Journal Pre-proof

No robust cortical morphometry differences and limited morphometry-phenotype associations in a UK Biobank irritable bowel syndrome phenotype

Hanbing Gao , Carolin V. Schneider , Tobias Seibel , Chi Zhang , Fengyu Cong , Klaus Mathiak , Kai Markus Schneider , David M.A. Mehler

PII: S1053-8119(26)00522-7
DOI: https://doi.org/10.1016/j.neuroimage.2026.122207
Reference: YNIMG 122207

To appear in: *NeuroImage*

Received date: 9 February 2026
Revised date: 14 July 2026
Accepted date: 7 September 2026

Please cite this article as: Hanbing Gao , Carolin V. Schneider , Tobias Seibel , Chi Zhang , Fengyu Cong , Klaus Mathiak , Kai Markus Schneider , David M.A. Mehler , No robust cortical morphometry differences and limited morphometry-phenotype associations in a UK Biobank irritable bowel syndrome phenotype, *NeuroImage* (2026), doi: https://doi.org/10.1016/j.neuroimage.2026.122207

Highlights:

- No robust average cortical morphometry differences were detected between irritable bowel syndrome and controls.
- Equivalence testing provided statistical evidence of negligible effect sizes.
- Weak brain-phenotype associations and mostly not irritable bowel syndrome-specific.

No robust cortical morphometry differences and limited morphometry-phenotype associations in a UK Biobank irritable bowel syndrome phenotype

Hanbing Gao[a,b], Carolin V. Schneider[c], Tobias Seibel[c], Chi Zhang[b], Fengyu Cong[b,d,e,f], Klaus Mathiak[a,g], Kai Markus Schneider[c,h,i,j], David M. A. Mehler[a,k,l*]

[a] Department for Psychiatry, Psychotherapy and Psychosomatics, Medical School, RWTH Aachen University, Aachen, Germany

[b] School of Biomedical Engineering, Faculty of Medicine, Dalian University of Technology, Dalian, China

[c] Department of Internal Medicine III, Gastroenterology, Metabolic Diseases and Intensive Care, University Hospital RWTH Aachen, Aachen, Germany

[d] Faculty of Information Technology, University of Jyvaskyla, Jyvaskyla, Finland

[e] Key Laboratory of Social Computing and Cognitive Intelligence, Ministry of Education, Dalian University of Technology, Dalian, China

[f] School of Software Engineering, Dalian University, Dalian, China

[g] JARA-Translational Brain Medicine, RWTH Aachen University, Aachen, Germany

[h] Department of Medicine I, Department of Gastroenterology and Hepatology, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany

[i] Center for Regenerative Therapies Dresden (CRTD), Technische Universität (TU) Dresden, Dresden, Germany

[j] Else Kroener Fresenius Center for Digital Health, Faculty of Medicine and University Hospital Carl Gustav Carus, TUD Dresden University of Technology, Dresden, Germany

[k] Institute of Translational Psychiatry, Medical Faculty, University of Münster, Münster, Germany

[l] School of Psychology, Cardiff University Brain Research Imaging Center (CUBRIC), Cardiff University, Cardiff, U.K.

* Corresponding author

E-Mail: dmehler@ukaachen.de (DMAM)

Abstract

Irritable bowel syndrome (IBS) is a common disorder of gut-brain interaction that has sparked interest in neural markers. Yet, previous neuroimaging findings remain inconsistent and often fail to replicate, likely due to small sample sizes that overestimate effects and overlook patient heterogeneity. Using normative modeling to quantify cortical deviations from MRI in a UK Biobank digestive health questionnaire (DHQ)-derived IBS phenotype and matched controls (N = 3 413 each) relative to a healthy reference sample (N = 8 661), we found no robust group differences. Further, equivalence tests provided evidence that both group and group × sex effects were statistically equivalent to negligible effect sizes across nearly all regions. Within IBS, negative associations between cortical gray matter deviations and symptom or pain severity were observed for some regions. However, no statistically meaningful IBS-specific effects and no sex moderation were detected. Overall, our findings challenge previously reported structural alterations in IBS, underscoring the need for replicability in adequately powered designs.

Keywords: neuroimaging replicability, cortical morphometry, normative modeling, UK Biobank, equivalence test, irritable bowel syndrome

1 Introduction

Irritable bowel syndrome (IBS) is a highly prevalent functional gastrointestinal disease characterized by recurrent abdominal pain, bowel dysfunction, and disturbed intestinal motility,

affecting approximately 4-9% of the global population, with estimates varying depending on the diagnostic criteria applied (Oka et al., 2020; Sulaimi et al., 2025). Beyond gastrointestinal symptoms, IBS is frequently accompanied by psychiatric comorbidities, particularly anxiety and depression, which have also been highlighted as psychological factors associated with IBS in a recent systematic umbrella review (Fond et al., 2014; Seeling et al., 2023; Sulaimi et al., 2025). Current brain-gut-microbiome (BGM) models conceptualize IBS as a multilevel disorder involving bidirectional interactions among central neural circuits, the autonomic and enteric nervous systems, immune and endocrine signaling, stress biology, and the gut microbiome (Mayer et al., 2015; Zheng et al., 2020; Schneider et al., 2022, 2023; Ng et al., 2023; Mayer et al., 2023). These interactions highlight the importance of investigating the relationship between the brain and IBS symptoms.

A considerable body of neuroimaging literature has investigated structural magnetic resonance imaging (MRI) data to identify neuroanatomical alterations potentially contributing to the development and persistence of IBS within the BGM model (Barazanji et al., 2022; Chua et al., 2017; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2015, 2014; Li et al., 2020, 2024; Mayer et al., 2023, 2015; Piché et al., 2013; Seminowicz et al., 2010; Sun et al., 2023; Weaver et al., 2016; Zhao et al., 2023). Two relatively consistent findings that have been reported in female IBS patients are reduced cortical thickness (CT) in the anterior insula and increased CT and gray matter (GM) volume in sensorimotor regions such as the precentral and postcentral gyrus, two areas that are critically involved in somatosensory processing and visceral pain modulation (Weaver et al., 2016). Beyond regional morphometric differences, structural

variations in the insula, prefrontal, and cingulate cortices have been associated with symptom severity, pain perception, and emotional dysregulation in IBS (Barazanji et al., 2022; Chua et al., 2017; Grinsvall et al., 2021; Labus et al., 2015; Li et al., 2020, 2024). However, substantial inconsistencies remain across previous reports, with some reporting increased measures or positive associations with clinical phenotypes relative to healthy controls (HCs), whereas others have observed reductions or negative associations (Barazanji et al., 2022; Chua et al., 2017; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2015, 2014; Li et al., 2020, 2024; Piché et al., 2013; Seminowicz et al., 2010; Weaver et al., 2016). These mixed findings render previous reports less likely to be replicable.

Crucially, when it comes to replicability, most previous neuroimaging studies in IBS have relied on small sample sizes—often fewer than 100 individuals (Weaver et al., 2016)—and were hence severely limited in their statistical power and prone to high measurement variability (Button et al., 2013; Mehler, 2019). When subjected to publication bias, small samples also increase the chance for inflated effect sizes and false-positive findings; a broader issue increasingly recognized across the neurosciences including neuroimaging research (Algermissen and Mehler, 2018). More recent large-scale evaluations have shown that several hundreds to thousands of samples are often required to yield replicable brain-behavior associations, as small cohorts frequently produce non-replicable correlations between brain structure and psychological traits (Goltermann et al., 2025; Kharabian Masouleh et al., 2019; Liu et al., 2023; Marek et al., 2022; Marek and Laumann, 2025). Importantly, previous structural neuroimaging studies in IBS relied solely on conventional null-hypothesis significance testing (NHST), which can provide

evidence for the presence of effects, but not evidence for the absence of effects that are considered large enough to be relevant (Algermissen and Mehler, 2018; Barazanji et al., 2022; Chua et al., 2017; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2015, 2014; Lakens, 2017; Lakens et al., 2018; Li et al., 2020, 2024; Mayer et al., 2023, 2015; Mehler et al., 2019; Piché et al., 2013; Seminowicz et al., 2010; Sun et al., 2023; Weaver et al., 2016; Zhao et al., 2023). Furthermore, most IBS studies have focused on females, likely due to a sampling bias resulting from the fact that women are 1.5-3.0 times more likely to develop IBS and 2-2.5 times more likely to seek medical care for IBS-related symptoms compared to men (Drossman et al., 1993; Mayer et al., 2023), leaving male patients underrepresented and potentially biasing group comparisons (Weaver et al., 2016). Together, these factors underscore the need for large, population-based samples and adequately powered statistical approaches to improve robustness, replicability, and evaluate the practical relevance of detected effect size.

The UK Biobank (UKB) provides large-scale neuroimaging, clinical, and demographic data from over 500 000 participants (Sudlow et al., 2015), offering the statistical sensitivity needed to overcome replicability challenges in neuroimaging studies. The availability of digestive health information, such as the digestive health questionnaire (DHQ), together with MRI and psychometric data, enables well-powered analyses of brain morphology and its associations with clinical features such as pain, anxiety, and depression. Moreover, the large healthy cohort provides an ideal sample for approaches such as normative modeling (NM) (Rutherford et al., 2023, 2022a) that allow individual-level deviations to be quantified based on a pretrained healthy reference model.

NM builds such a reference cohort model to estimate neuroanatomical deviations in patients while accounting for age, sex and scanner-related variability (Bayer et al., 2025; Bethlehem et al., 2020; Elad et al., 2021; Marquand et al., 2016; Rutherford et al., 2023, 2022b, 2022a; Verdi et al., 2024b; Zabihi et al., 2019). Successfully applied in psychiatric diseases (Bayer et al., 2025; Bethlehem et al., 2020; Elad et al., 2021; Verdi et al., 2024a; Zabihi et al., 2019), NM moves beyond group averages by capturing individual-specific morphometric variability and enabling analyses of how deviations relate to clinical features (Marquand et al., 2016; Rutherford et al., 2023, 2022b). Extending this approach to IBS offers a principled way to examine cortical morphology with greater precision and interpretability, including potential links to symptom severity and gut-brain interaction (Marquand et al., 2016; Rutherford et al., 2023, 2022b).

Here, we applied NM to UKB neuroimaging data to quantify individual-level cortical gray matter deviations in key morphological features, including CT, surface area (SA) and cortical volume (CV), in IBS and propensity-score-matched (PSM) HCs (Kline and Luo, 2022) relative to healthy reference models. Our first objective was to systematically investigate whether gray matter cortical alterations previously reported in IBS are replicable in a large, well-powered cohort, with a particular focus on brain regions implicated in pain perception and emotional processing. Second, we examined whether cortical deviations show brain-phenotype associations for IBS-related clinical features and assessed whether these associations were different in IBS compared to HCs to a statistically meaningful extent. Further, we explored if

potential morphometry variations or brain-phenotype associations were different between males and females as sex has been discussed as a potentially relevant factor in the IBS neuroimaging literature (Gupta et al., 2017; Jiang et al., 2013; Jin et al., 2019; Labus et al., 2014). Moving beyond conventional NHST, we followed up all primary hypothesis tests with equivalence tests, that allow us to reject a range of effect sizes that are considered too small to be statistically relevant (Senn, 2021). Moreover, for cases where the initial null hypothesis could not be rejected in the primary test, equivalence tests allow us to distinguish between "no evidence for an effect" and more informative statements about "evidence for no statistically relevant effects" (Campbell, 2024; Gerchen et al., 2021; Hanel and Mehler, 2019; Lakens et al., 2018; Walker and Nowacki, 2011). Finally, we conducted sensitivity analyses to assess whether the findings were influenced by IBS subtype heterogeneity or by the use of NM rather than conventional raw-measure case-control analyses. Together, using a large-scale UKB data sample, our study aims to provide robust insights into replicability of cortical morphometric alterations in IBS and to offer statistical evidence for rejecting meaningless, small effects.

In light of recent critiques that have highlighted risks that accompany the use of large datasets – including oversimplified phenotyping, selective analyses, and insufficient correction for multiple comparisons (Rudan et al., 2025; Suchak et al., 2025) – we aimed to address these concerns by using a transparent methodology, careful data selection, and robust statistical correction procedures. Specifically, we adopted these principles through a systematically designed normative modeling framework, following a guideline for reporting analyses of big data repositories open to the public (GRABDROP) (Rudan et al., 2025), conducting multiple

comparison corrections for all tests, and employing an established equivalence testing approach to evaluate the replicability and magnitude of IBS-related morphometric differences beyond conventional null hypothesis significance testing (Campbell, 2024; Gerchen et al., 2021; Hanel and Mehler, 2019; Lakens, 2017; Lakens et al., 2018; Mehler et al., 2019; Walker and Nowacki, 2011).

2 Materials and Methods

2.1 Study samples

2.1.1 UK Biobank dataset

Data were retrieved from the UK Biobank (UKB), a large-scale biomedical study with over 500 000 individuals. Brain imaging data were acquired at four sites using 3T scanners (Siemens Skyra) and processed using UKB's standardized pipeline for CT, SA and CV estimation, based on Destrieux parcellation (Destrieux et al., 2010), and quality control operated on behalf of UKB (Alfaro-Almagro et al., 2018).

Participants were eligible if they had complete brain MRI data from image Instance 2 (2014-2022) and available demographic information. To define the healthy and IBS groups, we excluded individuals with gastrointestinal disorders except IBS, major depressive disorder, neurological or psychiatric disorders, cerebrovascular disease, intracranial injury, neoplasms, or other major exclusionary medical conditions. Exclusion criteria were derived from ICD-10 diagnosis codes, OPCS-4 surgical procedure codes, self-reported non-cancer illness codes, DHQ-responses, mental health questionnaire responses, and UKB imaging quality control

information (See S1-S2 Tables for full criteria and corresponding data fields). After applying exclusion and data-quality criteria, 15 487 individuals with available demographic data and brain imaging measures remained for analysis.

