nature human behaviour

Supplementary information

https://doi.org/10.1038/s41562-024-02074-5

Neuropsychiatric polygenic scores are weak predictors of professional categories

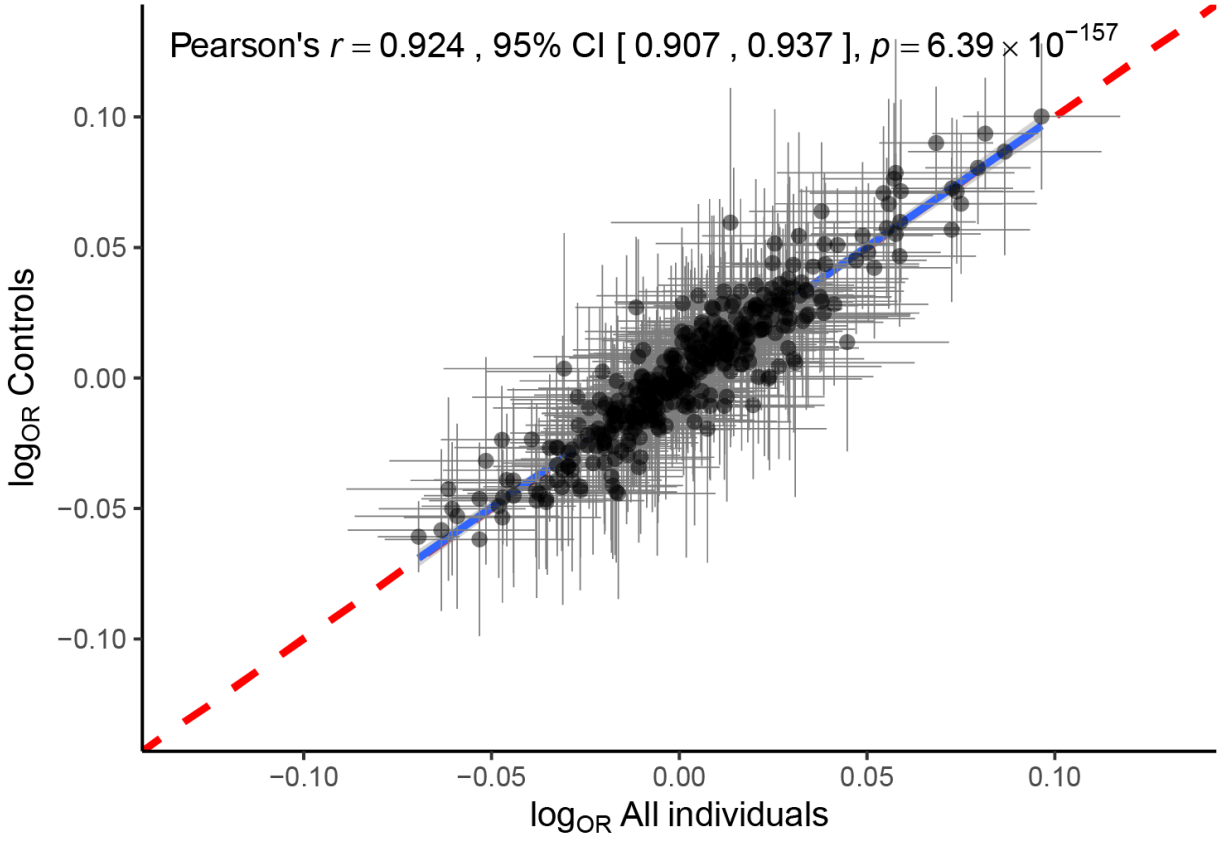
In the format provided by the authors and unedited

Supplementary Information

Table of contents

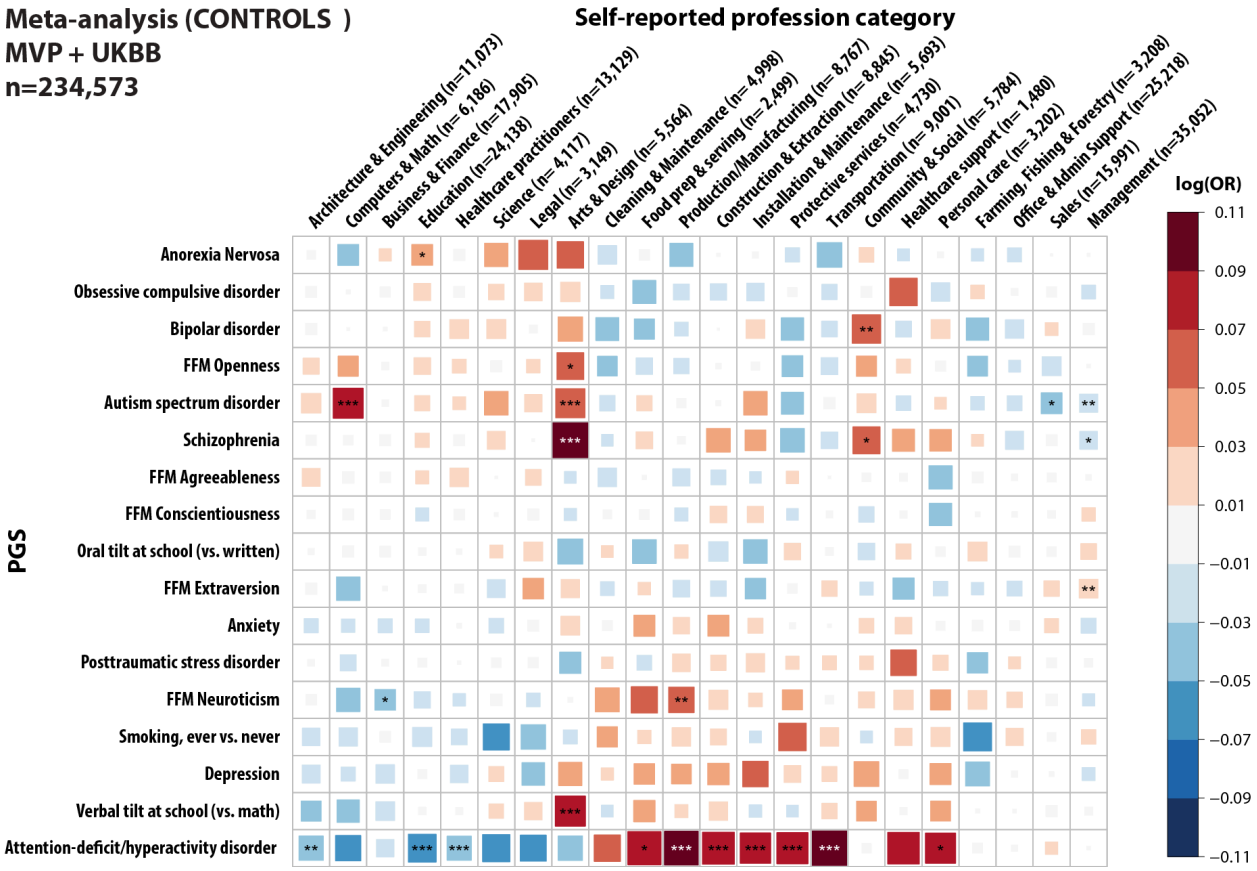
SUPPLEMENTARY FIGURES

Supplementary Figure 1



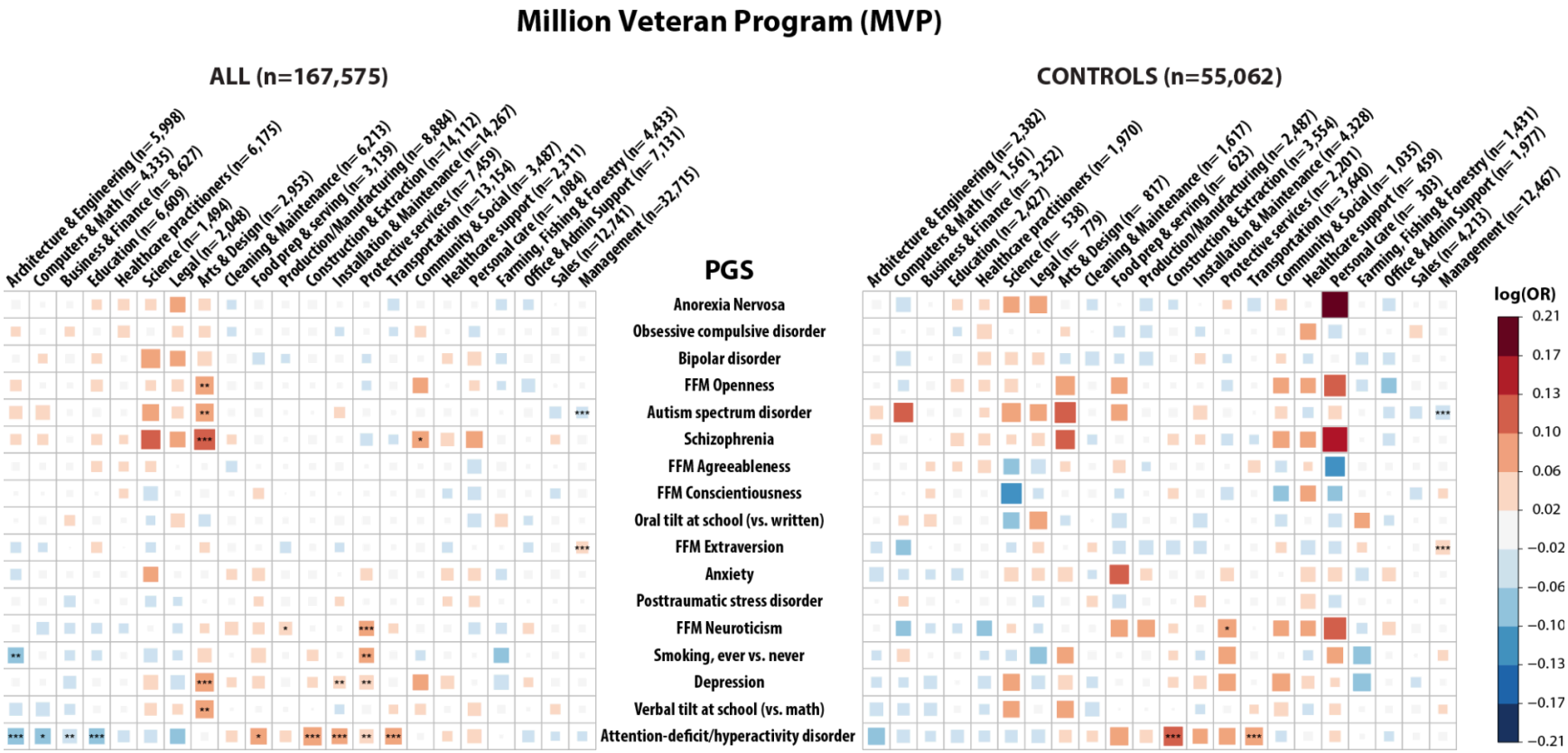
Supplementary Figure 1. Correlation analysis of log odds estimates (log_{OR}) for PGS trait - profession pairs while including ("All individuals"; y-axis) or excluding individuals with neuropsychiatric disorders ("Controls"; x-axis), corresponding to cohorts "ALL" and "CONTROLS". Pearson's *r* and two-tailed p value are provided, n = 374.

