Cell Reports, Volume *43*

Supplemental information

Performance reserves

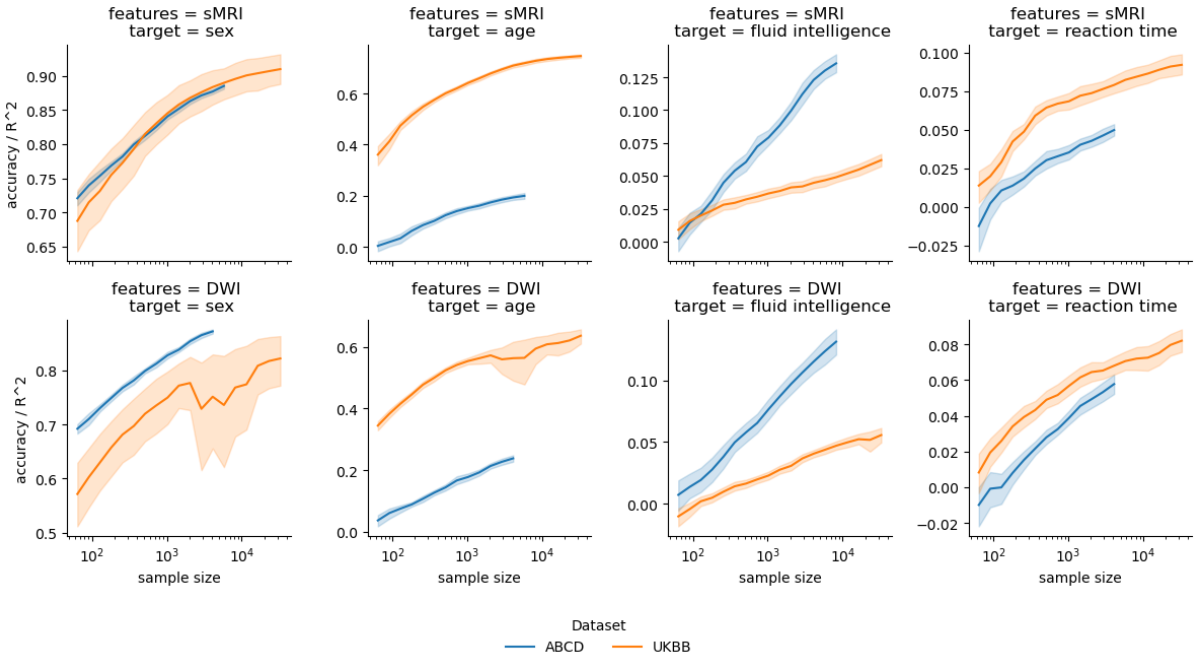
in brain-imaging-based phenotype prediction

Marc-Andre Schulz, Danilo Bzdok, Stefan Haufe, John-Dylan Haynes, and Kerstin Ritter