This study was conducted under UKB application ID: 71300, approved by the UKB Access Committee. The study was covered by the ethical approval for UKB studies from National Health Service (NHS) National Research Ethics Service on 17 June 2011 and extended on 18 June 2021 with written informed consent obtained from all participants (Data-field: 200). To ensure transparent reporting of secondary analyses of a large-scale public dataset, we followed the Guidelines for Reporting Analyses of Big Data Repositories Open to the Public (GRABDROP) (Rudan et al., 2025). The completed checklist is provided in the S1 File.

2.1.2 Definition of IBS and healthy groups

A total of 3 413 IBS cases were identified using a DHQ-derived approximation of Rome III IBS criteria based on self-reported digestive symptoms from the UKB (S3 Table) (Douglas A. Drossman, 2006; Eijsbouts et al., 2021). Matched HCs were selected using propensity score matching (PSM) to improve covariate overlap before case-control inference (Kline and Luo, 2022). IBS cases were matched 1:1 to HCs using nearest-neighbor matching without replacement, based on the logit of the propensity score estimated from age, sex, and MRI scanner site indicators. Covariate balance was evaluated using standardized effect sizes before and after matching. Although age, sex, and MRI scanner site were also included as covariates in the normative model, PSM served a distinct design-stage purpose. It reduced distributional imbalance between IBS cases and HCs in the analytic case-control sample, whereas the

normative model estimated expected brain morphometry and individual-level deviation scores conditional on these covariates. The remaining healthy participants served as the reference cohort for training and evaluating the NM framework, using an 8:2 split.

IBS subtypes were defined using UKB DHQ data based on standard Rome III definitions of constipation-predominant IBS (IBS-C), diarrhea-predominant IBS (IBS-D), mixed IBS (IBS-M), and unclassified IBS (IBS-U), according to the frequency of hard or lumpy stools versus loose, mushy or watery stools (S3 Table).

2.1.3 Clinical assessment

Clinical symptom data were collected through UKB self-reported questionnaires, including the IBS Symptom Severity Score (IBS-SSS) assessing abdominal pain intensity and frequency, abdominal distention, bowel habit satisfaction, and interference with activities of daily living (Francis et al., 1997); the Patient Health Questionnaire-12 (PHQ-12) somatic symptom score assessing non-digestive somatic symptoms (Spiller et al., 2010) (To avoid systematic sex-related bias, menstrual symptoms was excluded (Eijsbouts et al., 2021; Polster et al., 2019)); the Generalized Anxiety Disorder-7 (GAD-7) score (Spitzer et al., 2006); and PHQ-9 score (Kroenke et al., 2001) to assess the severity of anxiety and depressive symptoms. These questionnaires were used as continuous symptom-severity measures rather than diagnostic instruments. Non-zero scores in healthy controls reflect symptom burden rather than clinical diagnosis. Questionnaires and corresponding UKB field codes are available in S4 Table.

To evaluate temporal alignment between phenotype assessment and neuroimaging, we

calculated signed day differences between the MRI acquisition date and DHQ completion date. We summarized the median and interquartile range (IQR) of these intervals and the proportion of participants whose absolute DHQ-MRI interval was within 3 years.

2.1.4 Covariates and medication variables

Additional demographic, lifestyle, and medication variables were derived for sample characterization and covariate-balance assessment rather than for PSM or primary statistical models, because not all variables were completely available and would have reduced the analytic sample size. BMI and smoking status were extracted from UKB imaging assessment and questionnaire fields. Years of education were derived from education qualifications, with the highest qualification retained for each participant and converted into approximate years of education as follows: college/university degree = 16 years; A/AS levels = 12 years; O levels/GCSE/CSE = 10 years; NVQ/HND/HNC = 13 years; other professional qualification = 14 years; none of the above = 6 years; prefer not to answer = missing (Lennon et al., 2023). Daily alcohol intake was calculated based on alcohol consumption questionnaire fields from UKB category 100051 (Schneider et al., 2024). Medication conditions were derived from data field 6154 at the imaging visit. Participants endorsing aspirin, ibuprofen, or paracetamol were coded as using pain medication. Previous IBS diagnosis was derived from first occurrences of ICD-10 diagnosis code K58 in category 1712, requiring the diagnosis date to precede the imaging visit date for each participant.

2.1.5 Regions of interest

To improve statistical power and inform our hypotheses, we focused on 30 regions of interest

(ROIs) per hemisphere linked to IBS in prior studies (Grinsvall et al., 2021; Orand et al., 2015; Weaver et al., 2016). These included motor, cingulate, and insular cortices, among other regions, that are associated with pain processing, sensorimotor integration, emotional regulation, and interoception. The full list of ROIs is provided in S5 Table.

2.2 Normative modeling

2.2.1 Reference model construction

NMs were constructed using the Python-based PCNtoolkit (version 0.26) (Marquand et al., 2016; Rutherford et al., 2022b). Warped Bayesian linear regression was applied to predict brain measures from age, sex, and site, incorporating the SinhArcsinh warping function to account for distributional skewness and kurtosis (Fraza et al., 2021). Reference models were trained separately for each ROI and brain measure using 80% (N = 6 929) of the reference healthy dataset and evaluated on the remaining 20% (N = 1 733).

2.2.2 Model evaluation

Model performance was evaluated across the validation, matched HCs and IBS datasets using explained variance (EV), mean standardized log loss (MSLL), skewness, kurtosis, and correlation between true and predicted values (Fraza et al., 2021; Rutherford et al., 2022a, 2022b). These metrics were used to assess predictive accuracy, uncertainty calibration, and distributional properties of the resulting deviation scores. False discovery rate (FDR) correction was applied using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Model performance was then statistically compared across datasets to assess the generalizability of

the normative models.

2.2.3 Brain deviation scores

Deviation scores (z-scores) quantified individual deviations from the normative range for each ROI and brain measure (Fraza et al., 2021):

$$z_{nd} = \frac{y_{nd} - \hat{y}_{nd}}{\sqrt{\sigma_d^2 + (\sigma_*^2)_d}}$$

where y_{nd} represents the true value and $\hat{y}_{nd}$ stands for the predicted value for subject n's brain measure d. σ_d^2 and $(\sigma_*^2)_d$ represent the estimated noise variance and the variance attributable to modeling uncertainty.

2.3 Statistics and reproducibility

To assess case-control differences and their clinical relevance in IBS, we first applied general linear models (GLMs) to regional deviation scores across CT, SA and CV. Equivalence tests were then performed on the same contrasts to determine whether groups were statistically equivalent within a negligible effect range. Moreover, brain-phenotype association analyses were conducted to explore correlations between cortical deviation scores and clinical symptoms, including IBS symptom severity, pain severity and mental health measures such as depression and anxiety. Multiple testing was controlled for using FDR correction with the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995), applied separately within each family of comparisons.

2.3.1 Group-level analysis

GLMs were applied to examine group differences in brain structure deviations across ROIs for CT, SA and CV brain measures separately, including group (IBS vs. HCs), sex, and their interaction (group × sex) as fixed effects. As normative models already accounted for age, sex and scanner effects during training, these were not included as covariates in subsequent GLMs:

$$z_{nd} = \beta_0 + \beta_1(Group) + \beta_2(Sex) + \beta_3(Group \times Sex) + e_{nd}$$

where z_{nd} is the deviation score from the NM for subject n for regional brain measure d. FDR correction was conducted across ROIs separately for each morphometric measure (CT, SA, and CV). GLMs were implemented using the *ols* function from *statsmodels.formula.api* (version 0.14.0) (Seabold and Perktold, 2010).

2.3.2 Equivalence tests

Given that the relatively large sample size could detect even small effects, equivalence tests (Lakens et al., 2018) were conducted to evaluate whether brain structure deviations between groups (IBS - HCs) were statistically equivalent within prespecified ranges of effects considered marginally meaningful (Mehler et al., 2019). The two one-sided tests (TOST) procedure was applied using 90% confidence intervals (CIs). To distinguish statistical sensitivity from substantive interpretation, we used two smallest effect size of interest (SESOI) specifications. First, we used power-based SESOI bounds corresponding to the smallest detectable effect size at 95% statistical power ($\beta = 0.05$) (Campbell, 2024; Lakens, 2017): |d| = 0.088 for the main IBS-HC effect and |d| = 0.099 for the group × sex interaction effect. Second, we used a neuroimaging-informed small-effect bound of |d| = 0.15 as a conservative threshold below Cohen's conventional small effect of |d| = 0.20, while acknowledging that effects smaller

than conventional benchmarks may still be meaningful in large-scale behavioral and neuroimaging research (Cohen, 2013; Funder and Ozer, 2019; Marek et al., 2022; Yan et al., 2026). Equivalence was interpreted as evidence against effects at or above the corresponding SESOI bound. FDR correction was applied within each of the three families of comparisons (i.e., across all ROIs for each brain measure, separately).

2.3.3 Cortical brain-phenotype association analyses

To explore associations between cortical deviations and clinical symptoms, bivariate correlation analyses were conducted between deviations of CT, SA, CV and questionnaire scores (IBS-SSS, PHQ-12, GAD-7, PHQ-9; item details in S4 Table) across ROIs in individuals with complete questionnaire data (Table 1). Correlations were estimated separately for IBS and HCs groups, and the effect sizes of these correlations were evaluated using equivalence tests with SESOIs of |0.139| for IBS and |0.176| for HCs, corresponding to the smallest detectable effect sizes at 95% statistical power ($\beta = 0.05$) (Campbell, 2024; Lakens, 2017).

Group differences in correlation strength were assessed using Fisher's Z-transformation (Razali and Wah, 2011), with standardized effect size expressed as Cohen's q:

$$q = \frac{1}{2} log \frac{1 + r_1}{1 - r_1} - \frac{1}{2} log \frac{1 + r_2}{1 - r_2}$$

where r represents the correlation coefficient being compared. Equivalence tests for group-difference effect sizes were conducted with a SESOI of |0.112|, corresponding to the smallest detectable Cohen's q at 95% statistical power ($\beta = 0.05$) (Campbell, 2024; Lakens, 2017).

To explore sex moderation, DiD analyses compared the sex-related differences in correlation

strength between IBS and HCs (Baker et al., 2025; Bodner, 2017; Gupta et al., 2017; Hedges et al., 2023; Salminen et al., 2022):

$$\Delta q = q_{IBS} - q_{HC}$$

where q_{IBS} represent the sex differences between IBS females and IBS males (r_1: correlation coefficient of IBS females; r_2: correlation coefficient of IBS males), and q_{HC} represent such differences between HC females and HC males. Equivalence tests of DiD effect sizes used a SESOI of Cohen′s q = |0.256|, the smallest detectable Cohen's q at 95% statistical power ($\beta = 0.05$) (Campbell, 2024; Lakens, 2017).

The Shapiro-Wilk test was used to evaluate normality of the variables (Ramseyer, 1979). Pearson's correlation was applied if both variables were normally distributed; otherwise, Spearman's correlation was applied. FDR correction was applied within each family of comparisons for correlations, correlation differences, DiD analyses, and equivalence tests separately (i.e. across all ROI-based correlation analyses for each brain measure and each questionnaire, separately). Analyses were conducted using *scipy.stats* (version 1.10.0) and *pandas* (version 1.5.3) in Python (version 3.8.16) (McKinney, 2010; Virtanen et al., 2020).

2.3.4 Individual-level and multivariate exploratory analyses

To further assess whether normative modeling revealed individual-level or multiregional deviation patterns not captured by the primary ROI-wise IBS−HC analyses, we conducted a set of exploratory analyses using the full 180-dimensional deviation-score profiles across CT, SA, and CV.

First, for each ROI and modality, we calculated the proportion of HCs and IBS participants showing extreme deviations, defined as $|z| \geq 2$, and visualized these rates on cortical surface maps. ROI-wise IBS−HC comparisons of extreme-deviation rates were performed with false discovery rate (FDR) correction within each modality. Second, we examined multiregional deviation-profile structure using principal component analysis (PCA) on the full 180 deviation scores after median imputation and feature standardization. Group overlap between HCs and IBS was visualized using the first two principal components. Third, we examined whether symptom severity was related to individual extreme-deviation burden. For each modality, we derived burden metrics based on extreme deviations and estimated Pearson correlations between these metrics and symptom scales (IBS-SSS, PHQ-12, PHQ-9, and GAD-7) separately in HCs and IBS. Between-group differences in these associations were quantified using Cohen's q, defined as the difference between Fisher z-transformed correlations in IBS and HCs. FDR correction was applied across these association-difference tests.

2.3.5 IBS subtype sensitivity analyses

IBS subtype sensitivity analyses were conducted to evaluate whether heterogeneity across IBS subtypes could mask an overall IBS-HC effect. IBS cases were classified as IBS-C, IBS-D, IBS-M, or IBS-U according to Rome III criteria using DHQ-derived bowel-habit information (described in Section 2.1.2). For each cortical measure, linear models were fitted to deviation scores using the same model structure as in Section 2.3.1, with sex included as a covariate and group coded as a five-level factor: HC, IBS-C, IBS-D, IBS-M, IBS-U. HC was used as the

reference group. We tested the omnibus five-level group effect to assess whether any IBS subtype differed from HCs, and the five-level group × sex interaction to evaluate sex-specific subtype effects. Subtype heterogeneity was further evaluated using nested model comparisons, comparing a binary IBS-HC effect model against a five-level subtype model allowing separate effects for IBS-C, IBS-D, IBS-M, and IBS-U. The same nested-comparison approach was repeated for sex interaction terms to assess sex-specific subtype heterogeneity. FDR correction was performed separately within each modality for each analysis (CT, SA, CV).

2.3.6 Deviation score vs. conventional case-control sensitivity analyses

To improve interpretability to readers familiar with conventional case-control analyses, we additionally performed raw-measure sensitivity analyses using the original cortical morphometry measures in the same matched HCs and IBS cases. For each hemispheric ROI and modality, raw CT, SA, and CV values were z-standardized within the pooled HCs and IBS sample and modeled using GLMs with group, sex, and their interaction as predictors, and age and scanner site as covariates:

$$z(raw\ \ measure) \sim group \times sex + age + scanner\ \ site$$

where group and sex were coded as in the primary analyses, and scanner site was included as a categorical covariate. FDR correction was applied separately within each modality and effect. To directly evaluate concordance with the normative modeling-based results, we correlated the raw-measure group and group × sex effect sizes with the corresponding effect sizes from the primary NM-deviation analyses across ROIs using Pearson correlation.