Supplementary Figure 2



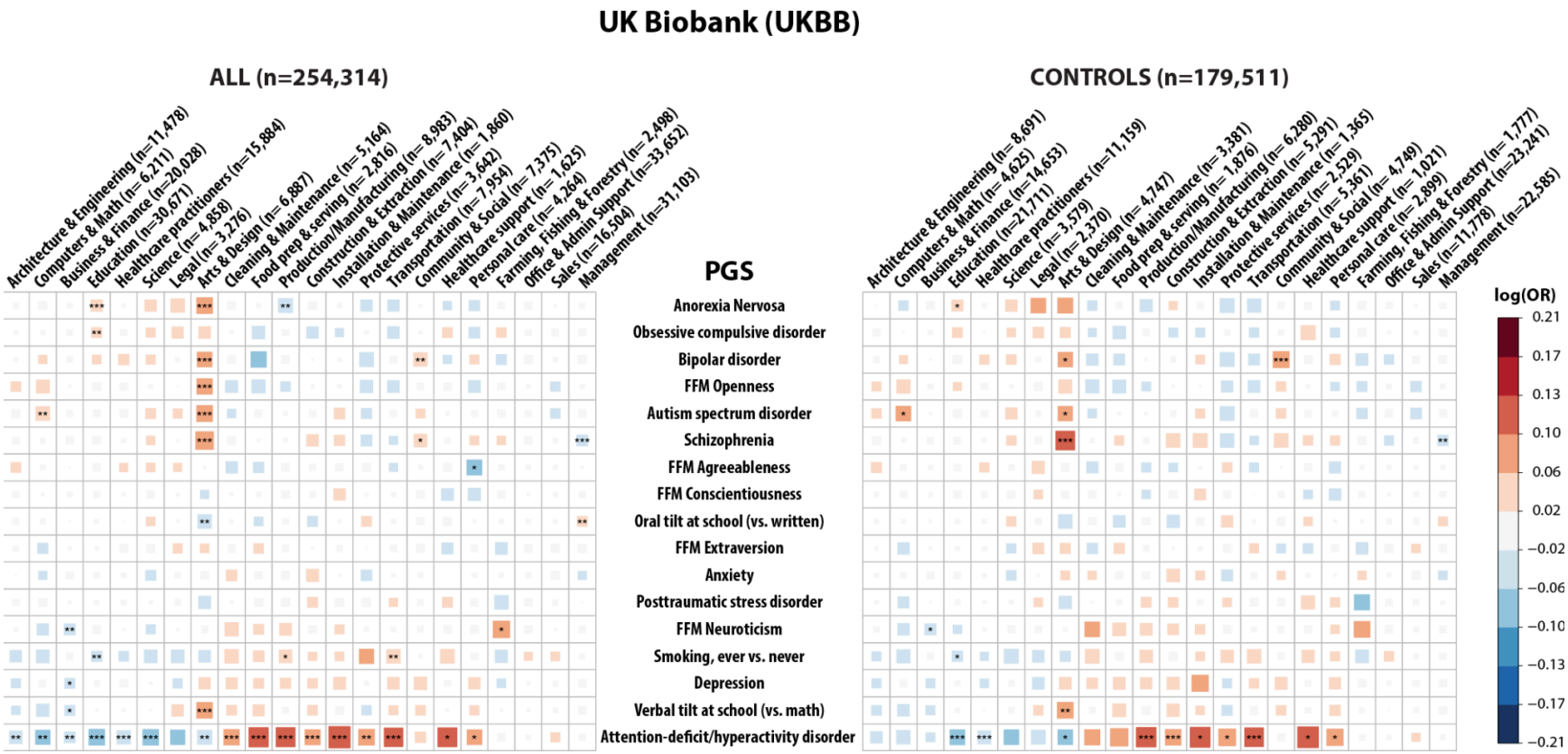
Supplementary Figure 2. Meta-analysis (fixed-effect inverse-variance weighted) of associations between neuropsychiatric trait PGS and distinct professional categories when excluding individuals with neuropsychiatric diagnoses (CONTROLS), as estimated by logistic regression in European ancestry individuals of both cohorts (MVP and UKBB; $\text{profession}_i \sim \text{PGS}_j + \text{age} + \text{age}^2 + \text{sex} + \text{ancestral PC1 to PC20} + \text{genotyping batch}$ (needed only for UKBB), for each of the 374 PGS_j-profession_i pairs). Plotted log odds denoted by both color and size, with statistical significance denoted by stars (***, ** and * correspond to Bonferroni-adjusted two-tailed p values of association equal or less than 0.001, 0.01 and 0.05 respectively). In spite of the reduction of sample sizes, cognitive traits retained statistically significant associations with about half of the profession categories identified in **Fig. 1**, suggesting genetic pleiotropy.

Supplementary Figure 3



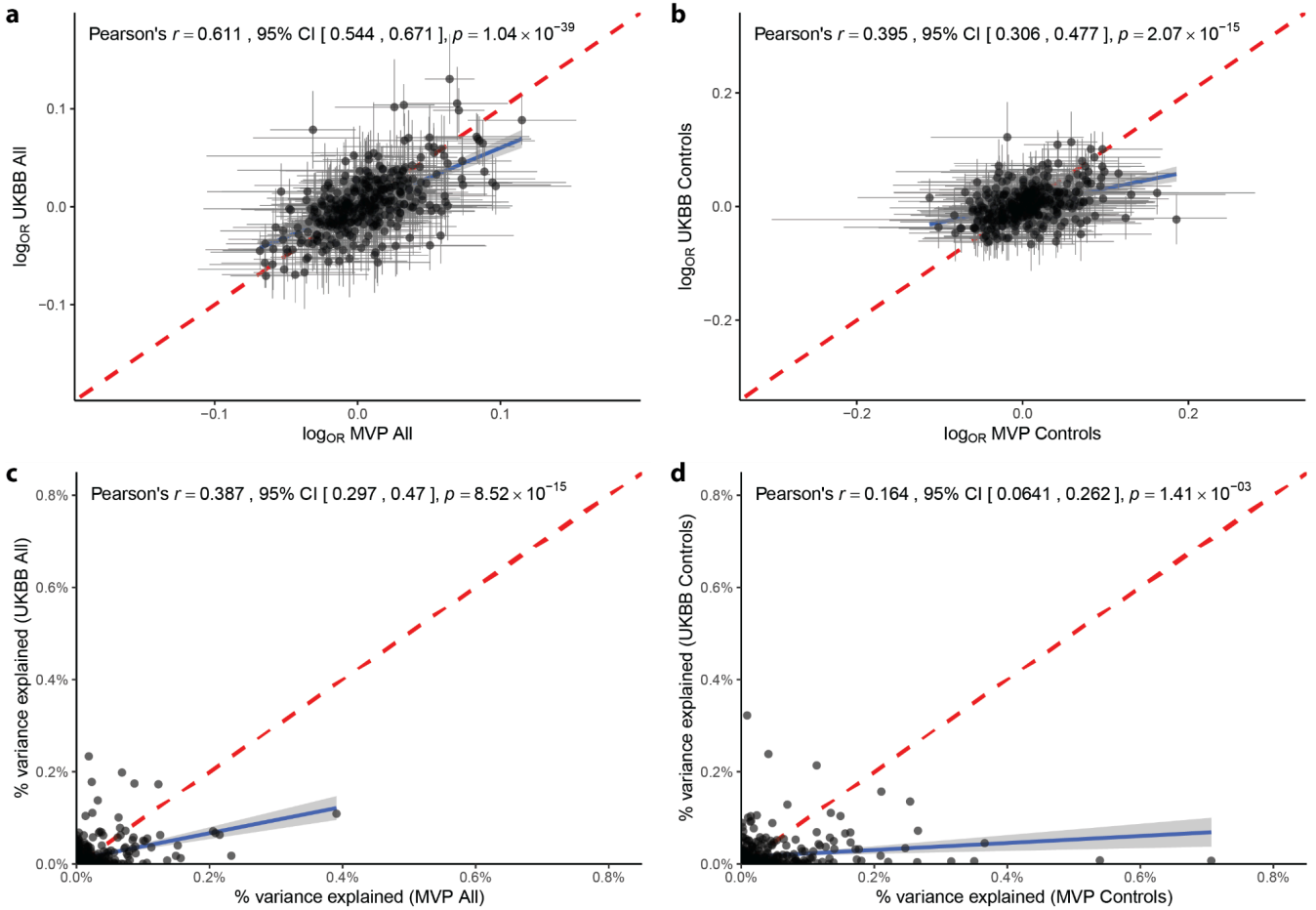
Supplementary Figure 3. Association between neuropsychiatric trait PGS and professions in the Million Veteran Program (MVP). Logistic regression analysis was performed as follows: $\text{profession}_i \sim \text{PGS}_j + \text{age} + \text{age}^2 + \text{sex} + \text{ancestral PC1 to PC20}$, for each of the 374 PGS_j-profession_i pairs. The CONTROLS group only comprises individuals of the ALL group without neuropsychiatric diagnoses. Plotted log odds (log(OR)) denoted by both color and size with statistical significance denoted by stars (***, ** and * correspond to Bonferroni-adjusted two-tailed p values of association equal or less than 0.001, 0.01 and 0.05 respectively).

Supplementary Figure 4



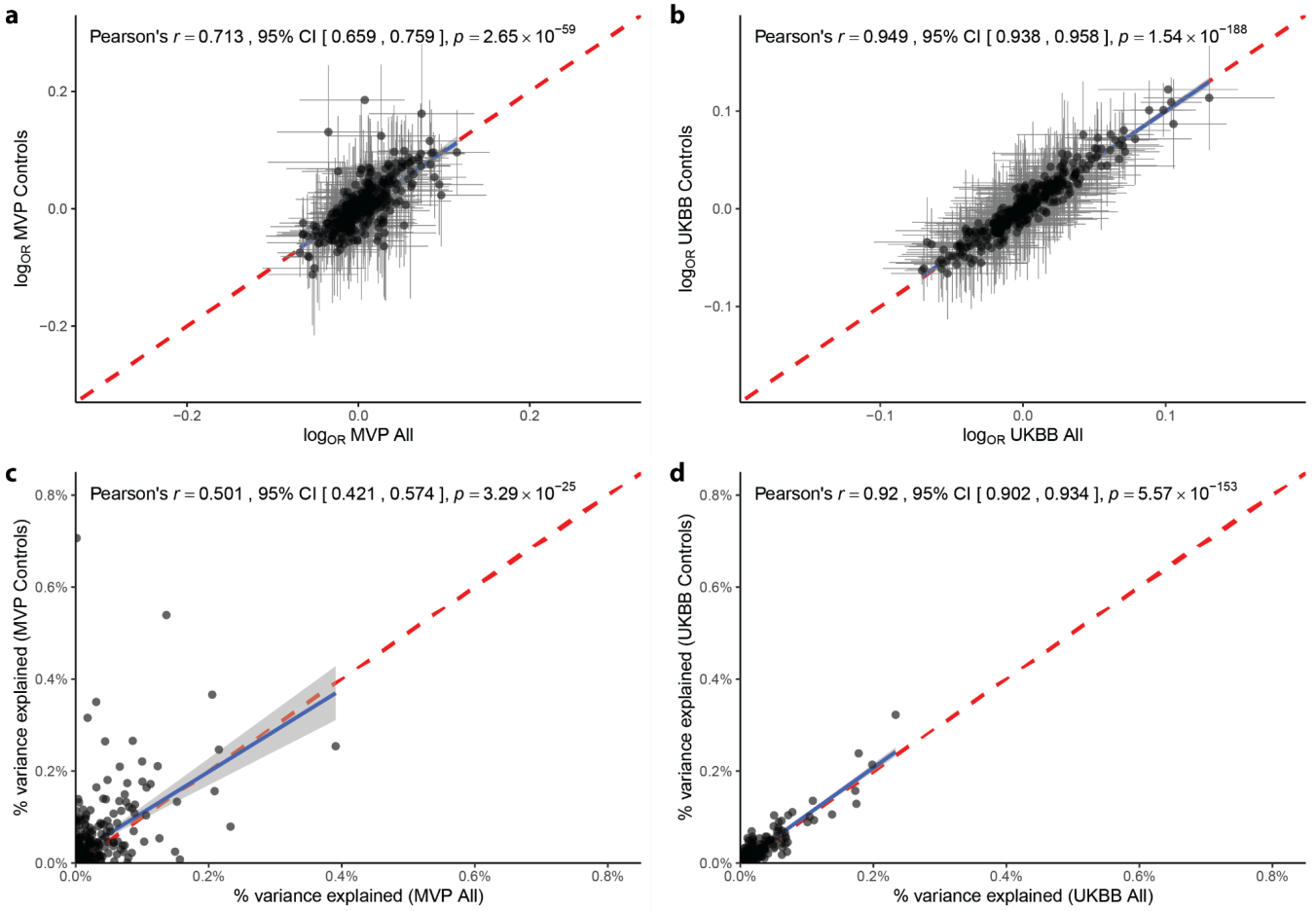
Supplementary Figure 4. Association between neuropsychiatric trait PGS and professions in the UK Biobank (UKBB). Logistic regression analysis was performed as follows: $\text{profession}_i \sim \text{PGS}_j + \text{age} + \text{age}^2 + \text{sex} + \text{ancestral PC1 to PC20} + \text{genotyping batch}$, for each of the 374 PGS_j-profession_i pairs. The CONTROLS group only comprises individuals of the ALL group without neuropsychiatric diagnoses. Plotted log odds (log(OR)) denoted by both color and size with statistical significance denoted by stars (***, ** and * correspond to Bonferroni-adjusted two-tailed p values of association equal or less than 0.001, 0.01 and 0.05 respectively).

