive at a consensus for diagnoses of DSM-5 disorders. We created a binary variable indicating if each participant was given any one of 13 anxiety-related diagnoses (Supporting Information Methods).

2.4  Replication and robustness

To replicate associations between GI symptoms and anxiety previously reported (Callaghan et al., 2020), in each cohort (LA, NYC, and HBN training data) we regressed continuous SCARED-P (parent-reported) scores on GI symptoms, sex, age, and interactions of GI symptoms and sex using the rstanarm package in R (Goodrich et al., 2018). To test the robustness of such associations to different reporters of anxiety symptoms in the HBN training data, we fit a similar linear regression model to continuous SCARED-C (child reported) scores, and a logistic regression model to determine whether participants were diagnosed through clinician consensus using the KSADS.
2.5 Predictive utility

**Step 1: Model selection using the HBN training set:** For each anxiety outcome, we selected the best-performing models across several regression model formulations for validation in our holdout sets (see Figure 1 and Step 2). The four model formulations included (1) a summed total GI score plus age and sex regressors, (2) each GI item as a separate regressor, plus age and sex regressors, (3) no GI terms with only age and sex regressors, and (4) an intercept-only model (i.e., predicting the group average level of anxiety for each person, included as a baseline comparison). For each formulation including a GI term (1–2), models contained terms for age and sex, and we also examined whether including interaction terms with age and sex (i.e., GI*age, GI*sex, GI*age*sex) would improve model performance (see Supporting Information Methods). We also used models from formulation 2 to examine which GI items were most predictive of anxiety (Figures S18–S22).

From each of the four model formulations, we selected the best-performing model from cross-validation in the training set, then fit the best-performing model variants to the entire training set (Table S5) before validation in the holdout sets (both the HBN holdout set and the combined LA + NYC data).

**Step 2: Model validation with holdout data:** For each selected linear and logistic regression model from Step 1, we examined out-of-sample predictive performance on the HBN holdout set, as well as the combined LA and NYC cohorts (for SCARED-P). Performance was assessed in Step 2 through bootstrapping and permutation testing with $q^2$ (proportion of variance predicted, see Supporting Information Methods), log loss, and receiver-operating characteristic curves (ROC, and corresponding area under the curve [AUC]) metrics. Because we made six comparisons with each set of models (all four selected model formulations against the permuted null, plus both GI model formulations against the no GI model), we used a Bonferroni adjusted alpha level of 0.0083 (.05/6) for each permutation test.

2.6 Exploratory analyses: GI symptoms as a predictor for broader psychopathology

We also asked whether GI symptoms could function as a predictor of general psychopathology risk within the HBN full dataset. We constructed logistic regression models using summed GI symptoms as well as age and sex covariates to explore associations between GI symptoms and the likelihood of any diagnosis, as well as each respective KSADS DSM-5 diagnosis that occurred in more than 5% of participants. In a separate set of models with only GI sum term (and without age and sex predictors), we examined predictive performance through repeated cross-validation (Figure S30).
**Stage 1: Model Selection**

<table>
<thead>
<tr>
<th>Training Data</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td></td>
</tr>
<tr>
<td>HBN Train</td>
<td></td>
</tr>
<tr>
<td>Parent-Proxy Report (SCARED-P)</td>
<td>Continuous</td>
</tr>
<tr>
<td>Clinician Consensus (KASAS)</td>
<td>Continuous</td>
</tr>
</tbody>
</table>

**Model Formulations**

1. GI Summed
   - Continuous
   - Clinical Cutoff

2. GI Individual Items
   - Continuous
   - Clinical Cutoff

3. No-GI
   - No interactions
   - No vars

4. Intercept Only
   - No interactions
   - No vars

**Model Variants**

- No-GI: better than a model with no-GI symptoms included
- Baseline: better than baseline threshold formulation, e.g., AUC = 0.5
- Null model: better than chance performance based on a permuted null distribution

One model variant was selected for each model formulation.