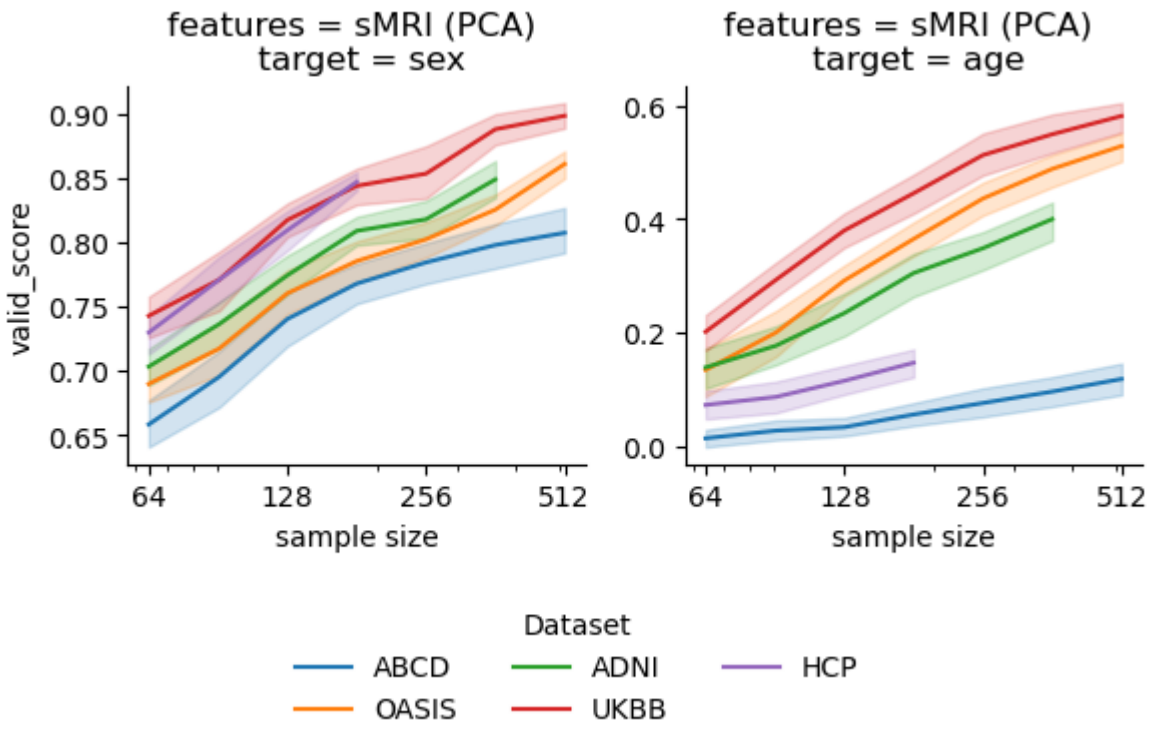
Performance reserves in brain-imaging-based phenotype prediction
-
Supplementary online material

Marc-Andre Schulz[1,2,*], Danilo Bzdok[3,4,5], Stefan Haufe[2,6,7,8], John-Dylan Haynes[2,8], Kerstin Ritter[1,2]

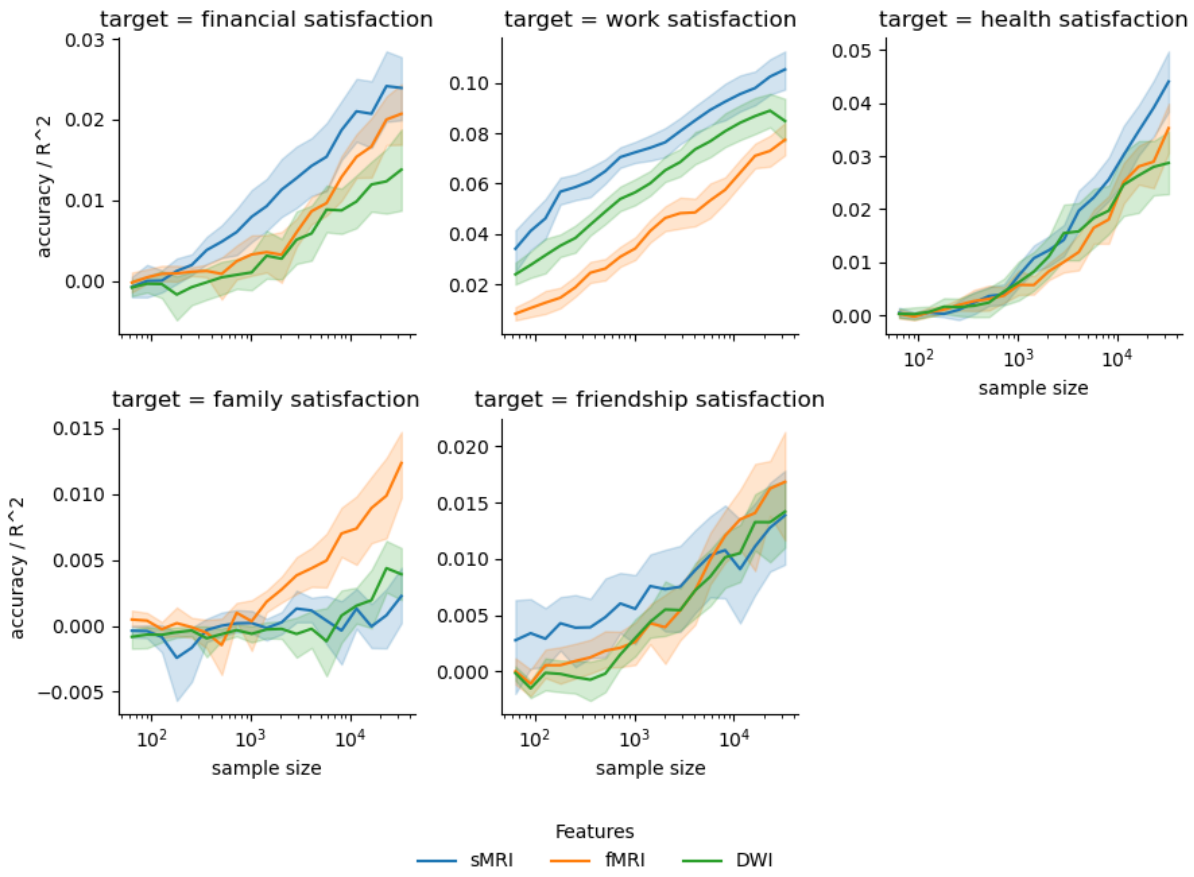
[1] Charité – Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Department of Psychiatry and Psychotherapy, Berlin, Germany
[2] Bernstein Center for Computational Neuroscience, Berlin, Germany
[3] McConnell Brain Imaging Centre (BIC), Montreal Neurological Institute (MNI), Faculty of Medicine, McGill University, Montreal, Canada
[4] Department of Biomedical Engineering, Faculty of Medicine, McGill University, Montreal, Canada
[5] Mila - Quebec Artificial Intelligence Institute, Montreal, Canada
[6] Technische Universität Berlin, Berlin, Germany
[7] Physikalisch-Technische Bundesanstalt, Berlin, Germany
[8] Charité – Universitätsmedizin Berlin (corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health), Department of Neurology, Berlin Center for Advanced Neuroimaging, Berlin, Germany
* Corresponding author: Marc-Andre Schulz - marc-andre.schulz@charite.de