Supplementary Figure 5



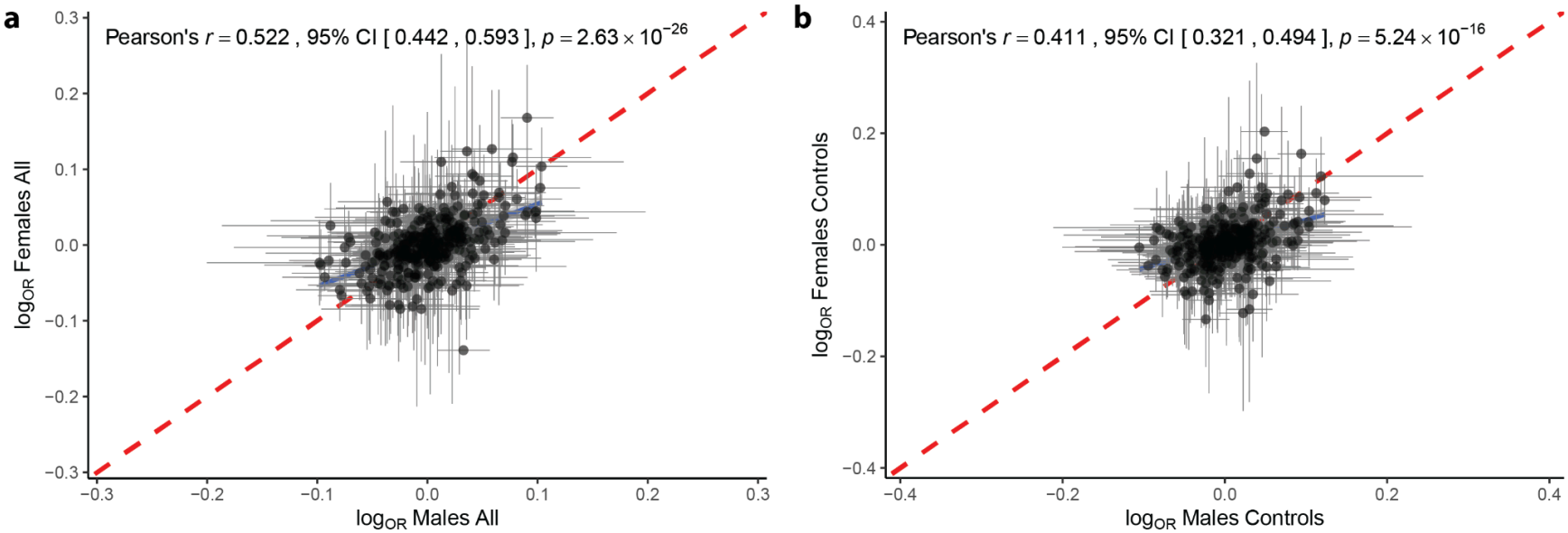
Supplementary Figure 5. Between-cohorts correlation analysis of log odds estimates (a-b) and fraction of variance explained (c-d) for trait PGS - profession pairs. We detected stronger correlation and a higher percentage of variance explained in ALL (a, c) compared to CONTROLS (b, d). Reported p values are two-tailed, n = 374. For panels (a) and (b), each dot represents the association log(OR) with corresponding 95% CI for individual cohorts. The blue line denotes the best-fit line estimated from a linear regression model, with the shaded region indicating the 95% CI for the regression line.

Supplementary Figure 6



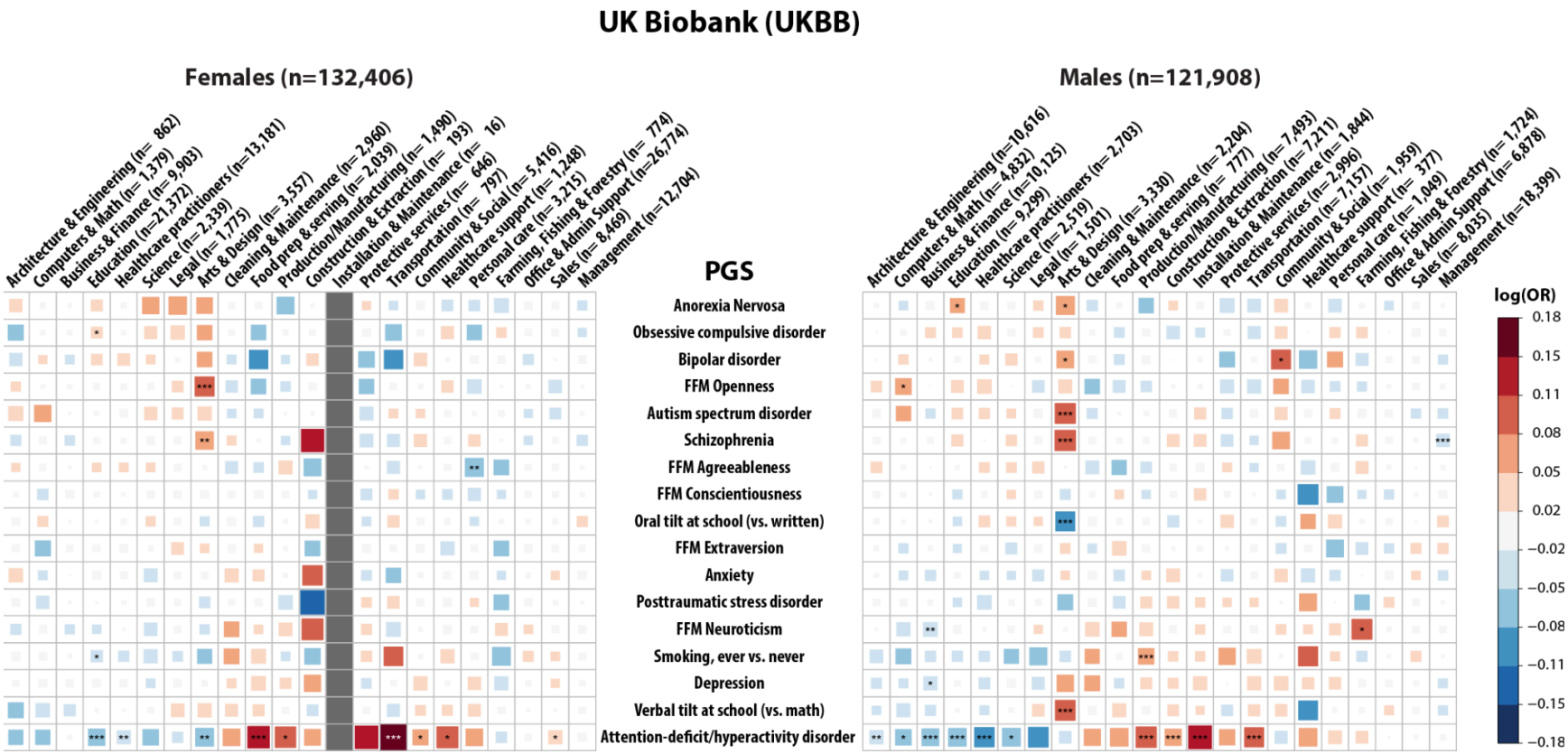
Supplementary Figure 6. Within-cohorts correlation analysis of log odds estimates (a-b) and fraction of variance explained (c-d) for PGS - profession pairs when comparing ALL and CONTROLS. The higher correlation coefficients observed in UKBB (b, d) likely reflect the greater proportion of the cohort retained after excluding individuals with neuropsychiatric diagnoses (CONTROLS) compared to MVP (a, c). Reported p values are two-tailed, n = 374. For panels (a) and (b), each dot represents the association log(OR) with corresponding 95% CI. The blue line denotes the best-fit line estimated from a linear regression model, with the shaded region indicating the 95% CI for the regression line.

Supplementary Figure 7



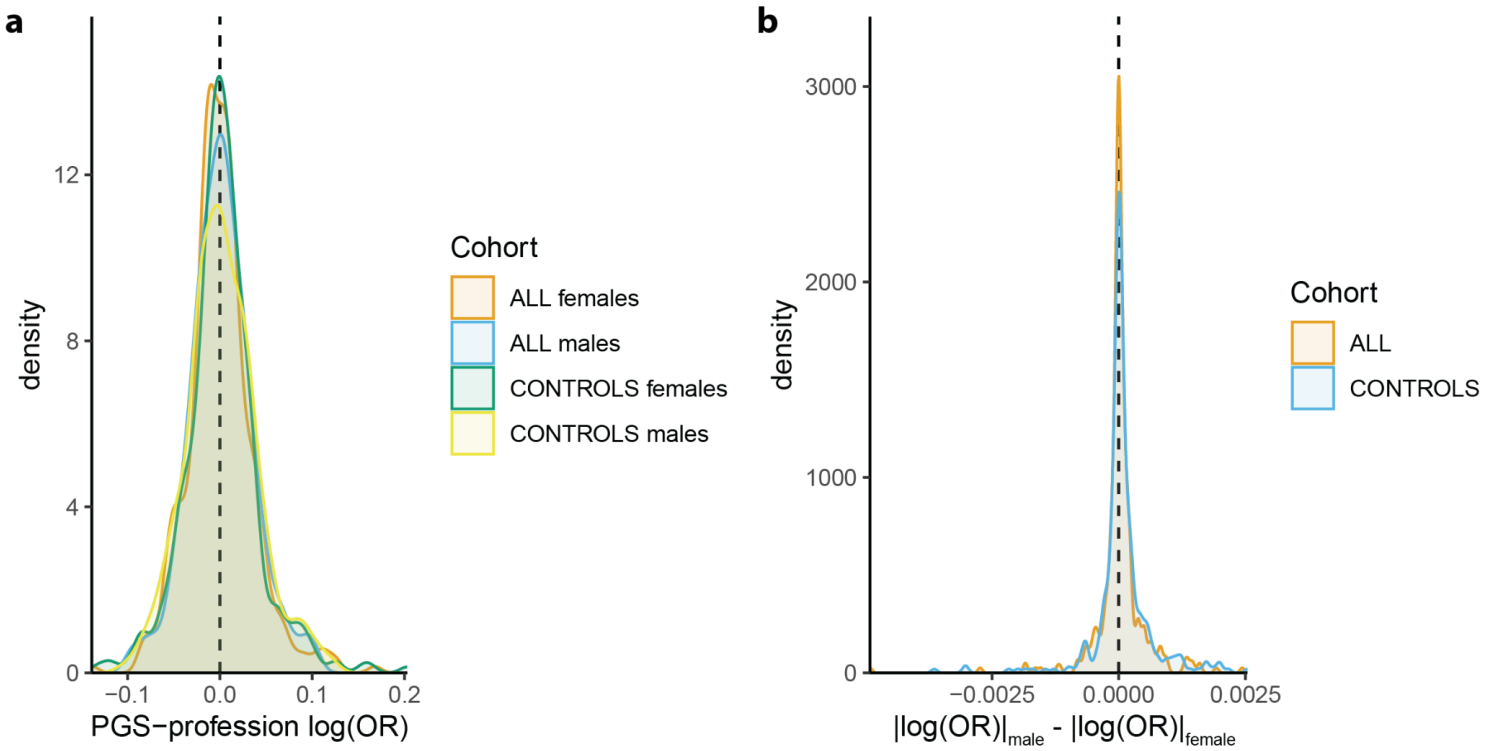
Supplementary Figure 7. Correlation analysis of log odds estimates within UKBB for PGS trait - profession pairs while considering female ("Females"; y-axis) or male ("Males"; x-axis) participants. Both "ALL" (a) and "CONTROLS" (b) are considered. Pearson's *r* and two-tailed p value are provided, n = 357. Each dot represents the association log(OR) with corresponding 95% CI. The blue line denotes the best-fit line estimated from a linear regression model, with the shaded region indicating the 95% CI for the regression line.