3 Results

A schematic overview of the study pipeline is presented in Fig 1. To test whether previously reported cortical morphometric differences in IBS can be replicated in a large population-based sample, we analyzed data from 3 413 IBS cases and age- and sex-matched HCs from the UKB. Normative models were trained using healthy reference data to quantify individual-level brain deviations in CT, SA, and CV. The resulting cortical deviation z-scores were used to assess both mean group differences and cortical brain-phenotype association differences between IBS and HCs, as well as interactions with sex. All primary statistical analyses were complemented by equivalence tests to evaluate the relevance of effect size estimates and aid interpretation of inconclusive primary test results in particular. Detailed methodological information on the analysis pipeline is provided in the Methods section.

Fig 1. **Schematic overview of the study design.** Structural MRI and clinical data from the UK Biobank were applied to investigate brain morphometric deviations in IBS within a large population cohort. A healthy reference sample (N = 8 661) was used to train normative models with age, sex, and scanner included as covariates (upper). Individual deviation scores were then computed for IBS cases and matched HCs (N = 3 413 each) across regions of interest (ROIs) and brain measures. Subsequent analyses assessed group-level differences in deviations and their associations with clinical phenotypes across ROIs.

3.1 Sample demographics and clinical characteristics

A total of 37 843 individuals were excluded due to exclusion criteria (e.g. other medical conditions) or data quality criteria (see Methods and exclusion criteria in S1 and S2 Tables for full details). The remaining 15 487 individuals with available demographic data and brain MRI measures for CT, SA and CV were included in the study. Based on inclusion criteria for IBS diagnosis (S3 Table), the final analytic sample comprised 3 413 individuals with IBS and 3 413 propensity score-matched HCs. Education and alcohol intake were incompletely available in the matched sample: education was available for 3 380 IBS cases and 3 384 HCs, and alcohol intake was available for 3 258 IBS cases and 3 310 HCs. Because not all participants completed the clinical questionnaires, brain-phenotype analyses included 2 712 IBS and 1 699 HCs with complete behavioral data. The remaining 8 661 healthy individuals served as the reference dataset for training and evaluating the brain-structure NM framework.

Table 1 summarizes the demographic, lifestyle, medication, and clinical characteristics of the matched case-control and IBS cases. IBS cases were further classified as IBS-C (N = 611), IBS-D (N = 877), IBS-M (N = 1 832), or IBS-U (N = 93). Matching substantially improved covariate balance for the prespecified matching variables. Age imbalance was reduced from a standardized mean difference (SMD) of 0.219 before matching to 0.073 after matching, and exact balance was achieved for sex, with the SMD reduced from 0.536 to <0.001. Hence, after matching, IBS and HCs groups showed only negligible differences in age, BMI, education, alcohol intake, smoking status, and pain medication use, with all absolute effect sizes below 0.1. The IBS group was defined using Rome III criteria based on self-reported UKB DHQ

symptoms; 21.9% of IBS cases also had a previous ICD-10 IBS diagnosis before imaging.

Table 1. **Demographic and clinical characteristics of the analysis samples**

Characteristic	HCs (N = 3 413)	IBS (N = 3 413)	Effect size	IBS Subtypes			
				IBS-C (N = 611)	IBS-D (N = 877)	IBS-M (N = 1 832)	IBS-U (N = 93)
Age, years	61.45 (7.47)	62.09 (7.61)	**-0.086**	62.64 (7.93)	61.92 (7.20)	61.94 (7.70)	63.09 (7.22)
Male sex, n (%)	921 (27.0%)	921 (27.0%)	0.000	79 (12.9%)	326 (37.2%)	487 (26.6%)	29 (31.2%)
BMI, kg/m^2	25.59 (4.17)	25.96 (4.50)	**-0.084**	25.04 (3.96)	26.13 (4.48)	26.25 (4.66)	24.62 (3.77)
Education, years	14.04 (2.65)	14.10 (2.56)	-0.022	14.04 (2.66)	14.34 (2.28)	13.99 (2.65)	14.29 (2.35)
Alcohol intake, g/day	12.88 (13.59)	12.50 (14.16)	0.027	9.35 (9.47)	15.06 (16.22)	12.35 (14.25)	12.00 (12.92)
Smoking status, n (%)			**0.067**				
Never	2397 (70.7%)	2173 (64.3%)		394 (65.1%)	554 (64.0%)	1162 (63.9%)	63 (68.5%)
Previous	909 (26.8%)	1112 (32.9%)		199 (32.9%)	290 (33.5%)	595 (32.7%)	28 (30.4%)
Current	83 (2.4%)	96 (2.8%)		12 (2.0%)	21 (2.4%)	62 (3.4%)	1 (1.1%)
Pain medication, n (%)	235 (6.9%)	323 (9.5%)	**0.047**	48 (7.9%)	71 (8.1%)	194 (10.6%)	10 (10.8%)
Previous IBS diagnosis, n (%)	0 (0.0%)	746 (21.9%)	**0.350**	120 (19.6%)	179 (20.4%)	433 (23.6%)	14 (15.1%)
IBS-SSS	3.03 (3.78)	17.98 (9.26)	**-1.961**	18.03 (9.47)	17.30 (9.14)	18.58 (9.15)	12.19 (8.61)
PHQ-12	3.51 (2.30)	6.49 (3.19)	**-1.034**	6.30 (3.01)	5.96 (3.21)	6.87 (3.18)	5.23 (3.12)
GAD-7	1.03 (1.87)	2.93 (3.79)	**-0.605**	2.80 (3.84)	2.71 (3.66)	3.14 (3.85)	1.89 (3.28)
PHQ-9	1.34 (1.90)	3.58 (3.94)	**-0.689**	3.46 (3.88)	3.23 (3.81)	3.85 (4.03)	2.43 (3.28)

1. Values are mean (SD) or n (%).

2. Effect sizes compare matched HCs and all IBS cases. For continuous variables, Cohen's d was calculated as mean(HC) – mean(IBS), so negative values indicate higher means in IBS. For categorical variables, effect sizes are reported as ϕ/Cramér's phi and are unsigned; direction should

be inferred from the percentages. Effect sizes marked in bold indicate corresponding p value < 0.05.

3. BMI, body mass index; HCs, healthy controls; IBS, irritable bowel syndrome; IBS-C, IBS constipation-predominant; IBS-D, IBS diarrhea-predominant; IBS-M, IBS mixed/alternating constipation and diarrhea; IBS-U, IBS unclassified; IBS-SSS, IBS symptom severity score; PHQ-12, somatic symptom severity; GAD-7, anxiety symptoms; PHQ-9, depressive symptoms.

As expected, individuals with IBS exhibited significantly higher scores on clinical measures of symptom severity (IBS-SSS), pain (PHQ-12), depression (PHQ-9), and anxiety (GAD-7) compared to HCs with large effect sizes (all $p < 0.05$; all |Cohen's d| > 0.6; item descriptions for each scale are provided in S4 Table). Similar patterns were observed across IBS subtypes descriptively, although IBS-U had the quite small sample size (N = 93).

DHQ-MRI timing was also examined to assess the temporal proximity between self-reported symptoms and imaging (S2 Fig). In the matched sample, DHQ completion occurred a median of 359 days before MRI in IBS participants (IQR: 699 days before to 138 days after) and 461 days before MRI in HCs (IQR: 739 days before to 2 days before). The absolute DHQ-MRI interval was within 3 years for 95.5% of IBS participants and 96.3% of HCs.

3.2 Quality control and evaluation of normative models

Normative models were fitted separately for both hemispheres across 30 ROIs selected based on prior MRI studies of IBS (see S5 Table for the ROI list). Model evaluation across the validation, matched HCs, and IBS datasets indicated stable performance across evaluation

metrics, including EV, MSLL, skewness, kurtosis, and correlations between observed and predicted values (S1 Fig; S6 Table), consistent with previous NM studies (Fraza et al., 2021; Rutherford et al., 2023, 2022a). Generalizability was further supported by the similarity of model performance metrics between the validation and matched HC datasets, both of which comprised healthy participants independent from the training set and were expected to represent the same reference population.

3.3 Structural brain deviations showed no mean group differences between IBS and HCs

GLMs including group, sex and their interaction as fixed effects found no FDR-corrected significant main effect of group for any brain measure across all ROIs (all |effect sizes| < 0.065, all $p_{FDR} > 0.05$; full model parameters are provided in S7 Table). Using the neuroimaging-informed SESOI of |d| = 0.15, all group effects were statistically equivalent to zero after FDR correction. Under the stricter power-based SESOI of |d| = 0.088, equivalence was observed for most measures: 54/60 ROIs for CT, 56/60 for SA, and 52/60 for CV. As shown in Fig 2.a, the 90% CIs for most regions fell well within the equivalence bounds ($max(p_{lower}, p_{upper})_{FDR} < 0.05$; CIs in black), indicating FDR-corrected equivalence, while a small subset of regions yielded inconclusive results where CIs overlapped the power-based SESOI boundary ($max(p_{lower}, p_{upper})_{FDR} > 0.05$; CIs in blue). Detailed equivalence test statistics are provided in S8 Table.

Fig 2. **Equivalence tests for group and group × sex effects across ROIs and brain measures.** Each point represents the effect size with 90% confidence intervals (CIs) for **a** the main group effect and **b**

the group × sex interaction effect across 60 ROIs (30 per hemisphere). Blue CIs indicate regions with inconclusive results after false discovery rate (FDR) correction, whereas black CIs indicate regions showing statistically significant equivalence after FDR correction. Dashed vertical lines represent the boundaries of the smallest effect size of interest (SESOI) for the power-based SESOI (black) and neuroimaging-informed SESOI (blue). Abbreviations: ROIs, regions of interest; CT, cortical thickness; SA, surface area; CV, cortical volume.

No FDR-corrected significant group × sex interaction effect was observed across ROIs (all |effect sizes| < 0.062, all p_{FDR} > 0.05; full model parameters in S9 Table). Using the neuroimaging-informed |d| = 0.15 SESOI, all tested interaction effects were statistically equivalent to zero after FDR correction. With the stricter power-based SESOI of |0.099|, equivalence was observed for nearly all interaction effects: 60/60 ROIs for CT, 57/60 for SA, 57/60 for CV (Fig 2.b). Only a small number of SA and CV measures in regions around the central sulcus yielded inconclusive effects. Detailed equivalence test results for interaction effects are summarized in S10 Table.

Overall, these findings suggest that neither the main group effect nor its interaction with sex revealed any replicable morphometric differences in IBS. Equivalence tests further indicated that nearly all effects were statistically equivalent and thus negligible, allowing small, likely meaningless differences overall to be conclusively rejected.

3.4 Limited evidence of IBS-specific morphometry-phenotype associations

Cortical gray matter deviations showed only limited associations with relevant clinical scales of IBS symptom severity (IBS-SSS), IBS-related pain (PHQ-12), depression (PHQ-9) and anxiety (GAD-7). Within the IBS group, deviation scores were negatively correlated with IBS-SSS sum scores across frontal, cingulate, and sensorimotor areas for CV and SA measures (-0.087 < r < -0.042; -0.175 < Cohen's d < -0.085; all $p_{FDR} < 0.05$). Likewise, significant negative correlations between PHQ-12 sum scores and CV/SA were also observed in IBS, particularly in postcentral, cingulate, and temporal cortices (-0.095 < r < -0.049, -0.191 < Cohen's d < -0.099, all $p_{FDR} < 0.05$). While these correlations were statistically significant and equivalence tests were non-significant (i.e., effect sizes were not equivalent to the SESOI; equivalence tests' $p_{FDR} > 0.05$), their CIs overlapped with the equivalence bounds, indicating that these effects remain inconclusive rather than clearly meaningful (S11 Table). Hence, the correlation results only provided limited evidence for cortical brain-phenotype associations. Corresponding brain maps with inconclusive effects are shown in S3 Fig. Further, in line with our expectations, no significant cortical brain-phenotype association was observed in HCs.

To assess group differences in correlation strengths between IBS and HCs, Fisher's Z tests were used. In primary NHST, several regions showed significantly stronger negative correlations in IBS than in HCs for IBS-SSS (Fig 3.a), including the right postcentral sulcus (CV, Cohen's q = 0.116; SA, Cohen's q = 0.122), left postcentral sulcus (SA, Cohen's q = 0.101), right postcentral gyrus (SA, Cohen's q = 0.107) and right supramarginal gyrus (SA, Cohen's q = 0.100). Further,

equivalence tests showed that these effect sizes were not statistically equivalent to the SESOI (p_{FDR} > 0.05; red CIs in Fig 3.b). However, their CIs still overlapped with the SESOI bounds, indicating that these group differences remain inconclusive rather than clearly meaningful (Fig 3.b). No significant group difference was observed for PHQ-12 (all Cohen's q < 0.1, p_{FDR} > 0.05) or for other clinical scales (S3 Fig). Complete results for within-group associations and association differences between groups are provided in S11 Table.

Fig 3. **Cortical brain-phenotype association differences between IBS and HCs. a** Primary test: Scatter plots illustrating cortical brain-phenotype relationships in regions where IBS and HCs differ significantly (p_{FDR} < 0.05). **b** Equivalence test: Forest plots showing effect sizes of between-group association differences and their 90% CIs relative to a SESOI of |0.112|. Regions in red indicate inconclusive effect sizes where the CIs overlap with the equivalence bounds (p_{FDR} > 0.05), whereas regions in black indicate effect sizes statistically equivalent to the SESOI. Abbreviation: CT, cortical thickness; SA, surface area; CV: cortical volume; ES, effect size; CI, confidence interval.

Finally, to evaluate whether detected associations were moderated by sex, we quantified group × sex differences using a difference-in-differences (DiD) framework and tested the equivalence of DiD effect sizes (Baker et al., 2025; Bodner, 2017; Hedges et al., 2023). Across regions, most effects were statistically equivalent to the SESOI, with mean DiD effect sizes clustering around zero; only a few regions were inconclusive (S4 Fig; complete results in S12 Table). Hence, overall, these results provide mostly conclusive evidence for the absence of a sex-moderation effect for IBS-specific cortical brain-phenotype associations.