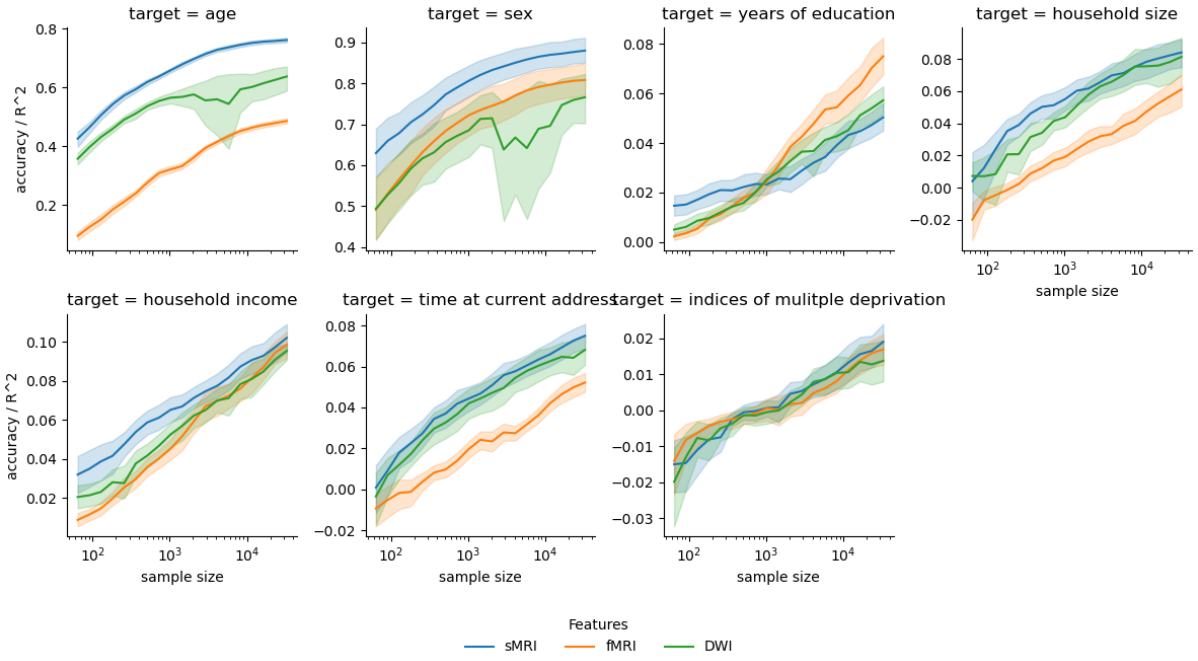
SFig 1: Comparative replication analysis of Adolescent Brain Cognitive Development (ABCD) and UK Biobank datasets illustrating power-law learning curves. The ABCD dataset (comprising adolescents) and the UK Biobank (comprising seniors) were evaluated for four key variables (age, sex, fluid intelligence, and reaction time) using available structural MRI and diffusion-weighted MRI (dMRI) imaging derived phenotypes (IDPs). Both datasets showed comparable scaling behavior for predicting sex and age, while fluid intelligence and reaction time displayed dataset-specific offsets. Notably, the ABCD dataset exhibited 15% explained variance for fluid intelligence using dMRI, surpassing UK Biobank's fMRI results. However, inherent cohort and data differences mandate cautious interpretation of direct comparisons.