Supplementary Figure 8



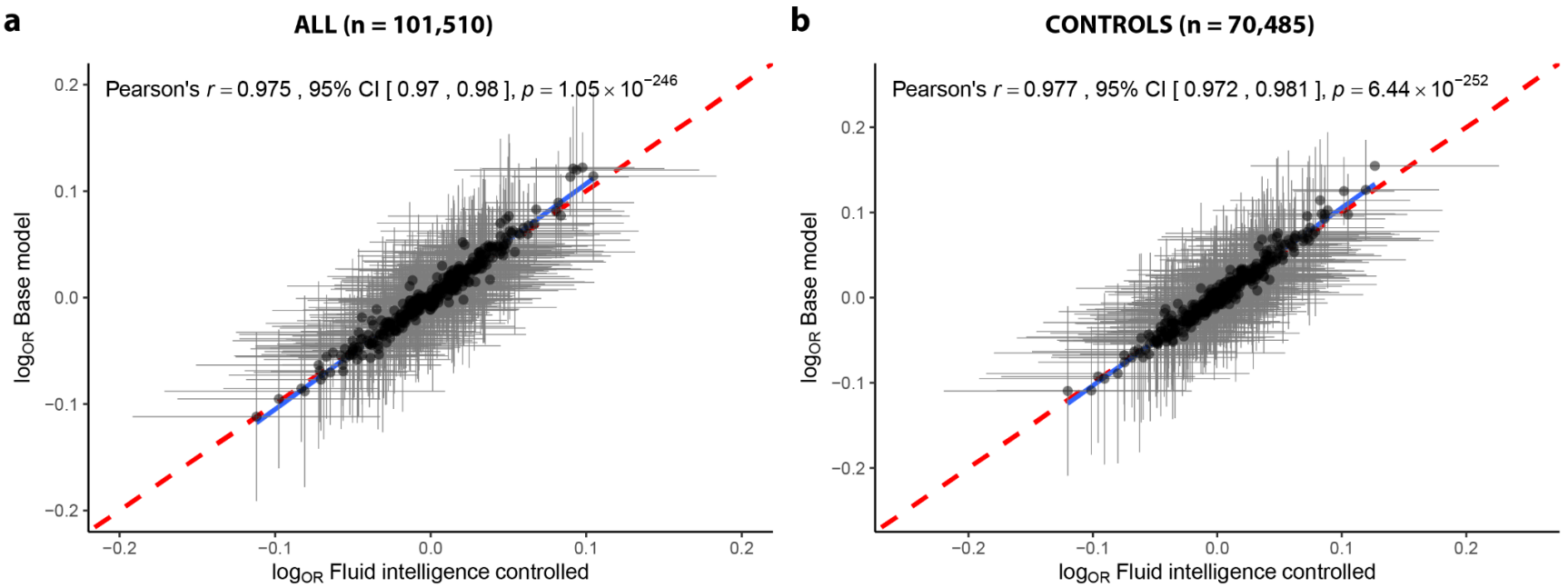
Supplementary Figure 8. Association between PGS and professions for females (left panel) and males (right panel) in the UK Biobank (UKBB). Logistic regression analysis was performed as follows: $\text{profession}_i \sim \text{PGS}_j + \text{age} + \text{age}^2 +$ ancestral PC1 to PC20 + genotyping batch, for each of the 357 and 374 PGS_j-profession_i pairs for females and males, respectively. Both analyses were performed in the ALL group using a base model without adjusting for sex. Plotted log odds (log(OR)) denoted by both color and size with statistical significance denoted by stars (***, ** and * correspond to Bonferroni-adjusted two-tailed p values of association equal or less than 0.001, 0.01 and 0.05 respectively). Grey boxes correspond to analyses that weren't well-powered ($n_{members} < 100$).

Supplementary Figure 9



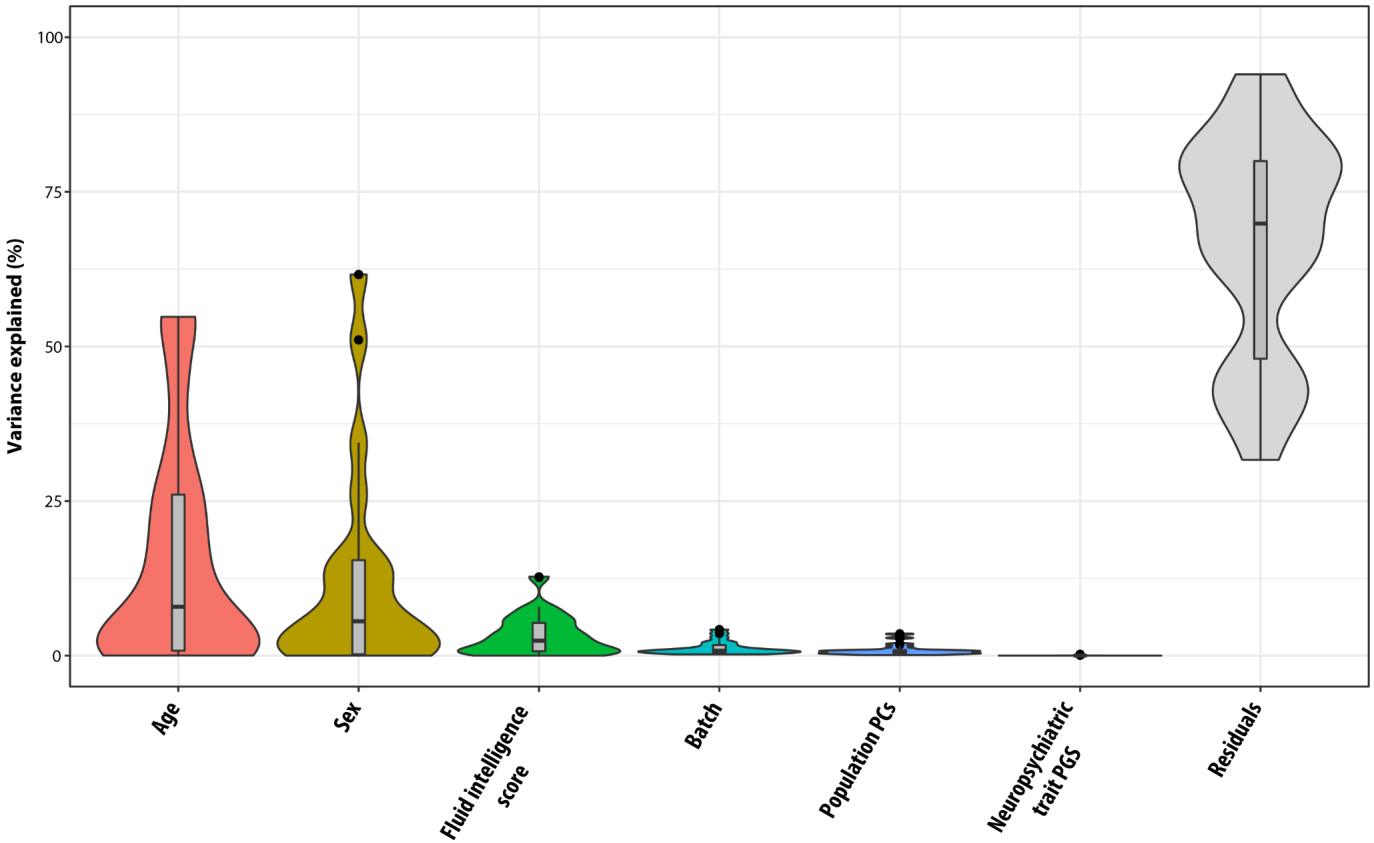
Supplementary Figure 9. Sex-specific density plots. PGS-profession associations here use a base model without adjusting for sex and are instead run for each sex separately; $\text{profession}_i \sim \text{PGS}_j + \text{age} + \text{age}^2 + \text{ancestral PC1 to PC20} + \text{genotyping batch}$, for each of the PGS_j-profession_i pairs. Sex-specific distributions of PGS-profession log(OR) for ALL and CONTROLS cohorts (a). Density plot of $|\log(OR)_{male}|$-$|\log(OR)_{female}|$ (b). Note in (b) that there is no significant difference in PGS-profession effect size estimates (ALL: two-tailed p value = 0.19, n = 235 PGS-profession pairs; CONTROLS: two-tailed p value = 0.55, n = 230 pairs; one-sample sign-test of $|\log(OR)_{male}|$-$|\log(OR)_{female}|$ where H_o is median = 0; only pairs with the same direction of effect across both females and males are considered).

Supplementary Figure 10



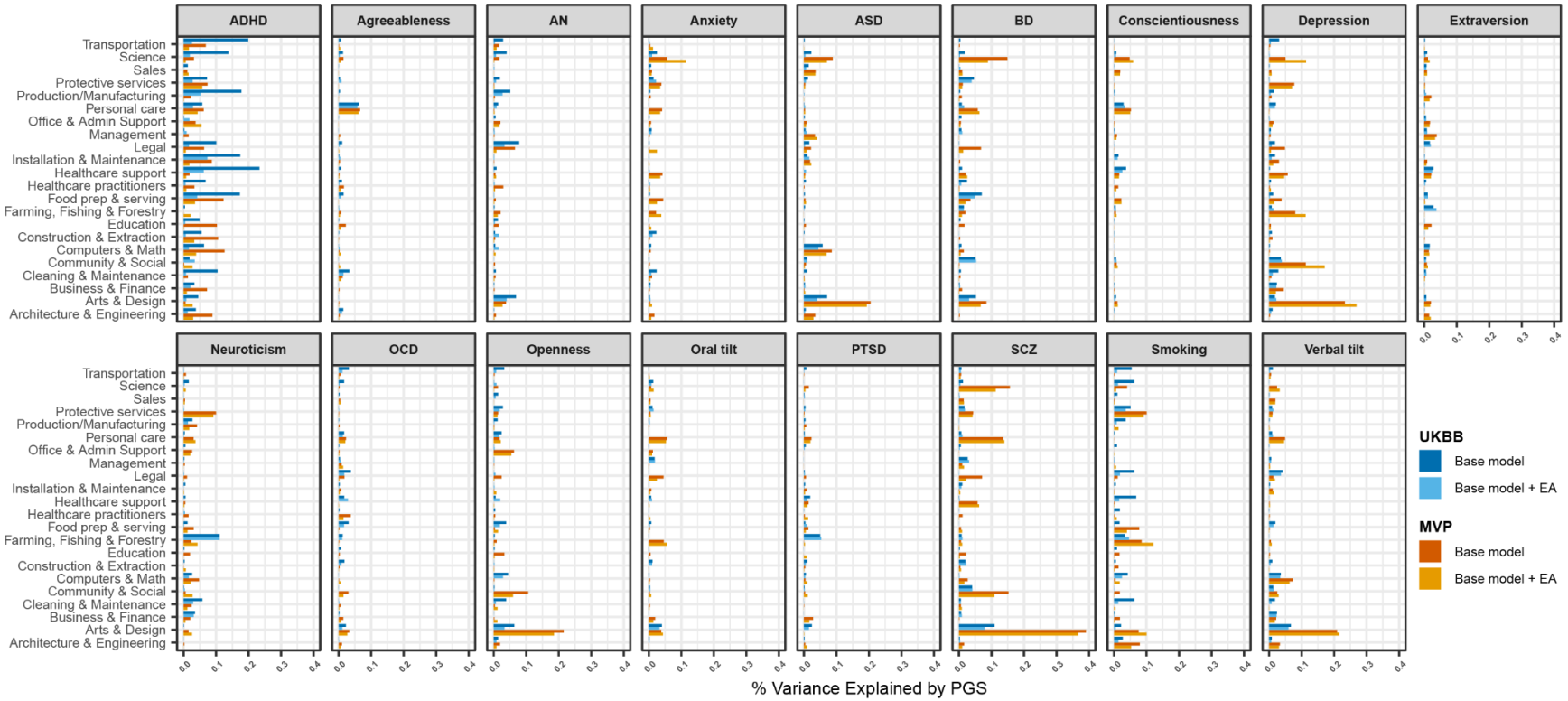
Supplementary Figure 10. Within-UKBB correlation analysis of log odds estimates for trait PGS - profession pairs while adjusting for fluid intelligence. Here, we are comparing the log odds between the UKBB base association model with (x-axis) and without (y-axis) additional adjustment for fluid intelligence for ALL (a) and CONTROLS (b). Of note, this subanalysis is only performed in individuals who have fluid intelligence information so the base model and intelligence-controlled model for each panel have the same sample size and can be directly compared. Reported p values are two-tailed, n = 374. Each dot represents the association log(OR) with corresponding 95% CI. The blue line denotes the best-fit line estimated from a linear regression model, with the shaded region indicating the 95% CI for the regression line.

Supplementary Figure 11



Supplementary Figure 11. Partitioning of variance into demographic covariates in UKBB when additionally considering fluid intelligence in the model (compare with **Fig. 5**). Fluid intelligence ranks 3rd after age (includes age and age^2) and sex. The data are visualized with Tukey-styled boxplots (line corresponds to median; box extends from the 25th to 75th percentiles; whiskers correspond to 1.5 times the interquartile range beyond the end of the box edges; dots correspond to outliers beyond the whiskers) on top of violin plots; n = 374 PGS-profession pairs.