3.5 Individual-level and multivariate exploratory analyses

Exploratory analyses did not reveal robust individual-level or multiregional deviation patterns beyond the primary ROI-wise null findings. ROI-level extreme-deviation maps showed broadly similar spatial distributions in HCs and IBS participants across CT, SA, and CV, and no ROI-wise IBS−HC differences in $|z| \geq 2$ extreme-deviation rates survived FDR correction. PCA of the full 180-dimensional deviation profiles showed strong overlap between HCs and IBS participants in the first two principal components, indicating no clear multiregional separation of groups. Analyses of symptom−extreme-deviation associations likewise did not reveal robust IBS-specific effects. No between-group association-difference test survived FDR correction, and the largest absolute effect size was negligible ($|q| = 0.077$; minimum FDR-corrected $p = 0.739$). Overall, these exploratory NM-specific analyses did not reveal evidence of increased abnormality burden, multiregional deviation-profile structure, or symptom-related deviation patterns that would alter the primary conclusion of negligible IBS−HC cortical macrostructural differences.

3.6 IBS subtype sensitivity analyses

IBS subtype sensitivity analyses were conducted to examine whether IBS subtype heterogeneity could have masked an overall IBS-HC effect. Across CT, SA, and CV, no five-level group omnibus effect, five-level group × sex interaction, nested subtype heterogeneity effect, or nested subtype × sex heterogeneity effect survived FDR correction within modality (S13 Table). These findings suggest that no robust subtype-specific cortical deviation patterns were detected

within the available DHQ-derived subtype labels.

3.7 Consistency between NM-based and conventional case-control analyses

Raw-measure sensitivity analyses converged with the primary NM-based findings. After adjustment for age and scanner site, conventional raw-measure GLMs identified no FDR-corrected significant IBS-HC main effects and no FDR-corrected significant group × sex interaction effects for CT, SA, or CV. Effect sizes estimated from raw-measure models were highly concordant with those from the NM-based deviation models, with high modality-specific correlations for both the main group effect (CT: r = 0.969; SA: r = 0.970; CV: r = 0.974) and the group × sex interaction effect (CT: r = 0.980; SA: r = 0.990; CV: r = 0.989; all $p < 0.001$). These results indicate that the case-control findings were not specific to the NM pipeline, while supporting NM as the primary framework for individualized age-, sex-, and site-adjusted deviation estimation. Full raw-measure model outputs and the raw-vs-NM effect size comparison are provided in the supplementary materials (S14 Table and S7 Fig).

4 Discussion

In this population-based study, we investigated structural brain deviations in IBS using NM and equivalence testing applied to UKB data. Despite previous reports of morphological brain differences in IBS (Barazanji et al., 2022; Chua et al., 2017; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2015, 2014; Li et al., 2020; Piché et al., 2013; Seminowicz et al., 2010; Weaver et al., 2016), we found no group-level differences in cortical measures between IBS

and matched HCs. In contrast, equivalence tests showed that effect sizes were within a range that can be considered statistically negligible or smaller than the neuroimaging-informed relevance threshold for nearly all ROIs (Fig 2, S8 Table), providing evidence for the absence of meaningful group differences, rather than merely a lack of significant findings (Alderson, 2004; Altman and Bland, 1995; Sedgwick, 2014). This approach allowed small, likely negligible differences to be statistically rejected, providing evidence for the absence of robust cortical macrostructural alterations in this UKB IBS phenotype. Our findings align with recent failed large-scale replication efforts in psychiatric and behavioral neuroscience, which have shown that many previously claimed brain-behavior associations fail to generalize when examined in adequately powered samples (Kharabian Masouleh et al., 2019; Flint et al., 2021; Marek et al., 2022; Winter et al., 2022; Liu et al., 2023; Belov et al., 2024; Goltermann et al., 2025; Marek and Laumann, 2025; Goya-Maldonado et al., 2026).

Prior structural MRI studies of IBS have frequently described alterations in regions involved in pain perception and emotional regulation, particularly the anterior insula, cingulate cortex, and somatosensory cortices (Chua et al., 2017; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2015; Li et al., 2020). However, reported effects across these studies have been partially inconsistent with regard to both the direction and magnitude of their effect sizes. Further, these studies have included mostly small sample sizes and single-site datasets. Notably, the sample size used here was more than ten times larger than those in prior studies and yet failed to detect significant group-level differences in cortical morphometry. In contrast, effect size estimates reported here were consistently small to negligible and most could be conclusively rejected

with inferential equivalence tests. Equivalence test results therefore allowed us to infer that most detected effects can be considered practically equivalent to effects below a prespecified relevance threshold (Fig 2, S8 Table). Null findings are informative only when supported by sufficient statistical power and appropriate analytical frameworks (Stein et al., 2024). Accordingly, by combining equivalence test with the high statistical sensitivity that was provided by the UKB, we were able to yield inferential evidence beyond traditional null-hypothesis testing (Campbell, 2024; Gerchen et al., 2021; Lakens, 2017; Lakens et al., 2018), showing that any average structural difference reported here between IBS and HCs is negligible relative to inter-individual variability. Importantly, subtype sensitivity analyses directly addressed the possibility that aggregating IBS-C, IBS-D, IBS-M, and IBS-U could mask subtype-specific cortical effects. These analyses found no significant group omnibus effects, group × sex interactions, nested subtype heterogeneity effects, or subtype × sex heterogeneity effects. Thus, within the available DHQ-derived subtype classification, subtype heterogeneity did not explain the absence of robust average IBS-HC cortical morphometry differences. Further, for the few instances where equivalence tests remained inconclusive (mainly for regions within somatosensory and cingulate cortices), effect size point estimates were still minimal (< 0.07). These descriptive findings thus further support our interpretation that small effects in the current study likely reflect residual noise rather than relevant, interpretable group differences. These estimates therefore contrast sharply to previously reported effect sizes, showing that very large sample sizes would be required to detect these reliably (Liu et al., 2023; Marek et al., 2022).

While several cortical regions across brain metrics showed small associations between cortical gray matter deviations and IBS symptom severity or pain severity within the IBS group (S3 Fig), these did not appear to be IBS-specific when compared to HCs (Fig 3, S4 Fig). For instance, IBS cases showed slightly stronger negative correlations between CV and SA deviations and both symptom severity and somatic pain severity ($-0.123 < \text{Cohen's } q < -0.11$), primarily within postcentral and parietal cortices (Fig 3.a). These group effects were statistically significant and not equivalent to the defined SESOI. However, their confidence intervals overlapped with the equivalence bounds and they can thus be considered statistically negligible (Goertzen and Cribbie, 2010; Walker and Nowacki, 2011). Hence, these findings only partially replicate the direction of earlier reports that have linked larger symptom burden to cortical gray matter morphological deviations in somatosensory and motor regions (Labus et al., 2015) and suggest that any symptom-related effects are subtle and not interpretable as IBS-specific. Moreover, exploratory DiD analyses found no statistically relevant sex differences for cortical brain-phenotype associations and most effect sizes ranged between 0 and 0.1. Follow-up equivalence tests provided conclusive evidence that most effects were statistically negligible (S12 Table, S4 Fig). These findings contrast earlier IBS studies that reported female-specific or sex-dependent neural findings based on small or female-dominant samples (Jiang et al., 2013; Jin et al., 2019). Taken together, these findings demonstrate that sex-specific cortical brain-phenotype associations in IBS are minimal once the sample size is sufficiently large to provide precise, reliable effect size estimates.

The consistency between conventional raw-measure analyses and NM-based deviation analyses

further supports the robustness of these findings. Raw-measure GLMs adjusted for age and scanner site showed no FDR-corrected IBS-HC group effects or group × sex interactions, and effect sizes were highly concordant with those from the NM-based analyses. This high concordance should not be interpreted as indicating that NM is unnecessary. Rather, it suggests that the present case-control conclusion was not specific to the NM pipeline. NM remains advantageous because it provides individualized deviation scores relative to age-, sex-, and site-conditioned healthy reference trajectories, thereby supporting analyses of inter-individual heterogeneity even when average group differences are minimal. In this sense, the raw-measure comparison bridges conventional case-control inference and future NM-based individualized analyses.

Overall, our results provide conclusive evidence for the absence of statistically relevant group effects and interaction effects in brain morphology for most hypotheses and show only limited evidence for group-specific brain phenotype associations. Moreover, this large-scale investigation provides precise effect size estimates, which are consistently small to negligible (< 0.1) across contrasts. These findings are at substantial odds with many earlier reports in IBS neuroimaging (Barazanji et al., 2022; Chua et al., 2017; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2015, 2014; Li et al., 2020; Piché et al., 2013; Seminowicz et al., 2010; Weaver et al., 2016). A plausible explanation for these discrepancies is that previous reports have been dominated by overestimations. These likely stem from a combination of several factors, including limited sample sizes, analytical variability, and publication bias, all of which characterize much of early neuroimaging research (Algermissen and Mehler, 2018; Button et

al., 2013; Liu et al., 2023; Marek et al., 2022). Indeed, previous reviews of neuroimaging studies in IBS indicate that most early investigations included fewer than 30 individuals, rendering their estimates vulnerable to sampling noise and overfitting (Mayer et al., 2023, 2015; Nisticò et al., 2022; Weaver et al., 2016; Zhao et al., 2023). Small studies have been shown to be at high risk of yielding inflated estimates of brain-behavior associations and false positives due to low power and selective reporting (Algermissen and Mehler, 2018; Button et al., 2013; Liu et al., 2023; Marek et al., 2022), as well as type S (Sign) and type M (Magnitude) errors (Gelman and Carlin, 2014). These statistical considerations can further explain inconsistencies among previous IBS neuroimaging reports. In consequence, large-scale replication attempts of these initially promising findings in samples with sufficient statistical sensitivity can test whether many previously reported abnormalities were driven more by sampling noise than by replicable IBS-related alterations. This also constitutes the main contribution of the present study, which to our knowledge is the first such attempt in the context of structural gut-brain neuroimaging research.

While there is broad consensus in clinical neuroimaging research fields for the need of large samples to gain sufficient statistical sensitivity (Algermissen and Mehler, 2018; Button et al., 2013; Nord et al., 2017; Szucs and Ioannidis, 2017), recent critiques directed at research using large-scale datasets such as UKB have pointed out risks. These include concerns about oversimplified, confounded phenotyping, selective reporting, and insufficient correction for multiple testing, which increase the chances for inflated, false positive findings (Rudan et al., 2025; Spick et al., 2025; Suchak et al., 2025). The current study addressed and prevented these

risks in several ways: To provide sufficient transparency and follow current best practice recommendations, we adopted the GRABDROP checklist for analyses of big data repositories (S1 File). Further, statistical concerns were addressed by applying robust normative modeling to account for confounders through deviations from healthy cohorts, adopting a transparent methodology with inclusion and exclusion criteria (S1-S3 Tables), and implementing strict false discovery correction across all analyses. Crucially, by employing equivalence testing, we moved beyond conventional null-hypothesis inference and used the statistical power of the available UKB dataset to provide evidence for the absence (i.e. null hypothesis) of the most inconclusive primary test results. We see this as one advantage of replication attempts in large databases and hence encourage others to combine NHST and equivalence testing and thereby fully capitalize on the high statistical sensitivity that large data sets entail, enabling researchers to provide both evidence for the presence and absence of (meaningful) effects (Stein et al., 2024).

Although this work addresses several limitations of previous studies, further aspects should be considered in the interpretation of our findings and conceptualization of future work. While UKB provided high statistical sensitivity to investigate morphological variations in IBS, IBS status was based on self-reported DHQ items mapped to Rome III symptom logic, without gastroenterologist confirmation, Rome IV reassessment, disease-duration information, detailed treatment status, or full characterization of comorbid disorders. Residual phenotypic heterogeneity therefore cannot be fully excluded and may still attenuate subtle morphometric effects. Subtype sensitivity analyses were also limited by the available DHQ-derived subtype

labels and unequal subtype sample sizes, particularly the small IBS-U group. Splitting the IBS sample into subtypes reduced statistical power for subtype-specific effects compared with the main IBS-HC analyses. Although these analyses did not detect subtype-specific cortical patterns, they do not establish equivalence across IBS subtypes. More deeply phenotyped and clinically confirmed cohorts may be needed to test whether subtype-specific neural alterations emerge under more precise diagnostic, symptom-state, or treatment-defined classifications. Further, questionnaire and MRI assessments were not always contemporaneous. Although the DHQ-MRI interval was within 3 years for most participants, questionnaire-derived phenotype and symptom measures may not perfectly reflect clinical status at imaging, which could attenuate brain-phenotype associations. This concern may be less pronounced for morphometric brain measures, because these macrostructural measures are relatively stable, but temporal mismatch remains an important limitation. In addition, the UKB cohort is not representative of the entire general population, consisting primarily of middle-aged and older adults from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Henrich et al., 2010; Rutherford et al., 2022a). This potential sampling bias may limit the generalizability of our findings as IBS prevalence, symptom presentation, and brain-gut interactions may differ across populations (Francisconi et al., 2016; Gerson and Gerson, 2010). Finally, our investigation was limited to gray matter structural deviations in cortical thickness, volume, and surface area. Therefore, the present findings should not be interpreted as evidence against central nervous system involvement in IBS or against the broader BGM framework. Structural MRI captures relatively stable macrostructural features, whereas IBS-related neural alterations may be functional, network-level, task- or symptom-state-dependent, stress-related, immune-

or microbiome-mediated, or longitudinal. Even within brain morphology, whether our findings also extend to other features such as vertex-wise cortical measures, subcortical regions, and white matter microstructure remains to be investigated. Previous IBS studies (Chen et al., 2021; Mayer et al., 2023, 2015) have also highlighted aberrations in functional brain networks in IBS. Functional imaging data, brain network metrics, and multimodal neuroimaging approaches could provide more sensitive measures to capture deviations related to concurrent symptomatology, compensatory mechanisms, or markers for clinical trajectories (Marquand et al., 2024). Future large-scale, multimodal studies combining structural and functional imaging with longitudinal symptom assessment, microbiome measures, inflammatory markers, and more detailed clinical phenotyping may be better positioned to identify reliable and replicable neural markers of IBS and its subtypes.