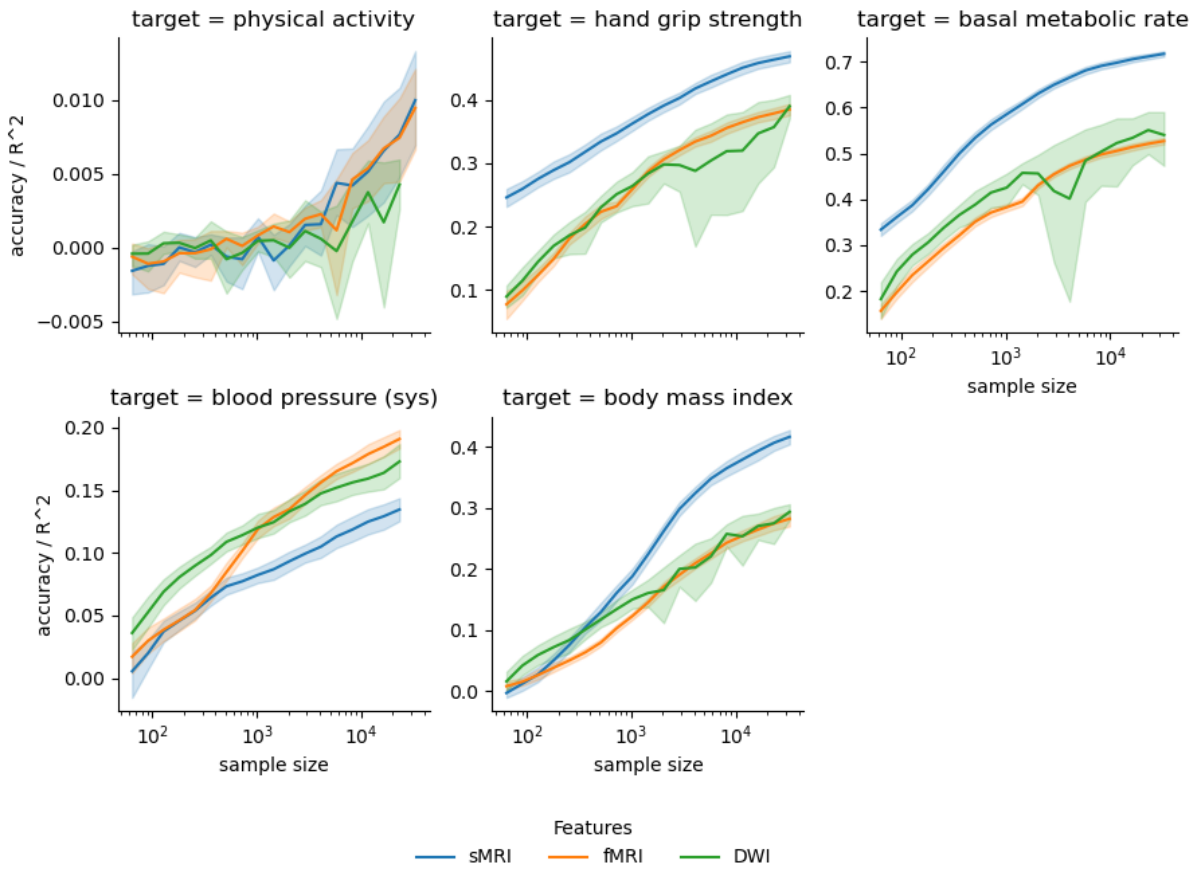
SFig 2: Robustness analysis across diverse neuroimaging datasets (OASIS, ADNI, HCP) for age and sex prediction using structural MRI data. Despite variations in sample sizes, age distributions, and sex ratios, a consistent pattern of learning curves gradually saturating is observed across datasets. Given the significant differences in available imaging-derived phenotypes (IDPs), a Principal Component Analysis (PCA) was performed to extract the first 100 principal components for each dataset, aiming to make feature spaces as comparable as possible. Slight offsets in pattern are attributed to unique dataset characteristics, such as differing sex ratios and age ranges. Lower R2 values for datasets with a narrow age range (ABCD and HCP) underscore the robustness and generalizability of the primary findings from the UK Biobank dataset.