Selected models move to Stage 2.

**Stage 2: Model Validation**

<table>
<thead>
<tr>
<th>Holdout Data</th>
<th>Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td></td>
</tr>
<tr>
<td>HBN Test</td>
<td></td>
</tr>
<tr>
<td>LA + NYC</td>
<td></td>
</tr>
</tbody>
</table>

**HBH Holdout**

- Child Self Report (SCARED-C) No model formulations
- Parent-Proxy Report (SCARED-P) Model formulations 1.6.2
- Clinician Consensus (KASAS) No model formulations 1.2

**LA + NYC Holdout**

- Parent Proxy (SCARED-P) Model formulations 1.6.2
- Parent Proxy (SCARED-P) Categorical: No model formulations

Boxes surrounded by a broken border were included only in the supplemental data section and are not represented in the charts to the right. As they were included in the preregistered analytical plan, they are included in this visualization for completeness.

4 models from each outcome measure (i.e., 20 models) were moved forward to stage 2 model validation.

---

**FIGURE 1** Flowchart summarizing our procedures and results for measuring the predictive validity of models predicting anxiety from gastrointestinal (GI) symptoms. **Step 1:** Model selection includes the models that were built on the HBN training dataset. There were four different model formulations built on the outcome measures: (1) GI summed (including age and sex), (2) GI individual items (including age and sex), (3) no GI (just age and sex), (4) intercept only. Within those model formulations, several model variants were tested which included models with no interactions between predictors, as well as models that interacted with GI symptoms, age, sex, or included three-way interactions with GI symptoms age and sex. One model variant was selected for each model formulation and carried through to Step 2. **Step 2:** Model validation occurred on the holdout data (Healthy Brain Network (HBN) test dataset, as well as the combined Los Angeles (LA) + New York City (NYC) cohorts). The LA and NYC cohorts were combined into a single holdout dataset due to highly similar demographics and to increase the sample size. Some of the models (i.e., those built on the clinical cutoff scores from SCARED-C and SCARED-P within the HBN training data—see boxes surrounded by broken lines) were tested within the HBN holdout data but were not discussed in the main manuscript and are not represented in the colored boxes to the bottom right of the graph. Colored boxes in the bottom right of the figure represent the relative level of evidence for the predictive validity of each model for each of the validation tests (> No-GI: better than a model with no GI symptoms included; > Baseline: better than baseline threshold formulation, e.g., AUC = 0.5; > Null model: better than chance performance based on a permuted null distribution). The level of evidence represented is relative, rather than absolute, meaning that relative to the other models, those highlighted in green provide the strongest evidence for predictive performance of GI models.
RESULTS

3.1 Replicability and robustness

Equivalent regression models fit to all three cohorts were consistent such that on average, higher total GI symptoms were associated with more parent-reported child anxiety symptoms ($\hat{\beta}_{\text{HBN}} = 3.88$, 95% CI [3.37, 4.38], $\hat{\beta}_{\text{LA}} = 5.02$, 95% CI [3.46, 6.52], $\hat{\beta}_{\text{NYC}} = 5.49$, 95% CI [2.85, 8.10]) as per the SCARED-P (Figure 2, Table S3). In addition, we found that associations between GI symptoms and anxiety were robust to responder (Figure 3), such that within the HBN training dataset higher GI symptoms were on average associated with higher anxiety behaviors on SCARED-C ($\hat{\beta} = 1.79$, 95% CI [1.07, 2.51]) and higher probability of diagnosis on the KSADS ($\hat{\beta} = 0.28$, 95% CI [0.19, 0.37]).

3.2 Predictive validity

Step 1: Model selection: Within each of the two model formulations that included GI terms, inclusion of age/sex interaction terms had little impact on model performance. Through cross-validation with the HBN training data, we found that both the best-performing GI summed and GI individual items models predicted anxiety scores on the SCARED-P better than the best-performing No-GI model or intercept-only model (Figure 4). There were not consistent performance differences between the best GI summed model and best GI individual items model. While GI models also tended to perform better than No-GI models during cross-validation with the SCARED-C and KSADS outcomes, they did not do so as consistently (Figures S15–S17).