5 Conclusions

In conclusion, we found no replicable group-level differences in cortical gray matter morphometry between IBS and HCs, nor any meaningful interactions with sex, after correction for multiple comparisons. Importantly, equivalence testing confirmed that nearly all regional effects were statistically negligible. Only a few cortical regions, including regions in somatosensory and cingulate areas, exhibited inconclusive effects close to the equivalence bounds. Subtype analyses also did not identify robust IBS subtype-specific cortical patterns. Although negative brain-phenotype associations were detected between cortical deviations and IBS symptom severity and pain severity across regions within the IBS group, no meaningful

IBS-specific differences relative to HCs were observed, with only a few regions showing inconclusive effects. In addition, no meaningful sex moderation was observed. By integrating normative modeling with equivalence testing in a large population-based cohort, this study fails to replicate prior reports of IBS-related cortical morphometric alterations, indicating that previously reported structural neuroimaging findings in IBS (Barazanji et al., 2022; Grinsvall et al., 2021; Jiang et al., 2013; Labus et al., 2014; Piché et al., 2013; Seminowicz et al., 2010) likely reflect inflated or sample-specific effects (Button et al., 2013; Marek et al., 2022; Mehler, 2019). Overall, our study provides rigorous statistical evidence that morphometric differences in IBS are likely very small under the present UKB phenotype and modeling framework. We therefore encourage future replication studies using adequately powered, clinically well-characterized datasets and multimodal frameworks to build a more consistent evidence base for clinically meaningful neural markers of IBS.

Data Availability

This study has been conducted using the UK Biobank Resource under Application 71300. All raw data are available from the UKB (www.ukbiobank.ac.uk). The raw and processed UKB MRI data are protected and are not openly available due to data privacy laws. Access can be obtained by applying for access (www.ukbiobank.ac.uk/enable-your-research/apply-for-access).

Code Availability

All software used is publicly available at the URL or references cited. Code supporting the findings of this study is available at https://github.com/hanbing-gao/NM_IBS.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 71300. UK Biobank data were accessed by H.G., C.V.S. and K.M.S. Copyright © 2025, NHS England. Reused with the permission of the NHS England and/or UK Biobank. All rights reserved. This work uses data provided by patients and collected by the NHS as part of their care and support.

Author contributions

H.G. contributed conceptualization, methodology, software, validation, formal analysis, investigation, writing-original draft, writing-review & editing, visualization; C.V.S. contributed conceptualization, resources, writing-review & editing; T.S. contributed methodology, validation, writing-review & editing; C.Z. contributed methodology, writing-review & editing, supervision, funding acquisition; F.C. contributed methodology, writing-review & editing, funding acquisition; K.M. contributed methodology, resources, writing-review & editing, supervision; K.M.S. contributed conceptualization, resources, writing-review & editing; D.M.A.M. contributed conceptualization, methodology, investigation, resources, writing-

review & editing, supervision, project administration, funding acquisition.

References

Alderson, P., 2004. Absence of evidence is not evidence of absence. BMJ 328, 476. https://doi.org/10.1136/bmj.328.7438.476

Alfaro-Almagro, F., Jenkinson, M., Bangerter, N.K., Andersson, J.L.R., Griffanti, L., Douaud, G., Sotiropoulos, S.N., Jbabdi, S., Hernandez-Fernandez, M., Vallee, E., Vidaurre, D., Webster, M., McCarthy, P., Rorden, C., Daducci, A., Alexander, D.C., Zhang, H., Dragonu, I., Matthews, P.M., Miller, K.L., Smith, S.M., 2018. Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. NeuroImage 166, 400–424. https://doi.org/10.1016/j.neuroimage.2017.10.034

Algermissen, J., Mehler, D.M.A., 2018. May the power be with you: are there highly powered studies in neuroscience, and how can we get more of them? Journal of Neurophysiology 119, 2114–2117. https://doi.org/10.1152/jn.00765.2017

Altman, D.G., Bland, J.M., 1995. Statistics notes: Absence of evidence is not evidence of absence. BMJ 311, 485. https://doi.org/10.1136/bmj.311.7003.485

Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., Sant'Anna, P.H.C., 2025. Difference-in-Differences Designs: A Practitioner's Guide. https://doi.org/10.48550/arXiv.2503.13323

Barazanji, N., Paul Hamilton, J., Icenhour, A., Simon, R.A., Bednarska, O., Tapper, S., Tisell, A., Lundberg, P., Engström, M., Walter, S., 2022. Irritable bowel syndrome in women: Association between decreased insular subregion volumes and gastrointestinal symptoms. NeuroImage: Clinical 35, 103128. https://doi.org/10.1016/j.nicl.2022.103128

Bayer, J.M.M., Van Velzen, L.S., Pozzi, E., Davey, C., Han, L.K.M., Bauduin, S.E.E.C., Bauer, J., Benedetti, F., Berger, K., Bonnekoh, L.M., Brosch, K., Buelow, R., Couvy-Duchesne, B., Cullen, K.R., Dannlowski, U., Dima, D., Dohm, K., Evans, J.W., Fu, C.H.Y., Fuentes-Claramonte, P., Godlewska, B.R., Goltermann, J., Gonul, A.S., Gotlib, I.H., Goya-Maldonado, R., Grabe, H.J., Groenewold, N.A., Grotegerd, D., Gruber, O., Hahn, T., Hall, G.B., Hamilton, J.P., Harrison, B.J., Hatton, S.N., Hermesdorf, M., Hickie, I.B., Ho, T.C., Jahanshad, N., Jamieson, A.J., Jansen, A., Kamishikiryo, T., Kircher, T.T.J., Klimes-Dougan, B., Kraemer, B., Kraus, A., Krug, A., Leehr, E.J., Leenings, R., Li, M., McIntosh, A., Medland, S.E., Meinert, S., Melloni, E., Mwangi, B., Nenadic, I., Okada, G., Oudega, M.L., Portella, M.J., Rodriguez, E., Romaniuk, L., Rosa, P.G.P., Sacchet, M.D., Salvador, R., Saemann, P.G., Shinzato, H., Sim, K., Simulionyte, E., Soares, J.C.,

Stein, F., Stein, D.J., Stolicyn, A., Straube, B., Strike, L.T., Teutenberg, L., Thomas-Odenthal, F., Thomopoulos, S.I., Usemann, P., Van Der Wee, N.J.A., Voelzke, H., Wagenmakers, M., Walter, M., Whalley, H.C., Whittle, S., Winter, N.R., Wittfeld, K., Wu, M.-J., Yang, T.T., Zarate, C.A., Zunta-Soares, G.B., Thompson, P.M., Veltman, D.J., Marquand, A.F., Schmaal, L., 2025. Dissecting heterogeneity in cortical thickness abnormalities in major depressive disorder: a large-scale ENIGMA MDD normative modelling study. https://doi.org/10.1101/2025.03.17.643677

Belov, V., Erwin-Grabner, T., Aghajani, M., Aleman, A., Amod, A.R., Basgoze, Z., Benedetti, F., Besteher, B., Bülow, R., Ching, C.R.K., Connolly, C.G., Cullen, K., Davey, C.G., Dima, D., Dols, A., Evans, J.W., Fu, C.H.Y., Gonul, A.S., Gotlib, I.H., Grabe, H.J., Groenewold, N., Hamilton, J.P., Harrison, B.J., Ho, T.C., Mwangi, B., Jaworska, N., Jahanshad, N., Klimes-Dougan, B., Koopowitz, S.-M., Lancaster, T., Li, M., Linden, D.E.J., MacMaster, F.P., Mehler, D.M.A., Melloni, E., Mueller, B.A., Ojha, A., Oudega, M.L., Penninx, B.W.J.H., Poletti, S., Pomarol-Clotet, E., Portella, M.J., Pozzi, E., Reneman, L., Sacchet, M.D., Sämann, P.G., Schrantee, A., Sim, K., Soares, J.C., Stein, D.J., Thomopoulos, S.I., Uyar-Demir, A., van der Wee, N.J.A., van der Werff, S.J.A., Völzke, H., Whittle, S., Wittfeld, K., Wright, M.J., Wu, M.-J., Yang, T.T., Zarate, C., Veltman, D.J., Schmaal, L., Thompson, P.M., Goya-Maldonado, R., 2024. Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. Sci Rep 14, 1084. https://doi.org/10.1038/s41598-023-47934-8

Benjamini, Y., Hochberg, Y., 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) 57, 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bethlehem, R.A.I., Seidlitz, J., Romero-Garcia, R., Trakoshis, S., Dumas, G., Lombardo, M.V., 2020. A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. Commun Biol 3, 486. https://doi.org/10.1038/s42003-020-01212-9

Bodner, T.E., 2017. Standardized Effect Sizes for Moderated Conditional Fixed Effects with Continuous Moderator Variables. Front. Psychol. 8, 562. https://doi.org/10.3389/fpsyg.2017.00562

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14, 365–376. https://doi.org/10.1038/nrn3475

Campbell, H., 2024. Equivalence testing for linear regression. Psychological Methods 29, 88–98. https://doi.org/10.1037/met0000596

Chen, N., Liu, G., Guo, M., Li, Y., Yao, Z., Hu, B., 2021. Calcarine as a bridge between brain function and structure in irritable bowel syndrome: A multiplex network analysis. Journal of Gastroenterology and Hepatology 36,

2408–2415. https://doi.org/10.1111/jgh.15382

Chua, C.S., Bai, C.-H., Shiao, C.-Y., Hsu, C.-Y., Cheng, C.-W., Yang, K.-C., Chiu, H.-W., Hsu, J.-L., 2017. Negative correlation of cortical thickness with the severity and duration of abdominal pain in Asian women with irritable bowel syndrome. PLOS ONE 12, e0183960. https://doi.org/10.1371/journal.pone.0183960

Cohen, J., 2013. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Routledge, New York. https://doi.org/10.4324/9780203771587

Destrieux, C., Fischl, B., Dale, A., Halgren, E., 2010. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. NeuroImage 53, 1–15. https://doi.org/10.1016/j.neuroimage.2010.06.010

Douglas A. Drossman, 2006. The Functional Gastrointestinal Disorders and the Rome III Process. Gastroenterology 130, 1377–1390. https://doi.org/10.1053/j.gastro.2006.03.008

Drossman, D.A., Li, Z., Andruzzi, E., Temple, R.D., Talley, N.J., Grant Thompson, W., Whitehead, W.E., Janssens, J., Funch-Jensen, P., Corazziari, E., Richter, J.E., Koch, G.G., 1993. U. S. Householder survey of functional gastrointestinal disorders. Digest Dis Sci 38, 1569–1580. https://doi.org/10.1007/BF01303162

Eijsbouts, C., Zheng, T., Kennedy, N.A., Bonfiglio, F., Anderson, C.A., Moutsianas, L., Holliday, J., Shi, J., Shringarpure, S., Agee, M., Aslibekyan, S., Auton, A., Bell, R.K., Bryc, K., Clark, S.K., Elson, S.L., Fletez-Brant, K., Fontanillas, P., Furlotte, N.A., Gandhi, P.M., Heilbron, K., Hicks, B., Hinds, D.A., Huber, K.E., Jewett, E.M., Jiang, Y., Kleinman, A., Lin, K.H., Litterman, N.K., Luff, M.K., McCreight, J.C., McIntyre, M.H., McManus, K.F., Mountain, J.L., Mozaffari, S.V., Nandakumar, P., Noblin, E.S., Northover, C.A.M., O'Connell, J., Petrakovitz, A.A., Pitts, S.J., Poznik, G.D., Sathirapongsasuti, J.F., Shastri, A.J., Shelton, J.F., Tian, C., Tung, J.Y., Tunney, R.J., Vacic, V., Wang, X., Zare, A.S., Voda, A.I., Kashyap, P., Chang, L., Mayer, E., Heitkemper, M., Sayuk, G.S., Ringel-Kulka, T., Ringel, Y., Chey, W.D., Eswaran, S., Merchant, J.L., Shulman, R.J., Bujanda, L., Garcia-Etxebarria, K., Dlugosz, A., Lindberg, G., Schmidt, P.T., Karling, P., Ohlsson, B., Walter, S., Faresjö, Å.O., Simren, M., Halfvarson, J., Portincasa, P., Barbara, G., Usai-Satta, P., Neri, M., Nardone, G., Cuomo, R., Galeazzi, F., Bellini, M., Latiano, A., Houghton, L., Jonkers, D., Kurilshikov, A., Weersma, R.K., Netea, M., Tesarz, J., Gauss, A., Goebel-Stengel, M., Andresen, V., Frieling, T., Pehl, C., Schaefert, R., Niesler, B., Lieb, W., Hanevik, K., Langeland, N., Wensaas, K.A., Litleskare, S., Gabrielsen, M.E., Thomas, L., Thijs, V., Lemmens, R., Van Oudenhove, L., Wouters, M., Farrugia, G., Franke, A., Hübenthal, M., Abecasis, G., Zawistowski, M., Skogholt, A.H., Ness-Jensen, E., Hveem, K., Esko, T., Teder-Laving, M., Zhernakova, A., Camilleri, M., Boeckxstaens, G., Whorwell, P.J., Spiller, R., McVean, G., D'Amato, M., Jostins, L., Parkes, M., 2021. Genome-wide analysis of 53,400 people with

irritable bowel syndrome highlights shared genetic pathways with mood and anxiety disorders. Nature Genetics 53, 1543–1552. https://doi.org/10.1038/s41588-021-00950-8

Elad, D., Cetin-Karayumak, S., Zhang, F., Cho, K.I.K., Lyall, A.E., Seitz-Holland, J., Ben-Ari, R., Pearlson, G.D., Tamminga, C.A., Sweeney, J.A., Clementz, B.A., Schretlen, D.J., Viher, P.V., Stegmayer, K., Walther, S., Lee, J., Crow, T.J., James, A., Voineskos, A.N., Buchanan, R.W., Szeszko, P.R., Malhotra, A.K., Keshavan, M.S., Shenton, M.E., Rathi, Y., Bouix, S., Sochen, N., Kubicki, M.R., Pasternak, O., 2021. Improving the predictive potential of diffusion MRI in schizophrenia using normative models—Towards subject-level classification. Human Brain Mapping 42, 4658–4670. https://doi.org/10.1002/hbm.25574