Supplementary Figure 12

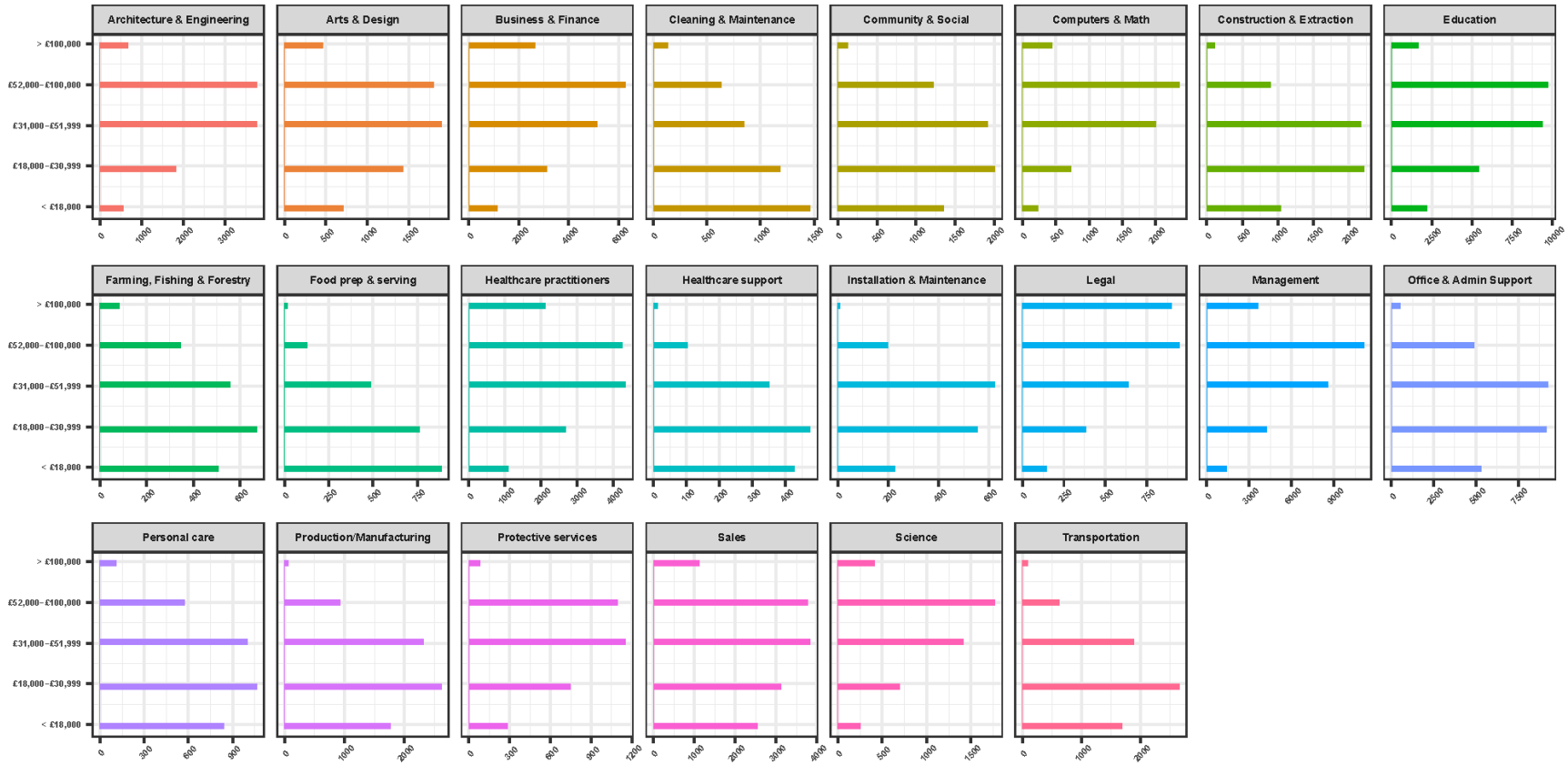


Supplementary Figure 12. Variance explained by neuropsychiatric PGS across profession categories. We are displaying results for the base model with or without adjustment for observed educational attainment in UKBB and MVP. Note the marked decrease in variance explained by ADHD PGS across profession categories.

Supplementary Figure 13a

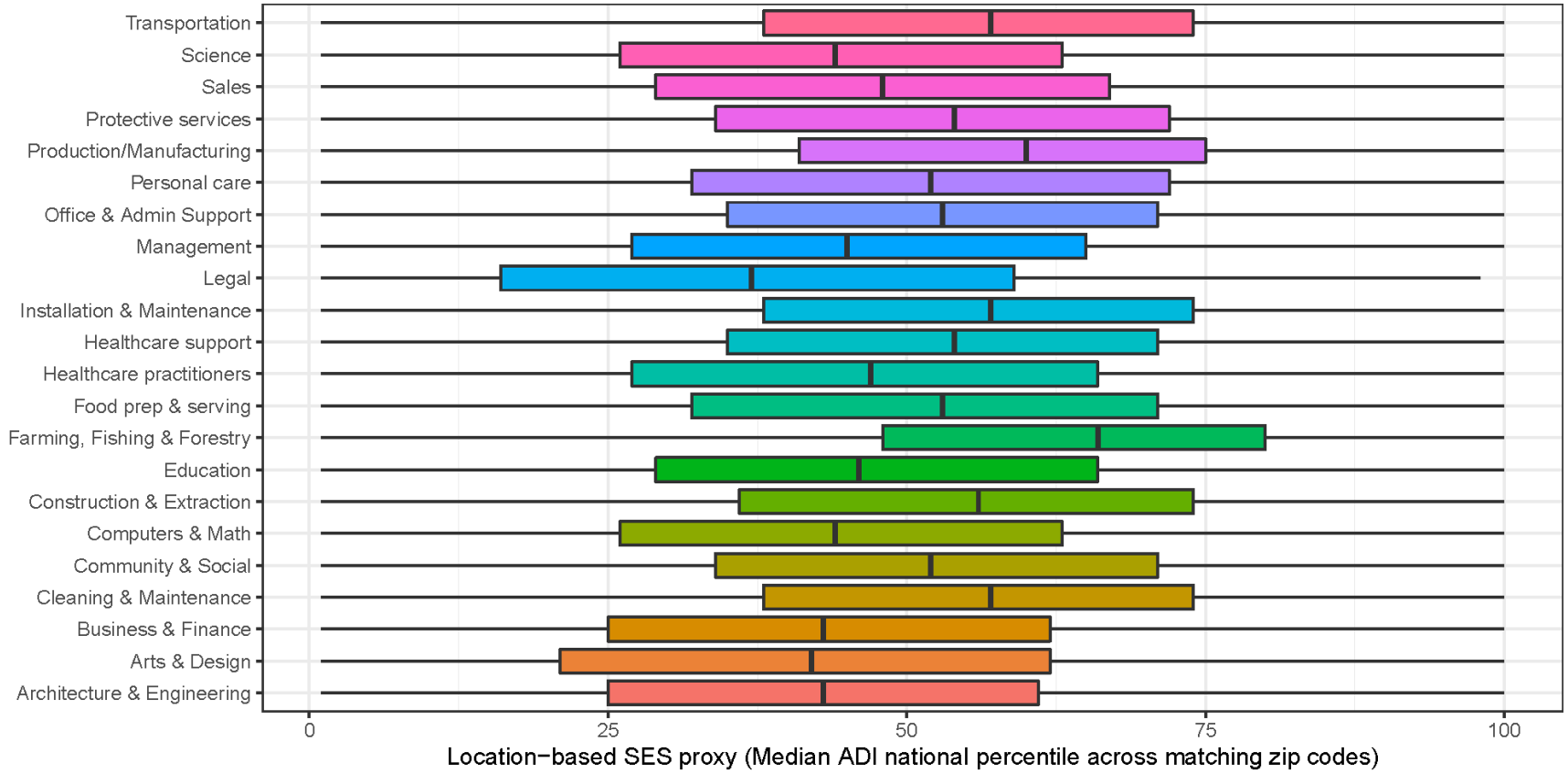
Supplementary Figure 13a. Distribution of income brackets across profession categories in MVP. Kruskal–Wallis rank sum test $\chi^2 = 13955$, df = 21, p value < 2.2×10^{-16}.

Supplementary Figure 13b



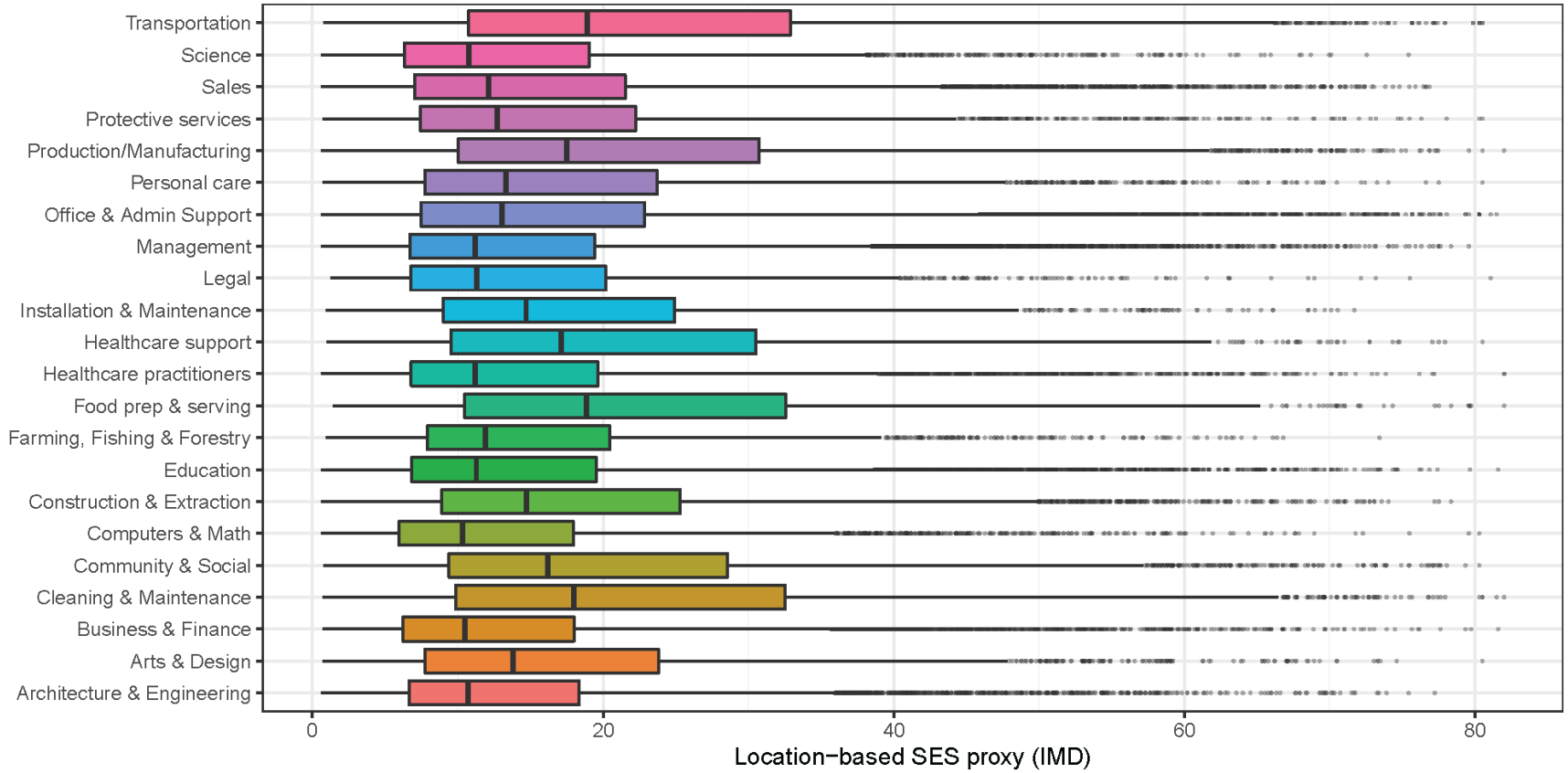
Supplementary Figure 13b. Distribution of income brackets across profession categories in UKBB. Kruskal–Wallis rank sum test $\chi^2 = 27557$, df = 21, p value < 2.2×10^{-16}.