Step 2: Model validation with holdout data: For parent-reported anxiety (as per the SCARED-P) in the HBN holdout data, GI summed ($q^2_{\text{median}} = 0.139$, 95% CI [0.036, 0.23]) and GI individual items ($q^2_{\text{median}} = 0.144$, 95% CI [0.034, 0.242]) models both predicted variance over baseline and outperformed the No-GI model ($q^2_{\text{median}} = 0.005$, 95% CI [−0.011, 0.018]) in predicting variance in anxiety ($p_{\text{summed}} = 0.0045$, $p_{\text{indiv. items}} = 0.0049$; Figure 5A). For child-reported anxiety, GI summed ($q^2_{\text{median}} = 0.073$, 95% CI [0.005, 0.135]) and GI individual items ($q^2_{\text{median}} = 0.071$, 95% CI [0.004, 0.138]), but not No-GI model formulations ($q^2_{\text{median}} = 0.045$, 95% CI [−0.005, 0.091]) reliably predicted variance in anxiety symptoms over a baseline $q^2$ of 0, though neither GI model formulation consistently outperformed the No-GI model ($p_{\text{summed}} = 0.0904$, $p_{\text{indiv. items}} = 0.1467$).

To assess whether the performance of any models was above chance, anxiety outcomes in the HBN holdout dataset were shuffled 10,000 times to calculate empirical null distributions and compared to the actual performance of each model. For parent-reported anxiety, both the GI sum model and the GI individual item model performed better than the null ($p_{\text{summed}} = 0.0001$, $p_{\text{indiv. items}} = 0.0001$), whereas the no-GI model did not ($p_{\text{no-GI}} = 0.0351$, Figure 5A), suggesting that only models including a GI term could predict parent-reported anxiety at above chance levels. For child-reported anxiety, the GI summed,
FIGURE 3  Robustness of associations between parent reports of child gastrointestinal (GI) symptoms and both child-reported and clinician-evaluated anxiety in the Healthy Brain Network (HBN) training set. All panels show model fitted predictions (solid lines) for an average participant of age 10.5 years with a 95% uncertainty interval (semi-transparent overlay). (a) Linear regression with parent-reported anxiety behaviors on Screen for Child Anxiety Related Disorders–Parent (SCARED-P) (orange) and child-reported anxiety behaviors on SCARED-C (green) as the outcomes. Points show individual participant scores on each measure. (b) Logistic regression with clinician-consensus anxiety diagnosis based on the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) as the outcome. Points indicate proportions of participants at each possible GI score with diagnoses, and point size indicates the number of participants in a given bin.

3.3 Exploratory analyses: GI symptoms as a predictor for broader psychopathology

While covarying for age and sex, GI symptoms were positively associated in the full HBN dataset with several mental health diagnoses, notably in the categories of anxiety, behavioral, elimination, and mood disorders (Figure 6A). Further, GI symptoms were negatively associated with the likelihood of participants having no mental health diagnoses. Additional cross-validation of logistic regression models with summed GI symptoms as the sole predictor (no age or sex predictors) revealed evidence that several diagnoses could be classified above a baseline AUC of 0.5 (Figure 6B).

