nature research

Corresponding author(s): Saige Rutherford
saige.rutherford@donders.ru.nl

Last updated by author(s): Jan 21, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All code in available on GitHub (https://github.com/predictive-clinical-neuroscience/PCNtoolkit-demo) and Zenodo (https://zenodo.org/record/5592153#.YenuRFjMJb8)
Data analysis	All analysis was run using python version 3.8 and the predictive clinical neuroscience python package version 0.20.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data in available on GitHub (https://github.com/predictive-clinical-neuroscience/PCNtoolkit-demo) and Zenodo (https://zenodo.org/record/5592153#.YenuRFjMJb8)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Cross-sectional samples, no repeated scans were included.
Research sample	Dataset includes a multi-site dataset combining publicly available data from the Human Connectome Project Young Adult, CAMCAN, and IXI samples.
Sampling strategy	N/A. All data used was secondary data analysis from public open data sets.
Data collection	N/A. All data used was secondary data analysis from public open data sets.
Timing	N/A
Data exclusions	Subjects had to have T1-weighted structural image that successfully finished Freesurfer surface reconstruction, and available demographic variables (age and sex).
Non-participation	N/A
Randomization	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Healthy subjects with no clinical diagnosis.
Recruitment	N/A. All data used was secondary data analysis from public open data sets.
Ethics oversight	Radboud University ethics board approved of this secondary data analysis study. All original data collection was approved by the ethics committee of the university where the original study was conducted.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	N/A
Design specifications	N/A
Behavioral performance measures	N/A

Acquisition

Imaging type(s)	Structural
Field strength	3.0T
Sequence & imaging parameters	High resolution T1-weighted MPRAGE images were used.
Area of acquisition	whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Freesurfer version 6.0
Normalization	recon-all
Normalization template	fsaverage template, Desikan-Killian atlas.
Noise and artifact removal	Data that did not successfully run Freesurfer surface reconstruction were excluded. Euler number, a scan quality metric, was used to exclude low-quality subjects.
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	Normative modeling using Bayesian linear regression to predict brain region Y, a n_subjects x 1 vector from X, a matrix with n_subjects x n_covariates (i.e., age, sex, site, data quality metric). There is a separate regression model for each brain region.
Effect(s) tested	The individual-level outputs of normative modeling are the predicted brain score, the predictive variance (separated into modeling and noise components), a deviation score (Z-score, how much each subject deviate from the normative range). The overall performance is evaluated out-of-sample by correlating predicted and true values, calculating explained variance, standardized mean squared error, and mean standardized log loss.

Specify type of analysis: ☐ Whole brain ☒ ROI-based ☐ Both

Anatomical location(s) | Desikan-Killian atlas was used.

Statistic type for inference (See Eklund et al. 2016)	ROI-level inference
Correction	N/A

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis | Normative modeling using Bayesian Linear Regression.