Supplementary Figure 14a



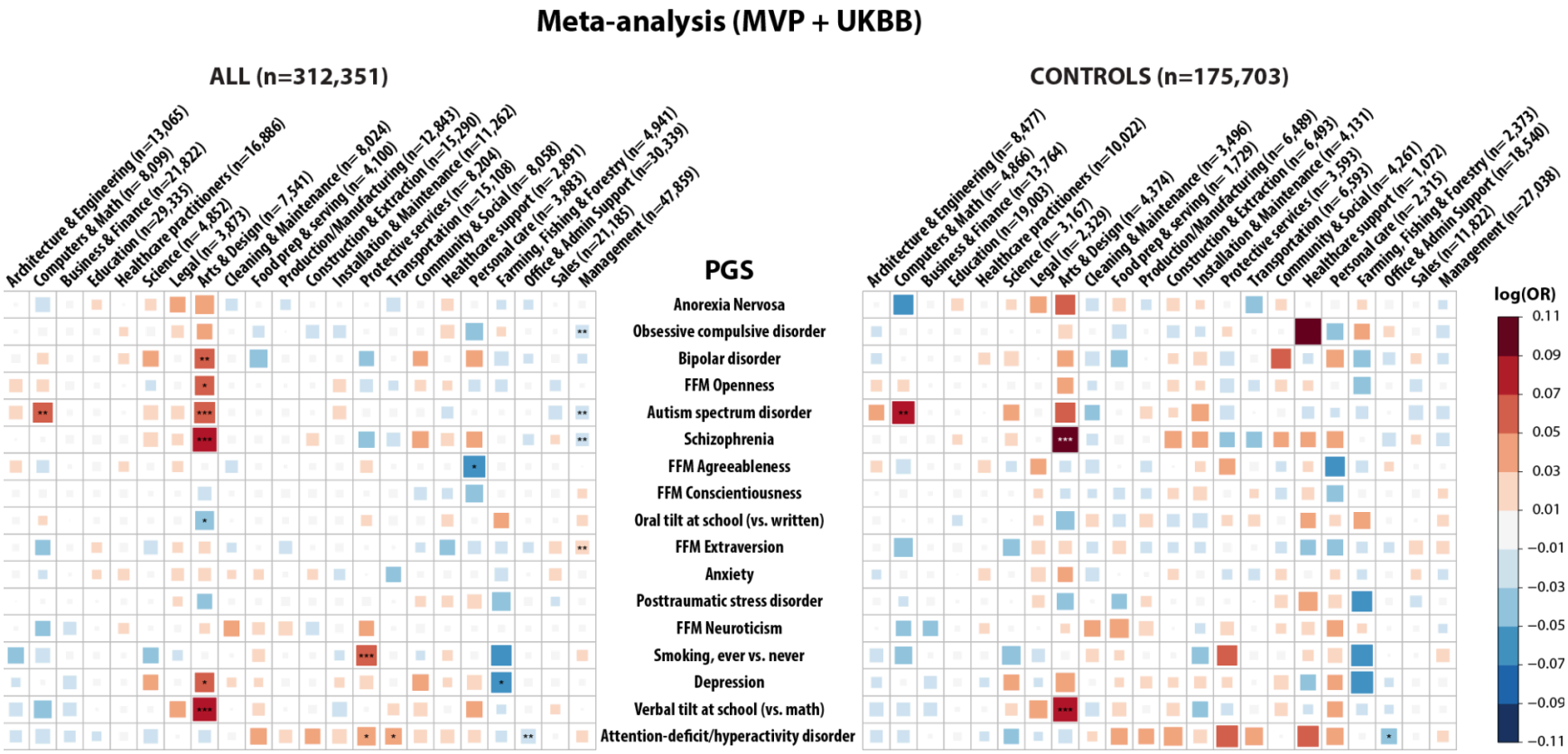
Supplementary Figure 14a. Boxplots for location-based SES in MVP. Area Deprivation Index (ADI) is visualized with Tukey-styled boxplots (line corresponds to median; box extends from the 25th to 75th percentiles; whiskers correspond to 1.5 times the interquartile range beyond the end of the box edges; dots correspond to outliers beyond the whiskers). For ADI, higher values correspond to greater levels of deprivation/disadvantage. Kruskal–Wallis rank sum test χ^2 = 4167.7, df = 21, p value < 2.2×10^{-16}, n = 116,610.

Supplementary Figure 14b



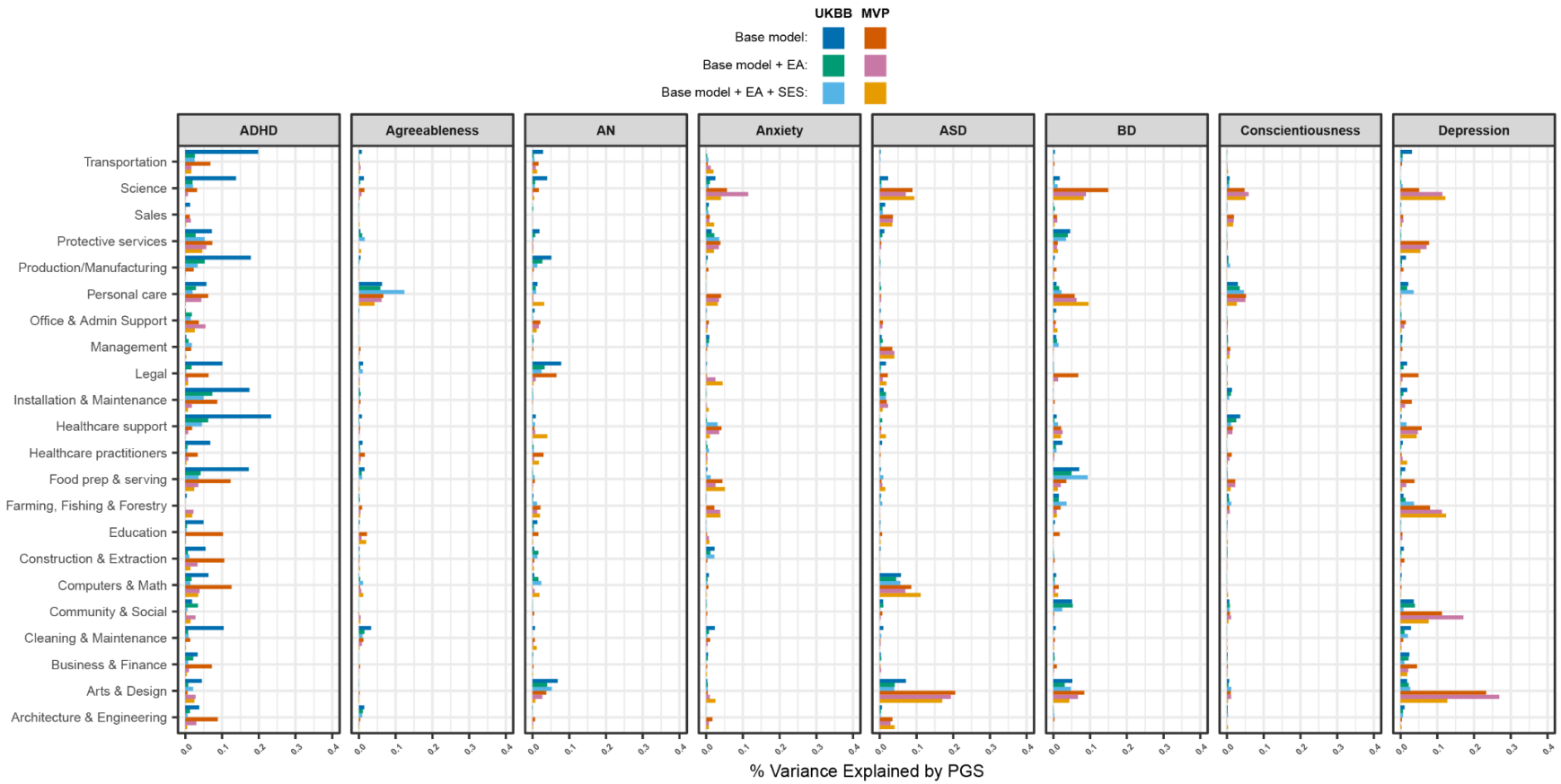
Supplementary Figure 14b. Boxplots for location-based SES in UKBB. The English Index of Multiple Deprivation (IMD) is visualized with Tukey-styled boxplots (line corresponds to median; box extends from the 25th to 75th percentiles; whiskers correspond to 1.5 times the interquartile range beyond the end of the box edges; dots correspond to outliers beyond the whiskers). For IMD, higher values correspond to greater levels of deprivation/disadvantage. This index is calculated separately for each neighborhood in England, and incorporates information related to income, employment, health and disability, education, housing, living environment, and crime. Kruskal–Wallis rank sum test χ^2 = 7877.3, df = 21, p value < 2.2×10^{-16}, n = 195,741.

Supplementary Figure 15



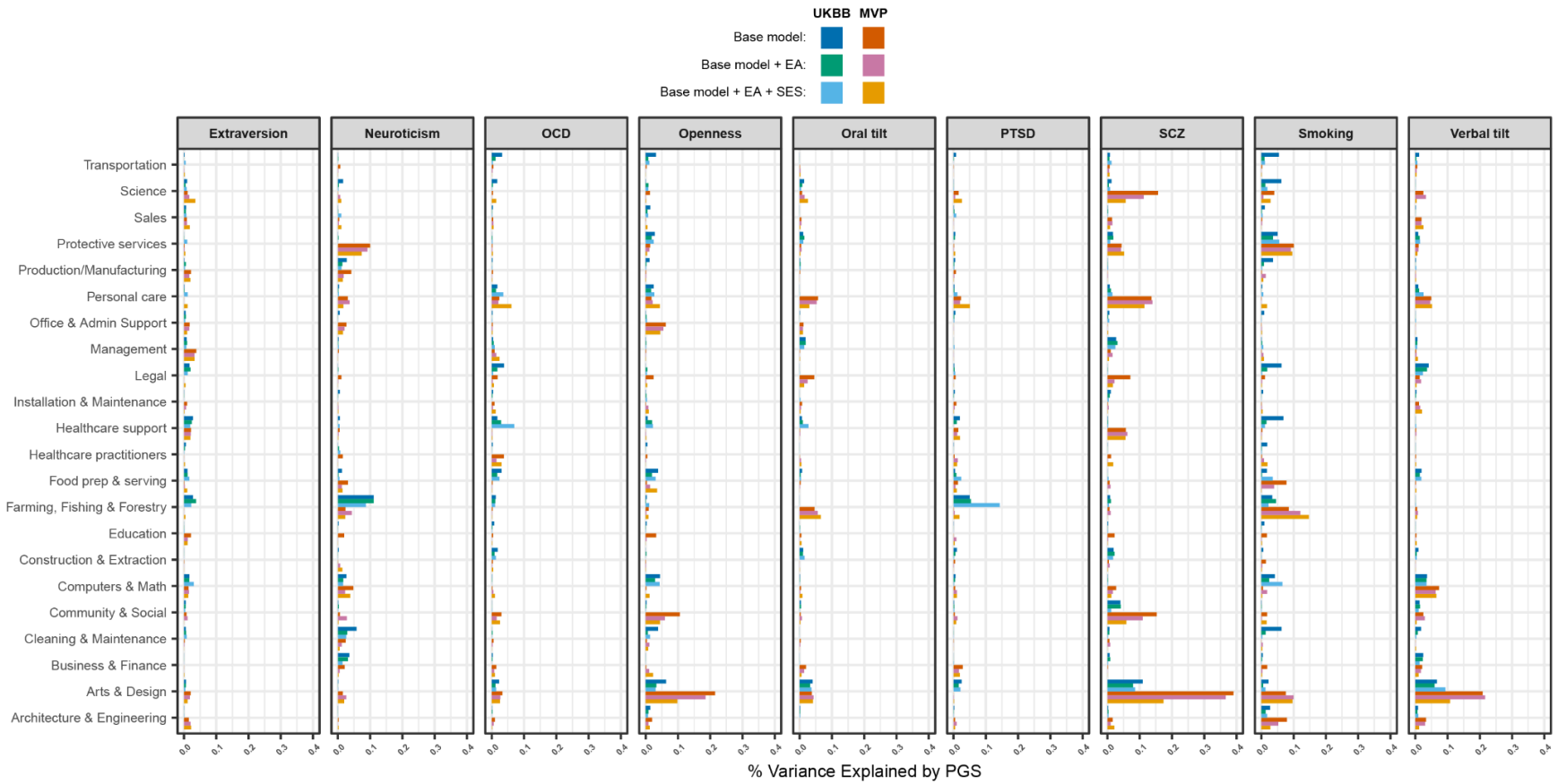
Supplementary Figure 15. Meta-analysis (fixed-effect inverse-variance weighted) of associations between neuropsychiatric trait PGS and professions while adjusting for years of education and SES proxies. The association model used here is the base model (e.g. **Fig. 1**) with additional covariates of years of education, income and location-based SES proxies, as follows: $\text{profession}_i \sim \text{PGS}_j$ + age + age^2 + sex + ancestral PC1 to PC20 + EduYears + income-range + SES-proxy + genotyping batch (needed only for UKBB), for each of the 374 PGS_j-profession_i pairs. The CONTROLS group only comprises individuals of the ALL group without a neuropsychiatric diagnosis. Plotted log odds (log(OR)) denoted by both color and size with statistical significance denoted by stars (***, ** and * correspond to Bonferroni-adjusted two-tailed p values of association equal or less than 0.001, 0.01 and 0.05 respectively, n = 374).