Flint, C., Cearns, M., Opel, N., Redlich, R., Mehler, D.M.A., Emden, D., Winter, N.R., Leenings, R., Eickhoff, S.B., Kircher, T., Krug, A., Nenadic, I., Arolt, V., Clark, S., Baune, B.T., Jiang, X., Dannlowski, U., Hahn, T., 2021. Systematic misestimation of machine learning performance in neuroimaging studies of depression. Neuropsychopharmacology 46, 1510–1517. https://doi.org/10.1038/s41386-021-01020-7

Fond, G., Loundou, A., Hamdani, N., Boukouaci, W., Dargel, A., Oliveira, J., Roger, M., Tamouza, R., Leboyer, M., Boyer, L., 2014. Anxiety and depression comorbidities in irritable bowel syndrome (IBS): a systematic review and meta-analysis. Eur Arch Psychiatry Clin Neurosci 264, 651–660. https://doi.org/10.1007/s00406-014-0502-z

Francis, C.Y., Morris, J., Whorwell, P.J., 1997. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 11, 395–402. https://doi.org/10.1046/j.1365-2036.1997.142318000.x

Francisconi, C.F., Sperber, A.D., Fang, X., Fukudo, S., Gerson, M.-J., Kang, J.-Y., Schmulson, M., 2016. Multicultural Aspects in Functional Gastrointestinal Disorders (FGIDs). Gastroenterology, Rome IV - Functional GI Disorders: Disorders of Gut-Brain Interaction 150, 1344-1354.e2. https://doi.org/10.1053/j.gastro.2016.02.013

Fraza, C.J., Dinga, R., Beckmann, C.F., Marquand, A.F., 2021. Warped Bayesian linear regression for normative modelling of big data. NeuroImage 245, 118715. https://doi.org/10.1016/j.neuroimage.2021.118715

Funder, D.C., Ozer, D.J., 2019. Evaluating Effect Size in Psychological Research: Sense and Nonsense. Advances in Methods and Practices in Psychological Science 2, 156–168. https://doi.org/10.1177/2515245919847202

Gelman, A., Carlin, J., 2014. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspect Psychol Sci 9, 641–651. https://doi.org/10.1177/1745691614551642

Gerchen, M.F., Kirsch, P., Feld, G.B., 2021. Brain-wide inferiority and equivalence tests in fMRI group analyses: Selected applications. Human Brain Mapping 42, 5803–5813. https://doi.org/10.1002/hbm.25664

Gerson, C.D., Gerson, M.-J., 2010. A Cross-Cultural Perspective on Irritable Bowel Syndrome. Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine 77, 707–712. https://doi.org/10.1002/msj.20218

Goertzen, J.R., Cribbie, R.A., 2010. Detecting a lack of association: An equivalence testing approach. British Journal of Mathematical and Statistical Psychology 63, 527–537. https://doi.org/10.1348/000711009X475853

Goltermann, J., Winter, N.R., Meinert, S., Grotegerd, D., Kraus, A., Flinkenflügel, K., Altegoer, L., Krieger, J., Leehr, E.J., Böhnlein, J., Bonnekoh, L.M., Richter, M., Hahn, T., Fisch, L., Gruber, M., Hermesdorf, M., Berger, K., Arolt, V., Brosch, K., Stein, F., Thomas-Odenthal, F., Usemann, P., Teutenberg, L., Hammes, V., Jamalabadi, H., Alexander, N., Straube, B., Jansen, A., Nenadić, I., Kircher, T., Opel, N., Dannlowski, U., 2025. Gray matter correlates of childhood maltreatment lack replicability in a multi-cohort brain-wide association study. Nat Commun 16, 8290. https://doi.org/10.1038/s41467-025-62374-w

Goya-Maldonado, R., Erwin-Grabner, T., Zeng, L.-L., Ching, C.R.K., Aleman, A., Amod, A.R., Basgoze, Z., Benedetti, F., Besteher, B., Brosch, K., Bülow, R., Colle, R., Connolly, C.G., Corruble, E., Couvy-Duchesne, B., Cullen, K., Dannlowski, U., Davey, C.G., Dols, A., Ernsting, J., Evans, J.W., Fisch, L., Fuentes-Claramonte, P., Gonul, A.S., Gotlib, I.H., Grabe, H.J., Groenewold, N.A., Grotegerd, D., Hahn, T., Hamilton, J.P., Han, L.K.M., Harrison, B.J., Ho, T.C., Jahanshad, N., Jamieson, A.J., Karuk, A., Kircher, T., Klimes-Dougan, B., Koopowitz, S.-M., Lancaster, T., Leenings, R., Li, M., Linden, D.E.J., MacMaster, F.P., Mehler, D.M.A., Meinert, S., Melloni, E., Mueller, B.A., Mwangi, B., Nenadić, I., Ojha, A., Okamoto, Y., Oudega, M.L., Penninx, B.W.J.H., Poletti, S., Pomarol-Clotet, E., Portella, M.J., Radua, J., Rodríguez-Cano, E., Sacchet, M.D., Salvador, R., Schrantee, A., Sim, K., Soares, J.C., Solanes, A., Stein, D.J., Stein, F., Stolicyn, A., Thomopoulos, S.I., Toenders, Y.J., Uyar-Demir, A., Vieta, E., Vives-Gilabert, Y., Völzke, H., Walter, M., Whalley, H.C., Whittle, S., Winter, N., Wittfeld, K., Wright, M.J., Wu, M.-J., Yang, T.T., Zarate, C., Veltman, D.J., Schmaal, L., Thompson, P.M., 2026. Classification of major depressive disorder using vertex-wise brain sulcal depth, curvature, and thickness with a deep and a shallow learning model. Mol Psychiatry 31, 1517–1529. https://doi.org/10.1038/s41380-025-03273-w

Grinsvall, C., Ryu, H.J., Van Oudenhove, L., Labus, J.S., Gupta, A., Ljungberg, M., Törnblom, H., Mayer, E.A., Simrén, M., 2021. Association between pain sensitivity and gray matter properties in the sensorimotor network in women with irritable bowel syndrome. Neurogastroenterology & Motility 33, e14027. https://doi.org/10.1111/nmo.14027

Gupta, A., Mayer, E.A., Fling, C., Labus, J.S., Naliboff, B.D., Hong, J., Kilpatrick, L.A., 2017. Sex-based differences in brain alterations across

chronic pain conditions. J of Neuroscience Research 95, 604–616. https://doi.org/10.1002/jnr.23856

Hanel, P.H., Mehler, D.M., 2019. Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. Public Underst Sci 28, 468–485. https://doi.org/10.1177/0963662519834193

Hedges, L.V., Tipton, E., Zejnullahi, R., Diaz, K.G., 2023. Effect sizes in ANCOVA and difference-in-differences designs. British Journal of Mathematical and Statistical Psychology 76, 259–282. https://doi.org/10.1111/bmsp.12296

Henrich, J., Heine, S.J., Norenzayan, A., 2010. The weirdest people in the world? Behavioral and Brain Sciences 33, 61–83. https://doi.org/10.1017/S0140525X0999152X

Jiang, Z., Dinov, I.D., Labus, J., Shi, Y., Zamanyan, A., Gupta, A., Ashe-McNalley, C., Hong, J.-Y., Tillisch, K., Toga, A.W., Mayer, E.A., 2013. Sex-Related Differences of Cortical Thickness in Patients with Chronic Abdominal Pain. PLoS ONE 8, e73932. https://doi.org/10.1371/journal.pone.0073932

Jin, Q., Duan, S., Li, G., Sun, L., Hu, Y., Hu, C., Zhao, J., von Deneen, K.M., Qian, L., Wang, H., Ji, G., Wu, K., Fan, D., Cui, G., Nie, Y., Zhang, Y., 2019. Sex-related differences in resting-state brain activity and connectivity in the orbital frontal cortex and insula in patients with functional constipation. Neurogastroenterology & Motility 31, e13566. https://doi.org/10.1111/nmo.13566

Kharabian Masouleh, S., Eickhoff, S.B., Hoffstaedter, F., Genon, S., Alzheimer's Disease Neuroimaging Initiative, 2019. Empirical examination of the replicability of associations between brain structure and psychological variables. eLife 8, e43464. https://doi.org/10.7554/eLife.43464

Kline, A., Luo, Y., 2022. PsmPy: A Package for Retrospective Cohort Matching in Python, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Presented at the 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, Glasgow, Scotland, United Kingdom, pp. 1354–1357. https://doi.org/10.1109/EMBC48229.2022.9871333

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med 16, 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Labus, J.S., Dinov, I.D., Jiang, Z., Ashe-McNalley, C., Zamanyan, A., Shi, Y., Hong, J.-Y., Gupta, A., Tillisch, K., Ebrat, B., Hobel, S., Gutman, B.A., Joshi, S., Thompson, P.M., Toga, A.W., Mayer, E.A., 2014. Irritable bowel syndrome in female patients is associated with alterations in structural brain networks. PAIN 155, 137. https://doi.org/10.1016/j.pain.2013.09.020

Labus, J.S., Van Horn, J.D., Gupta, A., Alaverdyan, M., Torgerson, C., Ashe-McNalley, C., Irimia, A., Hong, J.-Y., Naliboff, B., Tillisch, K., Mayer, E.A., 2015. Multivariate morphological brain signatures predict patients

with chronic abdominal pain from healthy control subjects. PAIN 156, 1545. https://doi.org/10.1097/j.pain.0000000000000196

Lakens, D., 2017. Equivalence Tests. Social Psychological and Personality Science 8, 355–362. https://doi.org/10.1177/1948550617697177

Lakens, D., Scheel, A.M., Isager, P.M., 2018. Equivalence Testing for Psychological Research: A Tutorial. Advances in Methods and Practices in Psychological Science 1, 259–269. https://doi.org/10.1177/2515245918770963

Lennon, M.J., Thalamuthu, A., Lam, B.C.P., Crawford, J.D., Sachdev, P.S., 2023. Genetically Predicted Blood Pressure and Cognition in Midlife: A UK Biobank Study. Hypertension 80, 2112–2121. https://doi.org/10.1161/HYPERTENSIONAHA.123.21612

Li, J., Yuan, B., Li, G., Lu, X., Guo, Y., Yang, Y., Liang, M., Ding, J., Zhou, Q., 2020. Convergent syndromic atrophy of pain and emotional systems in patients with irritable bowel syndrome and depressive symptoms. Neuroscience Letters 723, 134865. https://doi.org/10.1016/j.neulet.2020.134865

Li, Z., Ma, Q., Deng, Y., Rolls, E.T., Shen, C., Li, Y., Zhang, W., Xiang, S., Langley, C., Sahakian, B.J., Robbins, T.W., Yu, J.-T., Feng, J., Cheng, W., 2024. Irritable Bowel Syndrome is Associated With Brain Health by Neuroimaging, Behavioral, Biochemical, and Genetic Analyses. Biological Psychiatry, Circuit Mechanisms and Mood Disorders 95, 1122–1132. https://doi.org/10.1016/j.biopsych.2023.12.024

Liu, S., Abdellaoui, A., Verweij, K.J.H., van Wingen, G.A., 2023. Replicable brain–phenotype associations require large-scale neuroimaging data. Nat Hum Behav 7, 1344–1356. https://doi.org/10.1038/s41562-023-01642-5

Marek, S., Laumann, T.O., 2025. Replicability and generalizability in population psychiatric neuroimaging. Neuropsychopharmacol. 50, 52–57. https://doi.org/10.1038/s41386-024-01960-w

Marek, S., Tervo-Clemmens, B., Calabro, F.J., Montez, D.F., Kay, B.P., Hatoum, A.S., Donohue, M.R., Foran, W., Miller, R.L., Hendrickson, T.J., Malone, S.M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A.M., Earl, E.A., Perrone, A.J., Cordova, M., Doyle, O., Moore, L.A., Conan, G.M., Uriarte, J., Snider, K., Lynch, B.J., Wilgenbusch, J.C., Pengo, T., Tam, A., Chen, J., Newbold, D.J., Zheng, A., Seider, N.A., Van, A.N., Metoki, A., Chauvin, R.J., Laumann, T.O., Greene, D.J., Petersen, S.E., Garavan, H., Thompson, W.K., Nichols, T.E., Yeo, B.T.T., Barch, D.M., Luna, B., Fair, D.A., Dosenbach, N.U.F., 2022. Reproducible brain-wide association studies require thousands of individuals. Nature 603, 654–660. https://doi.org/10.1038/s41586-022-04492-9

Marquand, A., Rutherford, S., Wolfers, T., 2024. Chapter 11 - Normative modeling for clinical neuroscience, in: Kay, K. (Ed.), Computational and Network Modeling of Neuroimaging Data, Neuroimaging Methods and Applications. Academic Press, pp. 309–329. https://doi.org/10.1016/B978-0-443-13480-

7.00014-4

Marquand, A.F., Rezek, I., Buitelaar, J., Beckmann, C.F., 2016. Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. Biological Psychiatry 80, 552–561. https://doi.org/10.1016/j.biopsych.2015.12.023

Mayer, E.A., Labus, J.S., Tillisch, K., Cole, S.W., Baldi, P., 2015. Towards a systems view of IBS. Nat Rev Gastroenterol Hepatol 12, 592–605. https://doi.org/10.1038/nrgastro.2015.121

Mayer, E.A., Ryu, H.J., Bhatt, R.R., 2023. The neurobiology of irritable bowel syndrome. Molecular Psychiatry, 无 28, 1451–1465. https://doi.org/10.1038/s41380-023-01972-w

McKinney, W., 2010. Data Structures for Statistical Computing in Python, in: Proceedings of the 9th Python in Science Conference. Presented at the Python in Science Conference, Austin, Texas, pp. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Mehler, D., 2019. 11 - The replication challenge: Is brain imaging next?, in: Raz, A., Thibault, R.T. (Eds.), Casting Light on the Dark Side of Brain Imaging. Academic Press, pp. 67–71.