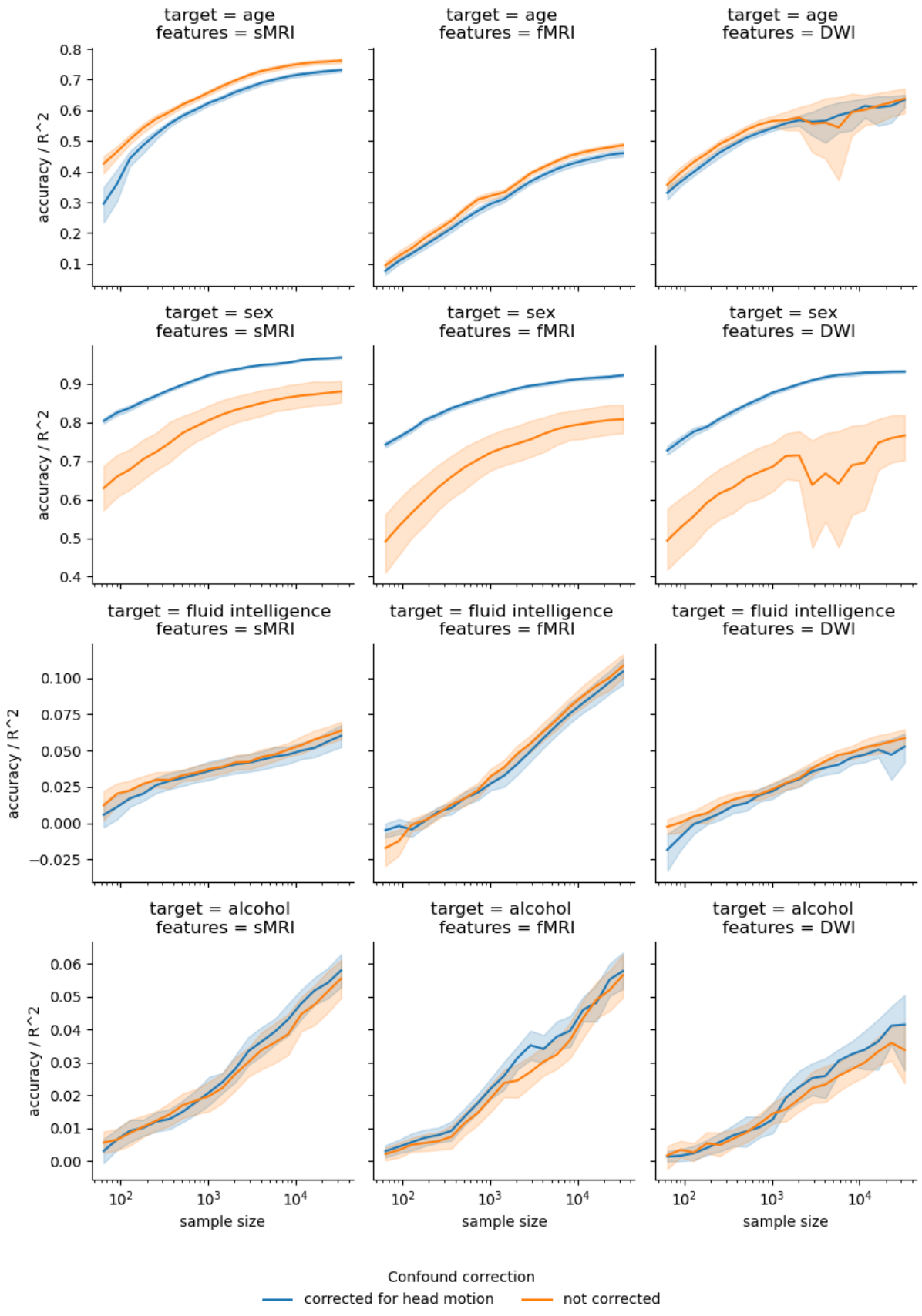
SFig 3: Expanded selection of target phenotypes. Comparative analysis of satisfaction measures across different brain imaging modalities. The figure comprises five subplots, each representing a different area of life satisfaction (family, friends, health, finance, work) as measured in the UK Biobank study. Predictive accuracies based on structural MRI, resting-state fMRI, and diffusion-weighted MRI data are displayed within each subplot. Satisfaction questions are answered on a consistent 6-point scale, allowing for comparability across domains. The 'Family' and 'Friends' subplots, representing social aspects of life, demonstrate higher predictive accuracy with fMRI data. Conversely, 'Health', 'Finance', and 'Work' satisfaction, pertaining to non-social aspects of life, show superior predictive accuracy with sMRI data. This pattern suggests that different imaging modalities may provide distinct information value for predicting aspects of mental health.