Supplementary Figure 16a



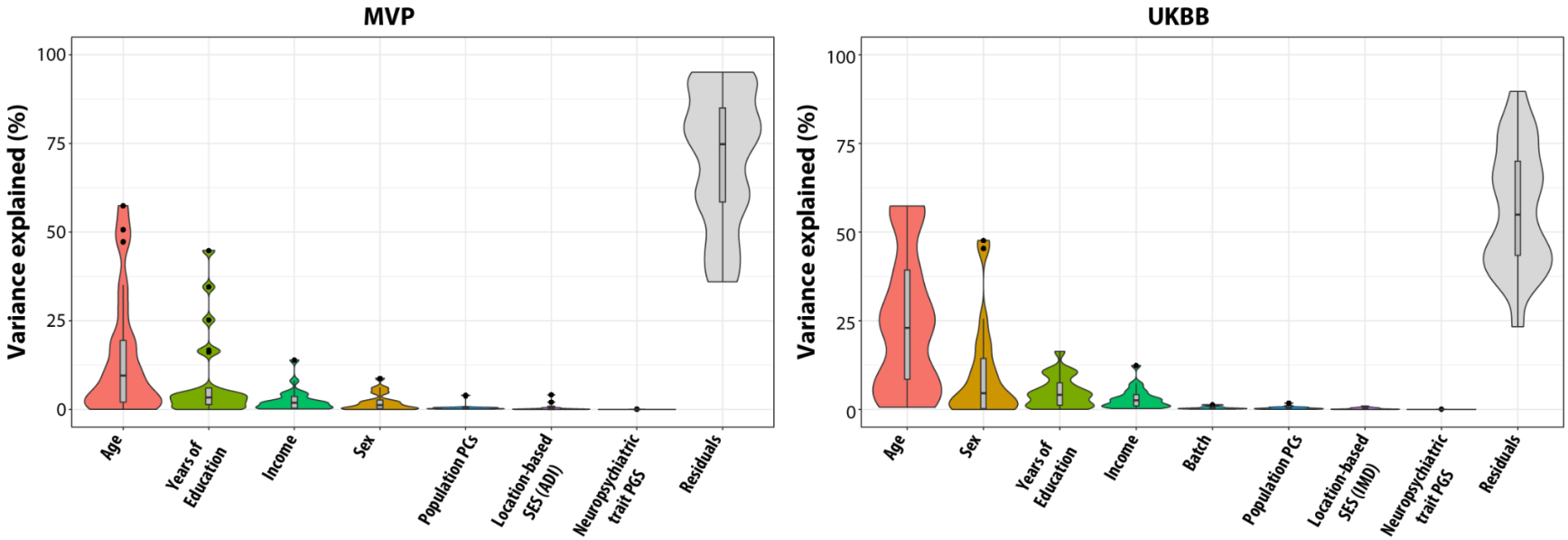
Supplementary Figure 16a. Variance explained by neuropsychiatric PGS across profession categories. We are displaying results for: 1) the base model, 2) the base model with adjustment for EA and 3) the base model with adjustment for EA, income and location-based SES. Note that most of the decrease in variance explained by ADHD PGS across profession categories happens with EA adjustment even if there is a further drop with SES.

Supplementary Figure 16b



Supplementary Figure 16b. Variance explained by neuropsychiatric PGS across profession categories. We are displaying results for: 1) the base model, 2) the base model with adjustment for EA and 3) the base model with adjustment for EA, income and location-based SES.

Supplementary Figure 17



Supplementary Figure 17. Partitioning of variance into demographic covariates in MVP and UKBB when the base model is expanded with these additional covariates: a) years of education, b) income, c) location-based SES proxies: ADI for MVP and IMD for UKBB. Compare with **Fig. 5**. The data are visualized with Tukey-styled boxplots (line corresponds to median; box extends from the 25th to 75th percentiles; whiskers correspond to 1.5 times the interquartile range beyond the end of the box edges; dots correspond to outliers beyond the whiskers) on top of violin plots; for each cohort n = 374 PGS-profession pairs.

DESCRIPTION OF SUPPLEMENTARY TABLES

Supplementary Table 1

Brief description: GWASs included for generation of PGS scores.

Column descriptions:

Column	Description
Trait	Neuropsychiatric trait
Citation	Citation for GWAS
Summary Statistics Source	Download location and file name of the GWAS summary stats file
Analysis used (MVP, UKBB, Both)	Target cohort: either “MVP” or “UKBB” or “Both”

Supplementary Table 2

Brief description: Description of sample sizes for each meta-analysis and cohort-specific analysis performed.

Column descriptions:

Column	Description
Conditions	Description of analysis conditions (covariates included, sex-specific analyses, and sample size restrictions for comparison)
MVP FULL	Sample size breakdown for full (ALL) MVP cohort-specific analyses
MVP CONTROLS	Sample size breakdown for MVP cohort-specific analyses, excluding participants with neuropsychiatric diagnoses (CONTROLS)
UKBB FULL	Sample size breakdown for full (ALL) UKBB cohort-specific analyses
UKBB CONTROLS	Sample size breakdown for UKBB cohort-specific analyses, excluding participants with

	neuropsychiatric diagnoses (CONTROLS)
Meta-Analysis FULL	Sample size breakdown for full (ALL) cohort Meta-analyses
Meta-Analysis CONTROLS	Sample size breakdown for Meta-analyses, excluding participants with neuropsychiatric diagnoses (CONTROLS)

Supplementary Table 3

Brief description: Meta-analysis (fixed-effect inverse-variance weighted) of PGS-profession association analysis for the base logistic regression model: $profession_i$ ~ PGS_j + age + age^2 + sex + ancestral PC1 to PC20 + genotyping batch (needed only for UKBB), for each of the 374 PGS_j-$profession_i$ pairs)

Column descriptions:

Column	Description
group	Either “FULL” for ALL cohort or “CONTROLS” for CONTROLS cohort
cohort	“MA” stands for Meta-analysis
trait_common_name	Neuropsychiatric trait name used for PGS score generation
Estimate	Effect size (log(OR))
Std.Error	Standard error of effect size
Pr(>|z|)	Two-tailed p value
z value	Z-score
n	Sample size
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Worktype	Profession category
n_controls	Sample size of individuals not within profession category

n_cases	Sample size of individuals within profession category
q	Bonferroni-adjusted two-tailed p value

Supplementary Tables 4-7

PGS-profession association analysis for:

- MVP ALL cohort results from the base logistic regression model (Supplementary table 4): $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20.
- UKBB ALL cohort results from the base logistic regression model (Supplementary table 5): $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20 + genotyping batch.
- MVP CONTROLS cohort results from the base logistic regression model (Supplementary table 6): $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20.
- UKBB CONTROLS cohort results from the base logistic regression model (Supplementary table 7): $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20 + genotyping batch.

Column descriptions:

Column	Description
bim	Genotype file used (for internal tracking)
trait	Abbreviated neuropsychiatric trait name used for PGS score generation (for internal tracking)
trait.simple	Short neuropsychiatric trait name used for PGS score generation
phi	PRS-CS global shrinkage parameter phi. If “auto” then phi was learned from the data using a fully Bayesian approach.
pheno	Profession category abbreviation
Estimate	Effect size (log(OR))
OR	Odds ratio
Std.Error	Standard error of effect size
z value	Z-score

Pr(>|z|)	Two-tailed p value
n_controls	Sample size of individuals not within profession category
n_cases	Sample size of individuals within profession category
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Nagelkerke	Nagelkerke R^2 for full model
var_PRS	Variance explained by specific neuropsychiatric trait PGS
var_PCi	Variance explained by PC i from ancestral PCA (where i is 1 to 20)
var_Batch	Variance explained by genotyping batch (only for UKBB)
var_sex	Variance explained by sex
var_age	Variance explained by age
var_age.sq	Variance explained by age^2
var_Residuals	Variance explained by residuals of the model
trait.full	Full neuropsychiatric trait name used for PGS score generation
Worktype	Profession category name
q	Bonferroni-adjusted two-tailed p value

Supplementary Table 8

Brief description: Bipolar disorder (BD) PGS-PheCode association analysis for expanded MVP ALL cohort with logistic regression analysis: BD diagnosis ~ BD PGS + age + age^2 + sex + ancestral PC1 to PC20.

Column descriptions:

Column	Description

bim	Genotype file used (for internal tracking)
trait	Trait GWAS used for PGS estimation
phi	PRS-CS global shrinkage parameter phi. If “auto” then phi was learned from the data using a fully Bayesian approach.
pheno	PheCode of respective neuropsychiatric disorder
Estimate	Effect size (log(OR))
OR	Odds ratio
Std.Error	Standard error of effect size
z value	Z-score
Pr(>|z|)	Two-tailed p value
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Nagelkerke	Nagelkerke R^2 for full model
trait.full	Full neuropsychiatric trait name used for PGS score generation

Supplementary Table 9

Brief description: Prevalence of neuropsychiatric diagnoses in MVP and UKBB

Column descriptions:

Column	Description
phecode	PheCode
phenotype	Phenotype description
Broad category	Phenotype group
GROUP	More DSM-compatible grouping designation
MVP	% for MVP cohort
UKBB	% for UKBB cohort

Supplementary Tables 10-13

Sex-specific PGS-profession association analysis for:

- UKBB ALL cohort results from female-only base logistic regression model (Supplementary table 10): $profession_i \sim PGS_j + age + age^2 +$ ancestral PC1 to PC20 + genotyping batch
- UKBB ALL cohort results from male-only base logistic regression model (Supplementary table 11): $profession_i \sim PGS_j + age + age^2 +$ ancestral PC1 to PC20 + genotyping batch
- UKBB CONTROLS cohort results from female-only base logistic regression model (Supplementary table 12): $profession_i \sim PGS_j + age + age^2 +$ ancestral PC1 to PC20 + genotyping batch
- UKBB CONTROLS cohort results from male-only base logistic regression model (Supplementary table 13): $profession_i \sim PGS_j + age + age^2 +$ ancestral PC1 to PC20 + genotyping batch

Column descriptions:

Column	Description
bim	Genotype file used (for internal tracking)
trait	Abbreviated neuropsychiatric trait name used for PGS score generation (for internal tracking)
trait.simple	Short neuropsychiatric trait name used for PGS score generation
phi	PRS-CS global shrinkage parameter phi. If “auto” then phi was learned from the data using a fully Bayesian approach.
pheno	Profession category abbreviation
Estimate	Effect size (log(OR))
OR	Odds ratio
Std.Error	Standard error of effect size
z value	Z-score
Pr(>|z|)	Two-tailed p value
n_controls	Sample size of individuals not within profession category

n_cases	Sample size of individuals within profession category
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Nagelkerke	Nagelkerke R^2 for full model
var_PRS	Variance explained by specific neuropsychiatric trait PGS
var_PCi	Variance explained by PC i from ancestral PCA (where i is 1 to 20)
var_Batch	Variance explained by genotyping batch
var_age	Variance explained by age
var_age.sq	Variance explained by age^2
var_Residuals	Variance explained by residuals of the model
trait.full	Full neuropsychiatric trait name used for PGS score generation
Worktype	Profession category name
q	Bonferroni-adjusted two-tailed p value

Supplementary Tables 14-21

PGS-profession association analysis for:

- UKBB ALL cohort results from fluid intelligence-adjusted logistic regression model (Supplementary table 14): $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20 + genotyping batch + fluid intelligence.
- UKBB ALL cohort results from base logistic regression model, with sample size restricted to match fluid intelligence-adjusted logistic regression model (Supplementary table 15): $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20 + genotyping batch.
- UKBB CONTROLS cohort results from fluid intelligence-adjusted logistic regression model (Supplementary table 16): $profession_i \sim PGS_j + age + age^2 +$ sex + ancestral PC1 to PC20 + genotyping batch + fluid intelligence.
- UKBB CONTROLS cohort results from base logistic regression model, with sample size restricted to match fluid intelligence-adjusted logistic regression

model (Supplementary table 17): $profession_i$ ~ PGS_j + age + age^2 + sex + ancestral PC1 to PC20 + genotyping batch.
- MVP ALL cohort results from EduYears-adjusted logistic regression model (Supplementary table 18): $profession_i$ ~ PGS_j + age + age^2 + sex + ancestral PC1 to PC20 + EduYears.
- UKBB ALL cohort results from EduYears-adjusted logistic regression model (Supplementary table 19): $profession_i$ ~ PGS_j + age + age^2 + sex + ancestral PC1 to PC20 + genotyping batch + EduYears.
- MVP CONTROLS cohort results from EduYears-adjusted logistic regression model (Supplementary table 20): $profession_i$ ~ PGS_j + age + age^2 + sex + ancestral PC1 to PC20 + EduYears.
- UKBB CONTROLS cohort results from EduYears-adjusted logistic regression model (Supplementary table 21): $profession_i$ ~ PGS_j + age + age^2 + sex + ancestral PC1 to PC20 + genotyping batch + EduYears.