Mehler, D.M.A., Edelsbrunner, P.A., Matić, K., 2019. Appreciating the Significance of Non-significant Findings in Psychology. J Eur Psychol Students 10, 1. https://doi.org/10.5334/e2019a

Ng, Q.X., Yau, C.E., Yaow, C.Y.L., Chong, R.I.H., Chong, N.Z.-Y., Teoh, S.E., Lim, Y.L., Soh, A.Y.S., Ng, W.K., Thumboo, J., 2023. What Has Longitudinal 'Omics' Studies Taught Us about Irritable Bowel Syndrome? A Systematic Review. Metabolites 13, 484. https://doi.org/10.3390/metabo13040484

Nisticò, V., Rossi, R.E., D'Arrigo, A.M., Priori, A., Gambini, O., Demartini, B., 2022. Functional Neuroimaging in Irritable Bowel Syndrome: A Systematic Review Highlights Common Brain Alterations With Functional Movement Disorders. J Neurogastroenterol Motil 28, 185–203. https://doi.org/10.5056/jnm21079

Nord, C.L., Valton, V., Wood, J., Roiser, J.P., 2017. Power-up: A Reanalysis of "Power Failure" in Neuroscience Using Mixture Modeling. J. Neurosci. 37, 8051–8061. https://doi.org/10.1523/JNEUROSCI.3592-16.2017

Oka, P., Parr, H., Barberio, B., Black, C.J., Savarino, E.V., Ford, A.C., 2020. Global prevalence of irritable bowel syndrome according to Rome III or IV criteria: a systematic review and meta-analysis. The Lancet Gastroenterology & Hepatology 5, 908–917. https://doi.org/10.1016/S2468-1253(20)30217-X

Orand, A., Gupta, A., Shih, W., Presson, A.P., Hammer, C., Niesler, B., Heendeniya, N., Mayer, E.A., Chang, L., 2015. Catecholaminergic Gene Polymorphisms Are Associated with GI Symptoms and Morphological Brain Changes in Irritable Bowel Syndrome. PLOS ONE 10, e0135910. https://doi.org/10.1371/journal.pone.0135910

Piché, M., Chen, J.-I., Roy, M., Poitras, P., Bouin, M., Rainville, P., 2013. Thicker Posterior Insula Is Associated With Disease Duration in Women With Irritable Bowel Syndrome (IBS) Whereas Thicker Orbitofrontal Cortex Predicts Reduced Pain Inhibition in Both IBS Patients and Controls. The Journal of Pain 14, 1217–1226. https://doi.org/10.1016/j.jpain.2013.05.009

Polster, A.V., Palsson, O.S., Törnblom, H., Öhman, L., Sperber, A.D., Whitehead, W.E., Simrén, M., 2019. Subgroups of IBS patients are characterized by specific, reproducible profiles of GI and non-GI symptoms and report differences in healthcare utilization: A population-based study. Neurogastroenterology & Motility 31, e13483. https://doi.org/10.1111/nmo.13483

Ramseyer, G.C., 1979. Testing the Difference between Dependent Correlations Using the Fisher Z. The Journal of Experimental Education 47, 307–310.

Razali, N.M., Wah, Y.B., 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. Journal of Statistical Modeling and Analytics 2, 21–33.

Rudan, I., Song, P., Adeloye, D., Campbell, H., 2025. Journal of Global Health's Guidelines for Reporting Analyses of Big Data Repositories Open to the Public (GRABDROP): preventing 'paper mills', duplicate publications, misuse of statistical inference, and inappropriate use of artificial intelligence. J Glob Health 15, 01004. https://doi.org/10.7189/jogh.15.01004

Rutherford, S., Barkema, P., Tso, I.F., Sripada, C., Beckmann, C.F., Ruhe, H.G., Marquand, A.F., 2023. Evidence for embracing normative modeling. eLife 12, 85082. https://doi.org/10.7554/eLife.85082

Rutherford, S., Fraza, C., Dinga, R., Mostafa Kia, S., Wolfers, T., Zabihi, M., Berthet, P., Worker, A., Verdi, S., Andrews, D., Han, L.K., Bayer, J.M., Dazzan, P., McGuire, P., Mocking, R.T., Schene, A., Sripada, C., Tso, I.F., Duval, E.R., Chang, S.-E., WJH Penninx, B., Heitzeg, M.M., Alexandra Burt, S., Hyde, L.W., Amaral, D., Wu Nordahl, C., Andreasssen, O.A., Westlye, L.T., Zahn, R., Ruhe, H.G., Beckmann, C., Marquand, A.F., 2022a. Charting brain growth and aging at high spatial precision. eLife 11, 72904. https://doi.org/10.7554/eLife.72904

Rutherford, S., Kia, S.M., Wolfers, T., Fraza, C., Zabihi, M., Dinga, R., Berthet, P., Worker, A., Verdi, S., Ruhe, H.G., Beckmann, C.F., Marquand, A.F., 2022b. The normative modeling framework for computational psychiatry. Nature Protocols 17, 1711–1734. https://doi.org/10.1038/s41596-022-00696-5

Salminen, L.E., Tubi, M.A., Bright, J., Thomopoulos, S.I., Wieand, A., Thompson, P.M., 2022. Sex is a defining feature of neuroimaging phenotypes in major brain disorders. Hum Brain Mapp 43, 500–542. https://doi.org/10.1002/hbm.25438

Schneider, C.V., Schneider, K.M., Raptis, A., Huang, H., Trautwein, C., Loomba,

R., 2024. Prevalence of at-risk MASH, MetALD and alcohol-associated steatotic liver disease in the general population. Alimentary Pharmacology & Therapeutics 59, 1271–1281. https://doi.org/10.1111/apt.17958

Schneider, K.M., Blank, N., Alvarez, Y., Thum, K., Lundgren, P., Litichevskiy, L., Sleeman, M., Bahnsen, K., Kim, J., Kardo, S., Patel, S., Dohnalová, L., Uhr, G.T., Descamps, H.C., Kircher, S., McSween, A.M., Ardabili, A.R., Nemec, K.M., Jimenez, M.T., Glotfelty, L.G., Eisenberg, J.D., Furth, E.E., Henao-Mejia, J., Bennett, F.C., Pierik, M.J., Romberg-Camps, M., Mujagic, Z., Prinz, M., Schneider, C.V., Wherry, E.J., Bewtra, M., Heuckeroth, R.O., Levy, M., Thaiss, C.A., 2023. The enteric nervous system relays psychological stress to intestinal inflammation. Cell 186, 2823-2838.e20. https://doi.org/10.1016/j.cell.2023.05.001

Schneider, K.M., Kim, J., Bahnsen, K., Heuckeroth, R.O., Thaiss, C.A., 2022. Environmental perception and control of gastrointestinal immunity by the enteric nervous system. Trends in Molecular Medicine 28, 989–1005. https://doi.org/10.1016/j.molmed.2022.09.005

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and Statistical Modeling with Python, in: Proceedings of the 9th Python in Science Conference. Presented at the Python in Science Conference, Austin, Texas, pp. 92–96. https://doi.org/10.25080/Majora-92bf1922-011

Sedgwick, P., 2014. Understanding why "absence of evidence is not evidence of absence." BMJ 349, g4751. https://doi.org/10.1136/bmj.g4751

Seeling, K.S., Hehl, L., Vell, M.S., Rendel, M.D., Creasy, K.T., Trautwein, C., Mehler, D.M.A., Keszthelyi, D., Schneider, K.M., Schneider, C.V., 2023. Comorbidities, biomarkers and cause specific mortality in patients with irritable bowel syndrome: A phenome-wide association study. United European Gastroenterology Journal 11, 458–470. https://doi.org/10.1002/ueg2.12397

Seminowicz, D.A., Labus, J.S., Bueller, J.A., Tillisch, K., Naliboff, B.D., Bushnell, M.C., Mayer, E.A., 2010. Regional Gray Matter Density Changes in Brains of Patients With Irritable Bowel Syndrome. Gastroenterology 139, 48-57.e2. https://doi.org/10.1053/j.gastro.2010.03.049

Senn, S., 2021. Probability, Bayes, P-values, Tests of Hypotheses and Confidence Intervals, in: Statistical Issues in Drug Development. John Wiley & Sons, Ltd, pp. 49–60. https://doi.org/10.1002/9781119238614.ch4

Spick, M., Onoja, A., Harrison, C., Stender, S., Byrne, J., Geifman, N., 2025. Quantifying new threats to health and biomedical literature integrity from rapidly scaled publications and problematic research. https://doi.org/10.1101/2025.07.07.25331008

Spiller, R.C., Humes, D.J., Campbell, E., Hastings, M., Neal, K.R., Dukes, G.E., Whorwell, P.J., 2010. The Patient Health Questionnaire 12 Somatic Symptom scale as a predictor of symptom severity and consulting behaviour in patients with irritable bowel syndrome and symptomatic diverticular disease. Aliment Pharmacol Ther 32, 811–820.

https://doi.org/10.1111/j.1365-2036.2010.04402.x

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. Archives of Internal Medicine 166, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stein, D.J., Honk, J. van, Thompson, P.M., 2024. The Importance and Limitations of Null Findings. American Journal of Psychiatry. https://doi.org/10.1176/appi.ajp.20240957

Suchak, T., Aliu, A.E., Harrison, C., Zwiggelaar, R., Geifman, N., Spick, M., 2025. Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US national health database. PLOS Biology 23, e3003152. https://doi.org/10.1371/journal.pbio.3003152

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Medicine 12, e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sulaimi, F., Ong, T.S.K., Tang, A.S.P., Quek, J., Pillay, R.M., Low, D.T., Lee, C.K.L., Siah, K.T.H., Ng, Q.X., 2025. Risk factors for developing irritable bowel syndrome: systematic umbrella review of reviews. BMC Med 23, 103. https://doi.org/10.1186/s12916-025-03930-5

Sun, Z., Wang, X., Feng, S., Xie, C., Xing, Y., Guo, L., Zhao, J., Ji, C., 2023. A review of neuroendocrine immune system abnormalities in IBS based on the brain-gut axis and research progress of acupuncture intervention. Front. Neurosci. 17, 934341. https://doi.org/10.3389/fnins.2023.934341

Szucs, D., Ioannidis, J.P.A., 2017. Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLOS Biology 15, e2000797. https://doi.org/10.1371/journal.pbio.2000797

Verdi, S., Rutherford, S., Fraza, C., Tosun, D., Altmann, A., Raket, L.L., Schott, J.M., Marquand, A.F., Cole, J.H., for the Alzheimer's Disease Neuroimaging Initiative, 2024a. Personalizing progressive changes to brain structure in Alzheimer's disease using normative modeling. Alzheimer's & Dementia alz.14174. https://doi.org/10.1002/alz.14174

Verdi, S., Rutherford, S., Fraza, C., Tosun, D., Altmann, A., Raket, L.L., Schott, J.M., Marquand, A.F., Cole, J.H., Initiative, for the A.D.N., 2024b. Personalizing progressive changes to brain structure in Alzheimer's disease using normative modeling. Alzheimer's & Dementia 20, 6998–7012. https://doi.org/10.1002/alz.14174

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J.,

Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Van Mulbregt, P., SciPy 1.0 Contributors, Vijaykumar, A., Bardelli, A.P., Rothberg, A., Hilboll, A., Kloeckner, A., Scopatz, A., Lee, A., Rokem, A., Woods, C.N., Fulton, C., Masson, C., Häggström, C., Fitzgerald, C., Nicholson, D.A., Hagen, D.R., Pasechnik, D.V., Olivetti, E., Martin, E., Wieser, E., Silva, F., Lenders, F., Wilhelm, F., Young, G., Price, G.A., Ingold, G.-L., Allen, G.E., Lee, G.R., Audren, H., Probst, I., Dietrich, J.P., Silterra, J., Webber, J.T., Slavič, J., Nothman, J., Buchner, J., Kulick, J., Schönberger, J.L., De Miranda Cardoso, J.V., Reimer, J., Harrington, J., Rodríguez, J.L.C., Nunez-Iglesias, J., Kuczynski, J., Tritz, K., Thoma, M., Newville, M., Kümmerer, M., Bolingbroke, M., Tartre, M., Pak, M., Smith, N.J., Nowaczyk, N., Shebanov, N., Pavlyk, O., Brodtkorb, P.A., Lee, P., McGibbon, R.T., Feldbauer, R., Lewis, S., Tygier, S., Sievert, S., Vigna, S., Peterson, S., More, S., Pudlik, T., Oshima, T., Pingel, T.J., Robitaille, T.P., Spura, T., Jones, T.R., Cera, T., Leslie, T., Zito, T., Krauss, T., Upadhyay, U., Halchenko, Y.O., Vázquez-Baeza, Y., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Walker, E., Nowacki, A.S., 2011. Understanding Equivalence and Noninferiority Testing. J GEN INTERN MED 26, 192–196. https://doi.org/10.1007/s11606-010-1513-8

Weaver, K.R., Sherwin, L.B., Walitt, B., Melkus, G.D., Henderson, W.A., 2016. Neuroimaging the brain-gut axis in patients with irritable bowel syndrome. WJGPT 7, 320. https://doi.org/10.4292/wjgpt.v7.i2.320

Winter, N.R., Leenings, R., Ernsting, J., Sarink, K., Fisch, L., Emden, D., Blanke, J., Goltermann, J., Opel, N., Barkhau, C., Meinert, S., Dohm, K., Repple, J., Mauritz, M., Gruber, M., Leehr, E.J., Grotegerd, D., Redlich, R., Jansen, A., Nenadic, I., Nöthen, M.M., Forstner, A., Rietschel, M., Groß, J., Bauer, J., Heindel, W., Andlauer, T., Eickhoff, S.B., Kircher, T., Dannlowski, U., Hahn, T., 2022. Quantifying Deviations of Brain Structure and Function in Major Depressive Disorder Across Neuroimaging Modalities. JAMA Psychiatry 79, 879–888. https://doi.org/10.1001/jamapsychiatry.2022.1780