4 DISCUSSION

Here we establish generalizability and predictive performance of GI–anxiety associations. We first attempted to replicate associations between GI symptoms and anxiety in several cohorts using different informants (parent, child, clinician) to demonstrate the generalizability of these associations. We then assessed predictive accuracy of GI symptom models in holdout datasets. Across three cohorts and with different informants, we found that GI distress was positively associated with youth anxiety, demonstrating a replicable and robust relationship. These findings are consistent with prior work from other groups within both adult and youth cohorts without enrichment for...
FIGURE 4  Cross-validation of linear regression models predicting Screen for Child Anxiety Related Disorders–Parent (SCARED-P) from gastrointestinal (GI) symptoms within the Healthy Brain Network (HBN) training data. (a) For each model formulation (GI summed, GI individual items, No-GI Term, and Intercept Only, represented by the different boxes) the median $q^2$, 80% uncertainty interval (thick line), and 95% uncertainty interval (thin line) from all cross-validated folds are plotted for each model variant. (b–d) $q^2$, calculated for the model variants selected in Step 1 across all folds (each data point represents a fold). Data points that fall above the line indicate a higher $q^2$, for the model represented on the y-axis, whereas those that fall below the solid line indicate higher $q^2$, from the model represented on the x-axis. Both the GI summed model (b) and the GI individual items model (c) outperformed the No-GI model, and there were no consistent differences in performance between the two types of GI models (individual items and summed; d).

early life adversity (Haug et al., 2002; Jones et al., 2017; Koloski et al., 2012; Waters et al., 2013). Models trained on one cohort (HBN training set) were also successful in predicting anxiety outcomes above chance levels in two holdout datasets (HBN holdout set and NYC + LA holdout set), although predictive performance was quite variable. Specifically, in the HBN holdout data, models predicted anxiety at above chance levels for all anxiety informants. Further, GI models outperformed equivalent models without GI terms for parent-reported child anxiety. However, models were less successful in predicting child-reported anxiety or clinician-consensus diagnoses beyond what could already be predicted based on age and sex. Similar findings emerged when the combined LA/NYC cohorts were used to test model performance. Together, these findings establish both generalizability and predictive validity of associations between GI symptoms and anxiety.

While most of the present work focused on anxiety, exploratory analyses indicated that GI symptoms were also associated with broader psychopathology within the full HBN cohort. Models with a sole GI term predicted a range of clinician-evaluated psychopathologies above a baseline AUC = 0.5, specifically generalized anxiety disorder, separation anxiety, specific phobias, major depressive disorder, dysthymia, oppositional defiant disorder, and encopresis. While these findings are preliminary and need to be confirmed in separate cohorts, links between GI symptoms and broader psychopathology, mood disorders in particular, are consistent with prior work (Bagayogo et al., 2013; Campo et al., 2004; Haug et al., 2002; Kennis et al., 2019; Okulate et al., 2004). While such findings indicate that these models may lack specificity in their ability to predict anxiety symptoms, they also suggest that somatic symptoms may be useful tools for transdiagnostic referrals or
FIGURE 5  Bootstrapped model predictive performance and null permuted distributions on Healthy Brain Network (HBN) holdout data for parent-report Screen for Child Anxiety Related Disorders–Parent (SCARED-P, panel a), child-report (SCARED-C, panel a), and clinician consensus Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) anxiety diagnoses (panels b–d). Bootstrapped distributions of model performance are shown in color (purple = GI summed (model formulation 1), blue = GI individual items (model formulation 2), red = No-GI (model formulation 3), green = Intercept only (model formulation 4)), and bootstrapped medians (colored points) are compared to the distribution of the best-performing null model (gray). (a) Bootstrapped distributions of $q^2$, metrics from resampling of the HBN holdout data for both parent-reported (SCARED-P) and child-reported (SCARED-C) anxiety symptoms. For the parent-reported anxiety, only GI models (GI summed and GI individual items) performed better than the null. For the child-reports, all models except for the intercept-only model tended to perform better than the null, with generally positive $q^2$, scores. (b) Receiver-operating characteristic curves for all model formulations for the full HBN holdout dataset. The $x$-axis represents specificity (proportion of those true diagnoses correctly predicted), and the $y$-axis represents sensitivity (proportion of true diagnoses correctly predicted). The diagonal black line represents chance. (c and d) Bootstrapped distributions of area under the curve (AUC) (panel C) and log loss metrics (panel d) from resampling of the HBN holdout data for clinician-consensus diagnoses on the KSADS. GI summed, GI individual items, and No-GI models treatment for children’s mental health (McGorry et al., 2018; Weersing et al., 2012), which may in fact be preferred given the high comorbidity in psychiatric presentations (Kessler et al., 1994). Indeed, comorbidity within the HBN cohort was high.