SFig 4: Comparative analysis of sociodemographic measures across different brain imaging modalities. The figure comprises multiple subplots, each representing a different sociodemographic measure (years of education, time at current address, household size, household income, Index of Multiple Deprivation) as identified in the UK Biobank study. Each subplot illustrates the predictive accuracy of structural MRI, resting-state fMRI, and diffusion-weighted MRI data for the corresponding sociodemographic measure. Patterns of predictive accuracy across imaging modalities vary between the sociodemographic measures, potentially reflecting the distinctive neural correlates or the varied impact of sociodemographic factors on brain structures and function.

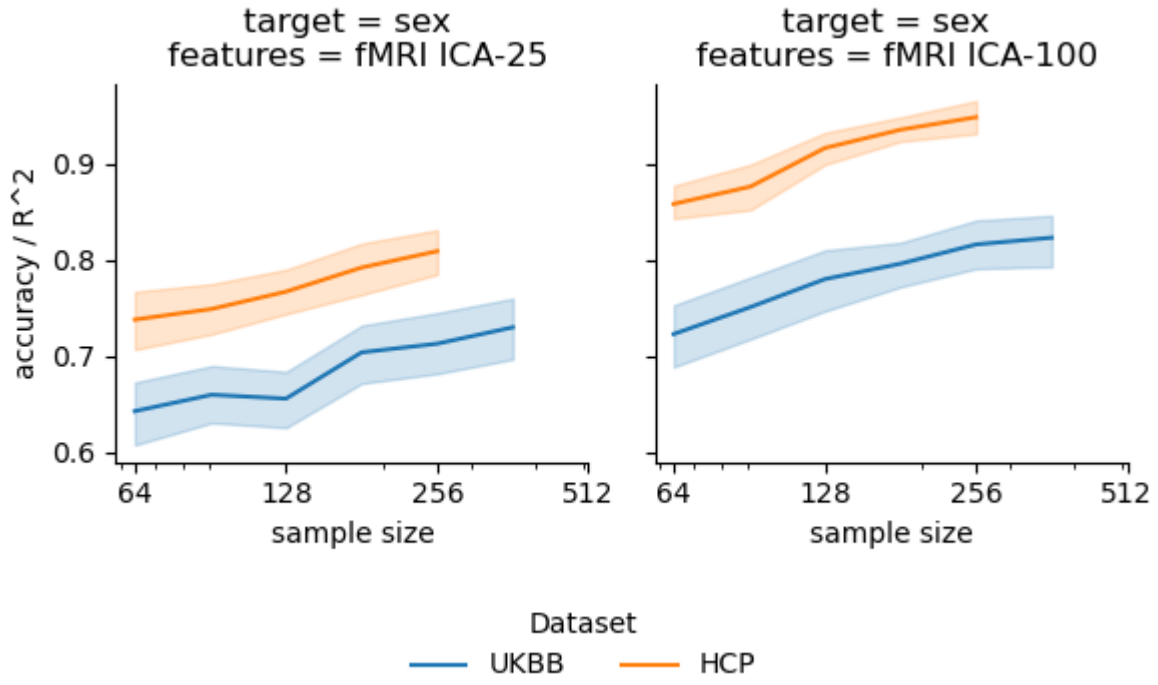


SFig 5: Comparative analysis of biological measures across different brain imaging modalities. The figure comprises multiple subplots, each representing a different biological measure (blood pressure, hand grip strength, level of physical activity / IPAQ activity group body mass index, basal metabolic rate, as provided by the UK Biobank study). For each biological measure, the corresponding subplot illustrates the predictive accuracy of structural MRI, resting-state fMRI, and diffusion-weighted MRI data. Most biological measures display an earlier saturation of prediction accuracy compared to mental health or sociodemographic target variables (SFig 3+4), likely due to more precise and objective measurements.