Yan, C.-G., Wang, Z.-H., Han, L.K.M., Alexander, N., Alnæs, D., Başgöze, Z., Bauduin, S.E.E.C., Bauer, J., Benedetti, F., Berger, K., Besteher, B., Borgers, T., Bülow, R., Cao, L.-P., Chen, G.-M., Chen, J.-S., Chen, S.-C., Chen, T., Chen, T.-L., Chen, X., Chen, Y.-R., Cheng, Y.-Q., Chu, Z.-S., Colle, R., Connolly, C.G., Corruble, E., Couvy-Duchesne, B., Cui, X.-L., Cullen, K.R., Dannlowski, U., Davey, C., Dima, D., Dols, A., Erwin-Grabner, T., Evans, J.W., Flinkenflügel, K., Fu, C.H.Y., Fuentes-Claramonte, P., Gao, Q.-L., Gong, Q.-Y., Gönül, A.S., Gotlib, I.H., Goya-

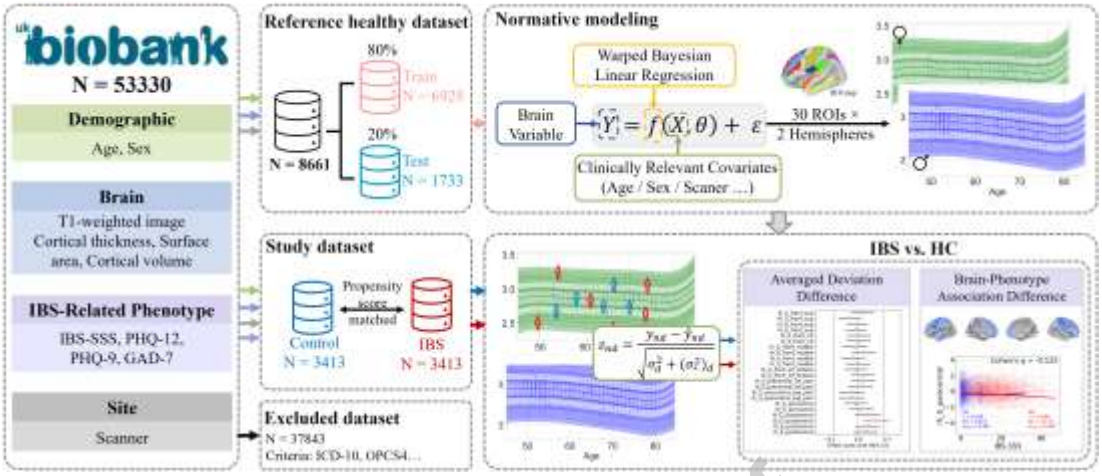
Maldonado, R., Grabe, H.J., Groenewold, N., Grotegerd, D., Guo, W.-B., Hahn, T., Hall, G.B., Hamilton, J.P., Harrison, B., Hatton, S., He, C.-C., Hermesdorf, M., Hickie, I.B., Ho, T.C., Hu, Z.-J.-Y., Huang, Q., Jahanshad, N., Jamalabadi, H., Jamieson, A., Ji, X.-L., Jia, F.-N., Kamishikiryo, T., Kircher, T., Koopowitz, S.-M., Kraus, A., Krug, A., Kuang, L., Lagopoulos, J., Lancaster, T., Landrø, N.I., Leehr, E.J., Li, M., Li, B.-J., Li, F., Li, H.-X., Li, T., Li, X.-Y., Liao, Y.-F., Lin, Z.-D., Linden, D.E.J., Liu, H.-L., Liu, X.-Y., Liu, Z.-N., Liu, Y.-S., Long, Y.-C., Lu, J.-P., Lu, B., MacMaster, F.P., Maglanoc, L.A., McIntosh, A.M., Medland, S., Mehler, D.M.A., Meinert, S., Mwangi, B., Ojha, A., Okada, G., Oudega, M.L., Penninx, B.W.J.H., Poletti, S., Pomarol-Clotet, E., Portella, M.J., Qiu, J., Reneman, L., Rise, H.S., Rodriguez-Cano, E., Sacchet, M.D., Salvador, R., Schlieper, H., Schrantee, A., Shan, X.-X., Shinzato, H., Si, T.-M., Sim, K., Soares, J.C., Stein, D.J., Straube, B., Strike, L.T., Sun, P.-F., Thomas-Odenthal, F., Thomopoulos, S.I., Thompson, P.M., Toenders, Y., Uyar, A., van der Wee, N.J.A., van der Werff, S.J.A., Veer, I.M., Veltman, D.J., Völzke, H., Wagenmakers, M., Walter, H., Walter, M., Wang, C.-Y., Wang, H.-L., Wang, H.-N., Wang, X., Wang, X.-Y., Wang, Y., Wang, Y.-W., Westlye, L.T., Whalley, H.C., Whittle, S., Winter, N.R., Wittfeld, K., Wu, M.-J., Wu, X.-P., Wu, X.-R., Wu, Y.-K., Wu, C.-N., Xie, C.-M., Xie, G.-R., Xie, P., Xu, X.-F., Xue, Z.-P., Yang, T.T., Yang, H., Yang, J., Yu, H., Yu, Y.-Q., Yuan, M.-L., Yuan, Y.-G., Zang, Y.-F., Zarate, C.A., Zhang, A.-X., Zhang, K.-R., Zhang, W., Zhao, J.-P., Zhong, W.-Y., Zhu, J.-J., Zunta-Soares, G.B., Zuo, X.-N., Pozzi, E., Schmaal, L., 2026. Vertex-wise cortical abnormalities in major depressive disorder from 64 cohorts from the DIRECT and ENIGMA MDD consortia. Nat. Mental Health 4, 1175–1186. https://doi.org/10.1038/s44220-026-00667-9

Zabihi, M., Oldehinkel, M., Wolfers, T., Frouin, V., Goyard, D., Loth, E., Charman, T., Tillmann, J., Banaschewski, T., Dumas, G., Holt, R., Baron-Cohen, S., Durston, S., Bölte, S., Murphy, D., Ecker, C., Buitelaar, J.K., Beckmann, C.F., Marquand, A.F., 2019. Dissecting the Heterogeneous Cortical Anatomy of Autism Spectrum Disorder Using Normative Models. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 4, 567–578. https://doi.org/10.1016/j.bpsc.2018.11.013

Zhao, M., Hao, Z., Li, M., Xi, H., Hu, S., Wen, J., Gao, Y., Antwi, C.O., Jia, X., Yu, Y., Ren, J., 2023. Functional changes of default mode network and structural alterations of gray matter in patients with irritable bowel syndrome: a meta-analysis of whole-brain studies. Front. Neurosci. 17, 1236069. https://doi.org/10.3389/fnins.2023.1236069

Zheng, D., Liwinski, T., Elinav, E., 2020. Interaction between microbiota and immunity in health and disease. Cell Res 30, 492–506. https://doi.org/10.1038/s41422-020-0332-7

Fig. 1



\

Fig. 2

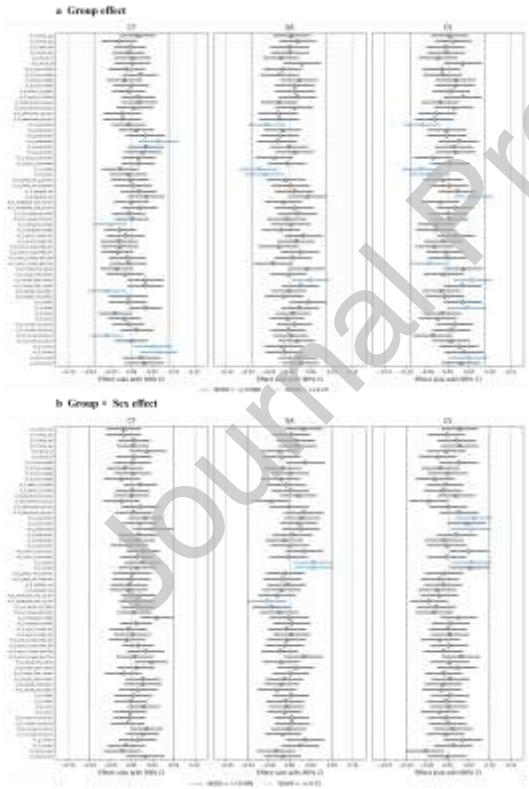
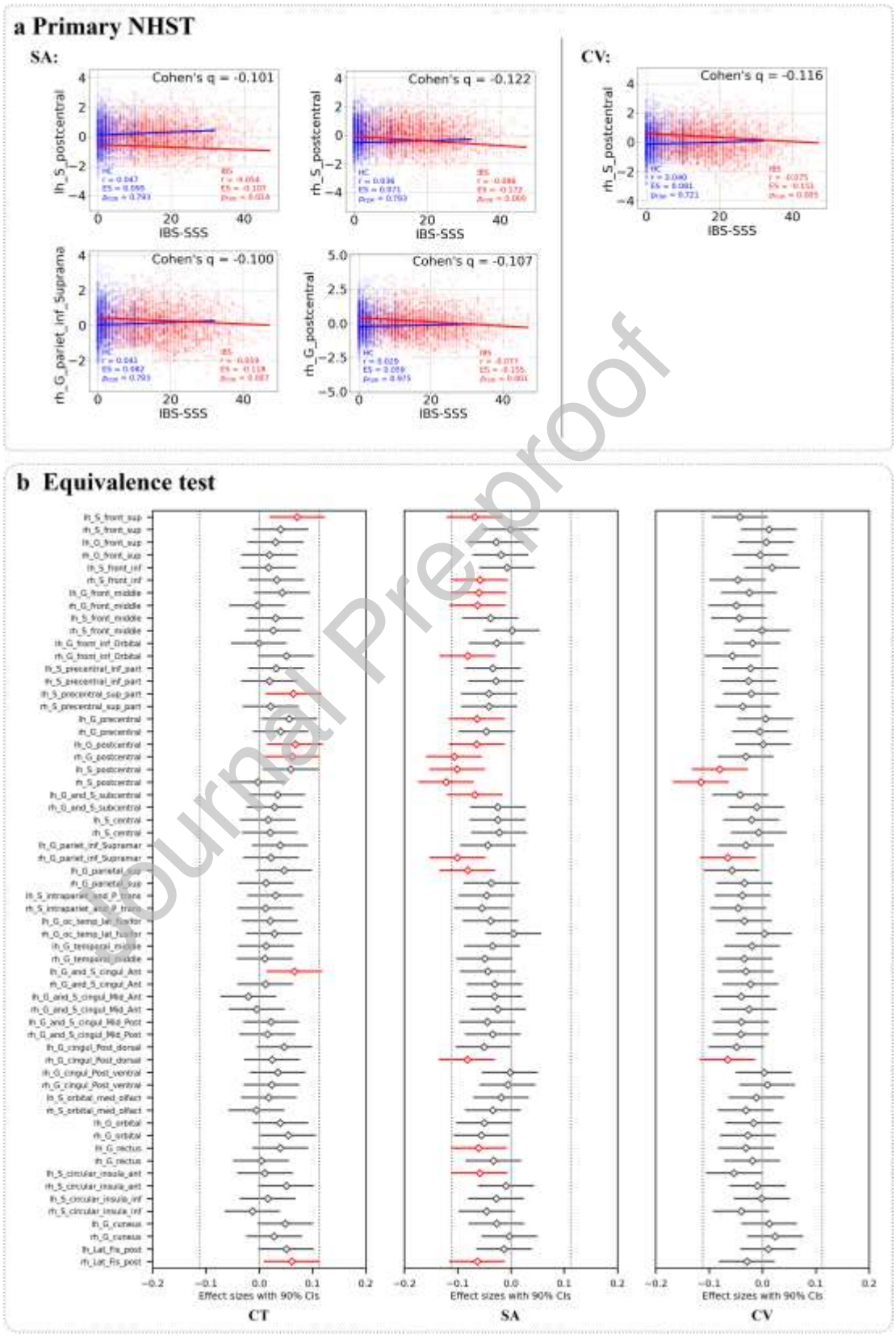
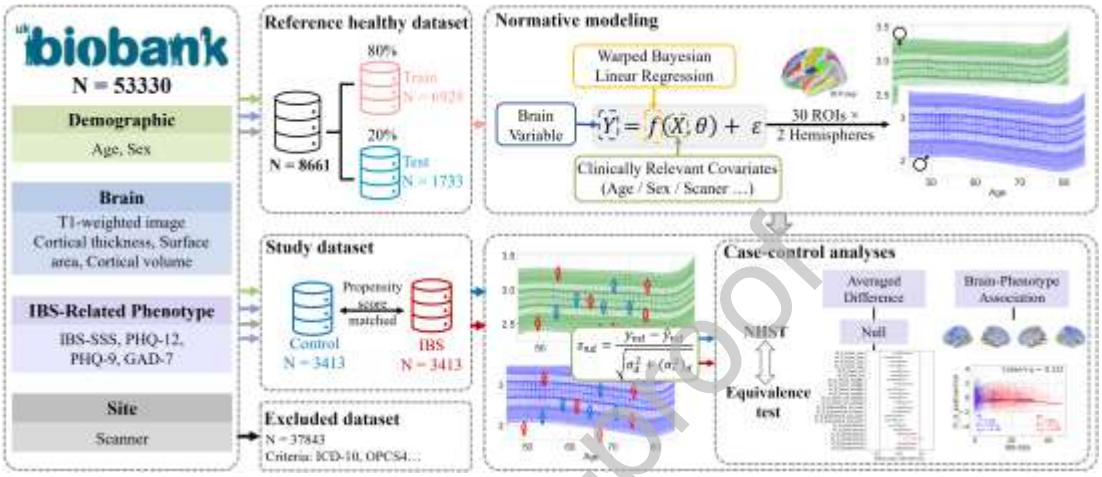


Fig. 3



Graphical abstract



Data Availability

This study has been conducted using UK Biobank Resource under Application 71300. All raw data are available from the UKB (www.ukbiobank.ac.uk). The raw and processed UKB MRI data are protected and are not openly available due to data privacy laws. Access can be obtained by applying for access (www.ukbiobank.ac.uk/enable-your-research/apply-for-access).

Code Availability

All software used is publicly available at the URL or references cited. Code supporting the findings of this study is available at https://github.com/hanbing-gao/NM_IBS.

Conflict of interest disclosure

The authors declare no competing interests.