Given the associations between GI distress and anxiety, routine screening for anxiety symptoms within all GI distress presentations by pediatricians, teachers, school nurses, and other child-facing professionals could likely help to increase appropriate referrals (Cunningham et al., 2018). However, because these individuals at initial access points for youth mental healthcare do not always have training, resources, or time to administer such mental health evaluations, using GI symptoms to assess psychopathological risks may be useful as a preliminary step. Particularly, because somatic symptoms may be readily observed by teachers and school nurses and are often assessed at primary care visits, increased attention to such symptoms as potential indicators of mental health risk may help improve community monitoring of youth mental health. Moreover, because anxiety symptom screening may miss youth who are exhibiting prodromal or subthreshold anxiety, as well as youth who lack insight to report anxiety symptoms, GI symptoms (either alone or in addition to anxiety screening) may help identify youth who would benefit from psychiatric referral but may be missed by mental distress screening.

Even with increasing awareness towards issues of mental health over time, many key gatekeepers to community youth mental health-care still express a lack of confidence in their preparedness to identify youth psychopathology and their abilities to provide adequate care (Green et al., 2019; Horwitz et al., 2015; Reineke et al., 2011). Based on the generalizability and predictive validity of GI–anxiety associations investigated here, more easily observable GI complaints might then be one tool for addressing difficulties in identifying children at risk for psychopathology, requiring little extra time and training. Thus, while many calls have been issued for increasing mental health competencies in both pediatrics and in schools, increasing widespread intuitions about relationships between physical and mental health symptoms serves as a relatively simple and low-cost measure. Even small increases in knowledge of the interconnected nature of mental and physical health may serve as an incremental addition to the skills and confidence of child-facing professionals in identifying those at risk for psychopathology and improve community mental healthcare for youth (Green et al., 2019).
FIGURE 6    Exploring gastrointestinal (GI) model performance for predicting a range of mental illness diagnoses (left y-axis). Numbers in parenthesis next to the name of each diagnosis (left y-axis) indicate the proportion of participants in the full Healthy Brain Network (HBN) cohort with that diagnosis. Each outcome is grouped into a diagnostic category (box), with the grey title indicating the diagnostic category (e.g., “behavioral disorders”). The beta estimate is indicated with a colored point and the lines around the point indicate the 95% uncertainty interval. (a) Estimated logistic regression betas for a summed GI symptom term (including age and sex covariate) in models fit to the entire HBN cohort, associated with likelihood of each respective outcome. Estimates are in log-odds. (b) Cross-validated area under the curve (AUC) estimates for each respective disorder from iterative rounds of cross-validation leaving out 25% of the HBN dataset. While the models in (a) included covariates for age and sex, cross-validated models only had a GI predictor, but no terms for age or sex to avoid inflated model performance based on age-related differences in specific psychopathologies. Points represent median AUC, and lines represent 95% uncertainty intervals across rounds of cross-validation.
A key consideration for these data is the possibility that referring youth with GI distress to psychiatric services might impede an appropriate referral to gastroenterology. Importantly, in this study, GI complaints are presumed nonorganic (questions asked about symptoms with “no known medical cause”). Nevertheless, as psychological comorbidities in GI presentations are common in both organic and nonorganic abdominal pain (Raymer et al., 1984), we suggest that psychiatric referrals may remain appropriate for even organic abdominal pain in youth and encourage concurrent screening for organic causes for GI pain.