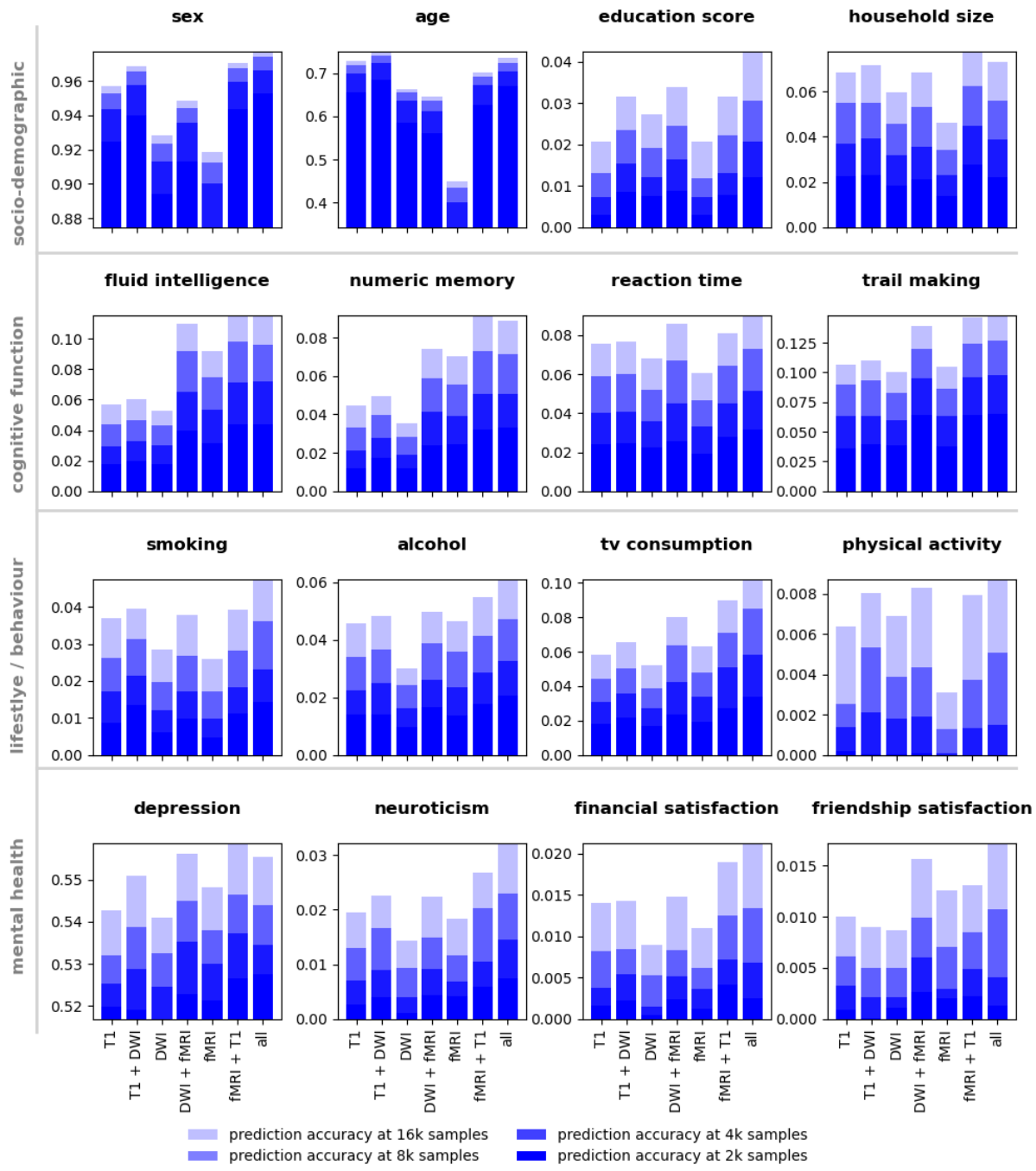


SFig 6: Analysis of the influence of data quality on learning curves and prediction performance, specifically highlighting the impact of mean absolute head motion (UKB field 24438). Despite initial preprocessing and motion corrections conducted using standard FSL

procedures, it was necessary to determine to what extent residual artifacts might have influenced model predictions. Our approach involved an additional linear correction for mean absolute head motion applied to the original imaging-derived phenotypes provided by the UK Biobank. The results indicate a slight enhancement in sex prediction performance post-correction, while the influence on other targets across all modalities was negligible. This suggests that the potential confounding effects of head motion on our analyses were likely minimal.



SFig 7: Comparison of resting-state fMRI prediction performance between the Human Connectome Project (HCP) dataset with ~30 minute scan duration and the UK Biobank (UKBB) with ~5 minute scan duration. Sex prediction based on functional connectivity features derived from ICA with 25 and 100 components. The HCP data demonstrated overall higher prediction accuracy (by ~5-10%) but retained a comparable saturation pattern as the UKBB data. This observation suggests that longer scan durations and higher data quality impact performance but not the fundamental scaling behavior. Caution is warranted when directly comparing cohorts due to differences in age distribution (HCP: 22-35 years, UKBB: 40-70 years) and other factors.



SFig 8: Alternative visualization of main text Figure 4. Augmenting single-modality feature spaces to incorporate multimodal input data can lead to improvements in prediction accuracy on par with doubling the sample size. Please refer to the main text for details.

a) main text analyses			
features	**DWI**	**fMRI**	**sMRI**
target	**max train size used**		
age	32768	32768	32768
alcohol	32768	32768	32768
depression	32768	32768	32768
education score	23170	23170	32768
financial satisfaction	32768	32768	32768
fluid intelligence	23170	23170	32768
friendship satisfaction	32768	32768	32768
household size	32768	32768	32768
neuroticism	23170	23170	23170
numeric memory	23170	16384	23170
physical activity	23170	23170	23170
reaction time	23170	23170	32768
sex	32768	32768	32768
smoking	32768	32768	32768
trail making	16384	16384	16384
tv consumption	32768	32768	32768

b) supplementary analyses			
features	**DWI**	**fMRI**	**sMRI**
target	**max train size used**		
age	32768	32768	32768
alcohol	32768	32768	32768
basal metabolic rate	32768	32768	32768
blood pressure systolic	23170	23170	23170
bmi	32768	32768	32768
depression	32768	32768	32768
edu	32768	32768	32768
family satisfaction	32768	32768	32768

body fat ratio	23170	23170	32768
fluid intelligence	32768	32768	32768
financial satisfaction	32768	32768	32768
friendship satisfaction	32768	32768	32768
grip strength	32768	32768	32768
health satisfaction	32768	32768	32768
household income	32768	32768	32768
household size	32768	32768	32768
multiple deprivation	32768	32768	32768
neuroticism	23170	32768	32768
numeric memory	23170	23170	23170
physical activity	23170	32768	32768
reaction time	32768	32768	32768
sex	32768	32768	32768
time at current address	32768	32768	32768
tobacco consumption	32768	32768	32768
trail making	23170	23170	23170
tv consumption	32768	32768	32768
work satisfaction	32768	32768	32768

STable 1: Maximum sample sizes used for training, shown per target phenotype and feature modality. Sample sizes differ due to missing imaging or target phenotype data in the UK Biobank. The supplementary analyses use a more recent release of the UK Biobank with a slightly increased sample size.