4.1 | Limitations

While the current work demonstrates the prerequisite generalizability of GI–anxiety associations for future development of screening instruments, we regard such models in their current form as valuable predictors of risk, but not stand-alone diagnostic screening tools. Though our models often performed above chance levels in predicting out-of-sample anxiety symptoms, such performance does not yet meet a level sufficient for clinical utility as a diagnostic screener. Our GI model AUC values in the range of 0.6–0.7 for classifying diagnoses and 10%–15% predicted variance for continuous symptoms at best, while reliably above chance, fall below standards for diagnostic screening algorithms (Dobrow et al., 2018). In fact, despite that GI–anxiety associations have been previously reported using similar measures within the LA cohort (Callaghan et al., 2020), GI model prediction of whether youth met clinical threshold on the SCARED-P was not reliably higher than chance in a holdout set of the combined LA + NYC cohorts (Figure S24). Nevertheless, that GI models demonstrate predictive ability above chance provides evidence for potential future uses of such approaches, particularly if more targeted GI symptom questions are used, or in combination with other potentially informative variables (such as other somatic distresses, sleep, early-life history).

A second limitation is that while GI models for parent-reported anxiety reliably outperformed models based on age and sex alone (No-GI models), this was not consistently true for child-reported anxiety or clinician-consensus diagnoses. One explanation for this is that because GI items were also parent-reported, better predictive performance for parent-reported anxiety outcomes may have been driven by parental reporting effects, rather than youth anxiety symptoms themselves. On the other hand, stronger relative performance for GI models for the parent-reported anxiety outcome could be attributable to the fact that parent-reported anxiety scores (SCARED-P) were not strongly associated with age, while child-reported anxiety tended to decrease with age, and clinician-consensus diagnoses tended to increase with age (see Tables S3 and S4). Accordingly, models based on age and sex alone performed somewhat better than the null for child-reported and clinician-diagnosed anxiety (see Figure 5). Consistent with prior work (Carter, 2015; Carter et al., 2011; Crystal et al., 1994), females also had higher average anxiety symptoms in all three cohorts and likelihood of anxiety diagnoses in the HBN training set (see Figure S4). Thus, despite the fact that including interaction terms between GI symptoms and age or sex did not meaningfully improve model performance during cross-validation within the HBN training data (see Figure 4, Figures S15 and S16), age and sex helped in predicting both child-reported and clinician-diagnosed anxiety outcomes.

Relatively, parent-reported GI items were taken from a screening measure for psychosocial development (CBCL), rather than from an instrument designed to assess GI distress (e.g., ROME IV). Because for each GI symptom, response options were limited to “not true,” “somewhat true,” or “very true,” this measure lacks the precision to measure nuanced differences in GI distress. Even when we summed all four of these GI items to create a scale from 0 to 8, scores above 3 were rare in all cohorts, and no participants scored above 4 in the LA or NYC cohorts (see truncated model fits in Figure 2). In addition, this four-item composite demonstrated poor internal consistency, which may have hampered its use in predictive models. Better performance might be expected in future models where predictors are designed to capture more specific GI distress. Future work could also leverage child-reported GI symptoms and/or biological measures associated with GI distress (e.g., the gut microbiota) to mitigate potential biases in parent reporting of child GI distress and improve prediction of youth anxiety.

Additionally, in the current study, we evaluate the ability of models to predict concurrent anxiety as an outcome, rather than future anxiety symptoms. While we believe there is value in this approach for informing community mental healthcare, work is also needed to evaluate model predictive performance for future anxiety in longitudinal datasets.

Another limitation is that models performed best in predictions of continuous anxiety symptoms in both holdout sets, rather than classification of binary anxiety diagnoses. While such performance does fit with spectrum-based conceptualizations of mental illness (i.e., Research Domain Criteria), future iterations of models that employ more specific assessments of GI distress should continue to monitor predictive performance on continuous and discrete outcomes.

A final concern regards the generalizability of the cohorts assessed here. The LA and NYC cohorts included youth who had experienced institutional or foster care abroad, both of which have been associated with GI distress and elevated anxiety (Bos et al., 2011; Bradford et al., 2012). The HBN cohort employed sampling strategies that lead to elevated rates of psychopathology relative to the population. In all cohorts, distributions of annual family income were well above local (NYC or LA) medians (see Table 1 and Figure S2), and there were more White participants than any other racial group (see Table 2).

An important consideration in relation to the racial and ethnic composition of the samples included here is that prior work has shown that somatic symptom expression in youth can vary by ethnicity. For example, Latino youth of some ethnocultural backgrounds express more somatic anxiety symptoms relative to European-American youth (Pina & Silverman, 2004; Varela & Hensley-Maloney, 2009; Varela et al., 2008). In our study, we did not find consistent evidence of differences in associations between GI symptoms and anxiety (see Figures S13 and S14) or model performance (see Figures S26–S29) as a function of ethnicity or race. However, the absence of such differences may have been a consequence of the relatively low numbers of non-White, and Hispanic or Latino youth (see Tables 1 and 2), or because the racial and...
Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field (Ambekar et al., 2009; Caplar et al., 2017; Dion et al., 2018; Dworkin et al., 2019; Maliniak et al., 2013). Here, we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman (Dworkin et al., 2009; Sood & Laahaprapan, 2018). By this measure (and excluding self-citations), our references contain 11.54% author of color (first)/author of color (last), 13.63% white author/author of color, 25.01% author of color/white author, and 49.83% white author/white author. This method is limited in that (a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and (b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

ACKNOWLEDGMENTS

The authors would like to thank Andrea Fields, Tricia Choy, Nicolas Camacho, Lisa Gibson, Anna Vannucci, and Nicole Gavrilova for their feedback and support in preparing this manuscript. The authors would also like to thank the Healthy Brain Network team for their efforts creating and managing open data repositories. This work was supported by funding from the National Science Foundation (Graduate Research Fellowship DGE 1644869) for PAB, the National Institute of Mental Health (R00MH113821 and R01MH091864), and the Brain & Behavior Research Foundation (24739) for BC, and the Dana Foundation for NT.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

CODE AND DATA AVAILABILITY STATEMENT

Code used for all data preprocessing and analyses is available at https://github.com/pab2163/giAnxiety. Further information on Data Usage Agreements for Healthy Brain Network data can be found at http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/Pheno_Access.html.

CITATION DIVERSITY STATEMENT

Recent work in several fields of science has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field (Ambekar et al., 2009; Caplar et al., 2017; Dion et al., 2018; Dworkin et al., 2020; Maliniak et al., 2013). Here, we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman (Dworkin et al., 2020; Zhou et al., 2020). By this measure (and excluding self-citations) of the first and last authors of our current paper), our references contain 33.63% woman(first)/woman(last), 17.55% man/woman, 24.78% woman/man, and 24.04% man/man. This method is limited in that (a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and (b) it cannot account for intersex, nonbinary, or transgender people. Second, we obtained predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color (Ambekar et al., 2009; Sood & Laahaprapan, 2018). By this measure (and excluding self-citations), our references contain 11.54% author of color (first)/author of color (last), 13.63% white author/author of color, 25.01% author of color/white author, and 49.83% white author/white author. This method is limited in that (a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and (b) it cannot account for Indigenous and mixed-race authors, or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

AUTHORS CONTRIBUTIONS

Paul Bloom and Bridget Callaghan planned and designed analyses, wrote the preregistration, analyzed the data, and drafted the manuscript. Niam Totten lead funding, recruitment, and data collection for two of the cohorts analyzed (LA and NYC cohorts), and contributed revisions to the manuscript. Michelle VanTieghem organized and cleaned the data for analysis and contributed revisions to the manuscript. Ian Douglas completed a code review for code used in analyses and contributed revisions to the analysis code and manuscript. All the authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

ORCID

Paul Alexander Bloom https://orcid.org/0000-0003-3970-5721
Bridget Callaghan https://orcid.org/0000-0001-5736-2510

REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.