Column descriptions:

Column	Description
bim	Genotype file used (for internal tracking)
trait	Abbreviated neuropsychiatric trait name used for PGS score generation (for internal tracking)
trait.simple	Short neuropsychiatric trait name used for PGS score generation
phi	PRS-CS global shrinkage parameter phi. If “auto” then phi was learned from the data using a fully Bayesian approach.
pheno	Profession category abbreviation
Estimate	Effect size (log(OR))
OR	Odds ratio
Std.Error	Standard error of effect size
z value	Z-score
Pr(>|z|)	Two-tailed p value
n_controls	Sample size of individuals not within profession category

n_cases	Sample size of individuals within profession category
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Nagelkerke	Nagelkerke R^2 for full model
var_PRS	Variance explained by specific neuropsychiatric trait PGS
var_PCi	Variance explained by PC i from ancestral PCA (where i is 1 to 20)
var_EduYears	Variance explained by observed years of education (only for EduYears-adjusted models)
var_Batch	Variance explained by genotyping batch (only for UKBB)
var_fluid_intelligence	Variance explained by fluid intelligence measure (only for fluid intelligence-adjusted models)
var_sex	Variance explained by sex
var_age	Variance explained by age
var_age.sq	Variance explained by age^2
var_Residuals	Variance explained by residuals of the model
trait.full	Full neuropsychiatric trait name used for PGS score generation
Worktype	Profession category name
q	Bonferroni-adjusted two-tailed p value

Supplementary Tables 22-23

ADHD PGS-profession mediation analysis (see Methods) for:

- MVP CONTROLS cohort results (Supplementary table 22)
- UKBB CONTROLS cohort results (Supplementary table 23)

Worktype	Profession category name

n_controls	Sample size of individuals not within profession category
n_cases	Sample size of individuals within profession category
Estimate_Total Effect	Effect size of total effect of ADHD PGS on profession category membership
95% CI Lower_Total Effect	2.5% Confidence interval of total effect of ADHD PGS on profession category membership
95% CI Upper_Total Effect	97.5% Confidence interval of total effect of ADHD PGS on profession category membership
p-value_Total Effect	Two-tailed p value of total effect of ADHD PGS on profession category membership
Estimate_ACME (average)	Effect size of average causal mediation effect of ADHD PGS on profession category membership mediated through EduYears
95% CI Lower_ACME (average)	2.5% Confidence interval of average causal mediation effect of ADHD PGS on profession category membership mediated through EduYears
95% CI Upper_ACME (average)	97.5% Confidence interval of average causal mediation effect of ADHD PGS on profession category membership mediated through EduYears
p-value_ACME (average)	Two-tailed p value of average causal mediation effect of ADHD PGS on profession category membership mediated through EduYears
Estimate_ADE (average)	Effect size of average direct effect of ADHD PGS on profession category membership
95% CI Lower_ADE (average)	2.5% Confidence interval of average direct effect of ADHD PGS on profession category membership
95% CI Upper_ADE (average)	97.5% Confidence interval of average direct effect of ADHD PGS on profession category membership
p-value_ADE (average)	Two-tailed p value of average direct effect of

	ADHD PGS on profession category membership
Estimate_Prop. Mediated (average)	Estimated proportion of effect mediated by EduYears
95% CI Lower_Prop. Mediated (average)	2.5% Confidence interval of proportion of effect mediated by EduYears
95% CI Upper_Prop. Mediated (average)	97.5% Confidence interval of proportion of effect mediated by EduYears
p-value_Prop. Mediated (average)	Two-tailed p value of proportion of effect mediated by EduYears
q-value_Total Effect	Bonferroni-adjusted two-tailed p value of total effect of ADHD PGS on profession category membership
q-value_ACME (average)	Bonferroni-adjusted two-tailed p value of average causal mediation effect of ADHD PGS on profession category membership mediated through EduYears
q-value_ADE (average)	Bonferroni-adjusted two-tailed p value of average direct effect of ADHD PGS on profession category membership
q-value_Prop. Mediated (average)	Bonferroni-adjusted two-tailed p value of proportion of effect mediated by EduYears

Supplementary Table 24

Brief description: Meta-analysis (fixed-effect inverse-variance weighted) of PGS-profession associations in MVP and UKBB as estimated by the following logistic regression: $profession_i \sim PGS_j + age + age^2 + sex +$ ancestral PC1 to PC20 + genotyping batch (for UKBB) + EduYears + income range + SES-proxy.

Column descriptions:

Column	Description
group	Either “FULL” for ALL cohort or “CONTROLS” for CONTROLS cohort
cohort	“MA” stands for Meta-analysis
trait_common_name	Neuropsychiatric trait name used for PGS score

	generation
Estimate	Effect size (log(OR))
Std.Error	Standard error of effect size
Pr(>|z|)	Two-tailed p value
z value	Z-score
n	Sample size
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Worktype	Profession category
n_controls	Sample size of individuals not within profession category
n_cases	Sample size of individuals within profession category
q	Bonferroni-adjusted two-tailed p value

Supplementary Tables 25-28

PGS-profession association analysis for:

- MVP ALL cohort results from the EduYears, income, and SES proxy-adjusted logistic regression model (Supplementary table 25): $profession_i \sim PGS_j$ + age + age^2 + sex + ancestral PC1 to PC20 + EduYears + income range + SES-proxy
- UKBB ALL cohort results from the EduYears, income, and SES proxy-adjusted logistic regression model (Supplementary table 26): $profession_i \sim PGS_j$ + age + age^2 + sex + ancestral PC1 to PC20 + genotyping batch + EduYears + income range + SES-proxy
- MVP CONTROLS cohort results from the EduYears, income, and SES proxy-adjusted logistic regression model (Supplementary table 27): $profession_i \sim PGS_j$ + age + age^2 + sex + ancestral PC1 to PC20 + EduYears + income range + SES-proxy
- UKBB CONTROLS cohort results from the EduYears, income, and SES proxy-adjusted logistic regression model (Supplementary table 28): $profession_i \sim PGS_j$ + age + age^2 + sex + ancestral PC1 to PC20 + genotyping batch + EduYears + income range + SES-proxy

Column descriptions:

Column	Description
bim	Genotype file used (for internal tracking)
trait	Abbreviated neuropsychiatric trait name used for PGS score generation (for internal tracking)
trait.simple	Short neuropsychiatric trait name used for PGS score generation
phi	PRS-CS global shrinkage parameter phi. If “auto” then phi was learned from the data using a fully Bayesian approach.
pheno	Profession category abbreviation
Estimate	Effect size (log(OR))
OR	Odds ratio
Std.Error	Standard error of effect size
z value	Z-score
Pr(>|z|)	Two-tailed p value
n_controls	Sample size of individuals not within profession category
n_cases	Sample size of individuals within profession category
2.5 %	2.5% Confidence interval
97.5 %	97.5% Confidence interval
Nagelkerke	Nagelkerke R^2 for full model
var_PRS	Variance explained by specific neuropsychiatric trait PGS
var_PCi	Variance explained by PC i from ancestral PCA (where i is 1 to 20)
var_income	Variance explained by self-reported income range
var_ADI	Variance explained by Area Deprivation Index (only for MVP)

var_IMD	Variance explained by English Index of Multiple Deprivation (only for UKBB)
var_EduYears	Variance explained by observed years of education (only for EduYears-adjusted models)
var_Batch	Variance explained by genotyping batch (only for UKBB)
var_sex	Variance explained by sex
var_age	Variance explained by age
var_age.sq	Variance explained by age^2
var_Residuals	Variance explained by residuals of the model
trait.full	Full neuropsychiatric trait name used for PGS score generation
Worktype	Profession category name
q	Bonferroni-adjusted two-tailed p value

Supplementary Table 29

Brief description: Occupation information in the MVP's "Lifestyle Survey" questionnaire. Table is provided as in the questionnaire.

Supplementary Table 30

Brief description: Mapping from MVP "Baseline Survey" questionnaire to ISCED level to EduYears phenotype.

Supplementary Table 31

Brief description: Income range information from the MVP's "Baseline Survey" and "COVID-19 Survey". Table is provided as in the questionnaires.

Supplementary Table 32

Brief description: Mapping from UKBB self-reported baseline diagnoses codes to phecodes.

Column descriptions:

Column	Description
Diagnosis code	UKBB self-reported baseline diagnosis code
Diagnosis code description	Description of UKBB self-reported baseline diagnosis code
Phecode	PheCode it maps to
Phecode description	Description of PheCode it maps to

Supplementary Table 33

Brief description: Mapping from UKBB Mental Health Questionnaire codes to phecodes.

Column descriptions:

Column	Description
Mental Health Questionnaire code	UKBB mental health questionnaire diagnosis code
Code description	Description of UKBB mental health questionnaire diagnosis code
Phecode	PheCode it maps to
Phecode description	Description of PheCode it maps to

Supplementary Table 34

Brief description: Mapping from UKBB job code reported at baseline Verbal Interview to MVP occupation category.

Column descriptions:

Column	Description
Job code	UKBB Job code
Job code description	Description of UKBB Job code
MVP occupation category	Corresponding MVP profession category (NA if not mapped to a MVP category used in this study)

Supplementary Table 35

Brief description: Mapping from UKBB Baseline questionnaire to ISCED level to EduYears phenotype.

Supplementary Table 36

Brief description: Income range information from the UKBB's Touch-Screen Questionnaire. Table is provided as in the questionnaire.

VA Million Veteran Program Core Acknowledgement

MVP Program Office

- Program Director - Sumitra Muralidhar, Ph.D.

US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Associate Director, Scientific Programs - Jennifer Moser, Ph.D.

US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Associate Director, Cohort Management & Public Relations - Jennifer E. Deen, B.S.

US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

MVP Executive Committee

- Co-Chair: J. Michael Gaziano, M.D., M.P.H.

VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Co-Chair: Sumitra Muralidhar, Ph.D.

US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Jean Beckham, Ph.D.

Durham VA Medical Center, 508 Fulton Street, Durham, NC 27705

- Kyong-Mi Chang, M.D.

Philadelphia VA Medical Center, 3900 Woodland Avenue, Philadelphia, PA 19104

- Philip S. Tsao, Ph.D.

VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

- Shiuh-Wen Luoh, M.D., Ph.D.

VA Portland Health Care System, 3710 SW US Veterans Hospital Rd, Portland, OR 97239

US Department of Veterans Affairs, 810 Vermont Avenue NW, Washington, DC 20420

- Juan P. Casas, M.D., Ph.D., Ex-Officio

VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

MVP Principal Investigators

- J. Michael Gaziano, M.D., M.P.H.

VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130

- Philip S. Tsao, Ph.D.

VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304

MVP Operations

- MVP Executive Director – Juan P. Casas, M.D., Ph.D.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Director of Regulatory Affairs – Lori Churby, B.S.
VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
- MVP Cohort Management Director – Stacey B. Whitbourne, Ph.D.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- MVP Recruitment/Enrollment Director - Jessica V. Brewer, M.P.H.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Director, VA Central Biorepository, Boston – Mary T. Brophy M.D., M.P.H.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Executive Director for MVP Biorepositories - Luis E. Selva, Ph.D.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- MVP Informatics, Boston – Shahpoor (Alex) Shayan, M.S.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Director, MVP Data Operations/Analytics, Boston – Kelly Cho, M.P.H., Ph.D.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Director, Center for Computational and Data Science (C-DACS) & Genomics Core – Saiju Pyarajan Ph.D.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Director, Molecular Data Core – Philip S. Tsao, Ph.D.
VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
- Director, Phenomics Data Core – Kelly Cho, M.P.H, Ph.D.
VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
- Director, VA Informatics and Computing Infrastructure (VINCI) – Scott L. DuVall, Ph.D.
VA Salt Lake City Health Care System, 500 Foothill Drive, Salt Lake City, UT 84148
- MVP Coordinating Centers
    - o Cooperative Studies Program Clinical Research Pharmacy Coordinating Center, Albuquerque – Todd Connor, Pharm.D.; Dean P. Argyres, B.S., M.S.
    New Mexico VA Health Care System, 1501 San Pedro Drive SE, Albuquerque, NM 87108
    - o Genomics Coordinating Center, Palo Alto – Philip S. Tsao, Ph.D.
    VA Palo Alto Health Care System, 3801 Miranda Avenue, Palo Alto, CA 94304
    - o MVP Boston Coordinating Center, Boston - J. Michael Gaziano, M.D., M.P.H.
    VA Boston Healthcare System, 150 S. Huntington Avenue, Boston, MA 02130
    - o MVP Information Center, Canandaigua – Brady Stephens, M.S.
    Canandaigua VA Medical Center, 400 Fort Hill Avenue, Canandaigua, NY 14424

Current MVP Local Site Investigators

- Atlanta VA Medical Center (Peter Wilson, M.D.)
1670 Clairmont Road, Decatur, GA 30033
- Bay Pines VA Healthcare System (Rachel McArdle, Ph.D.)
10,000 Bay Pines Blvd Bay Pines, FL 33744
- Birmingham VA Medical Center (Louis Dellitalia, M.D.)
700 S. 19th Street, Birmingham AL 35233
- Central Western Massachusetts Healthcare System (Kristin Mattocks, Ph.D., M.P.H.)
421 North Main Street, Leeds, MA 01053
- Cincinnati VA Medical Center (John Harley, M.D., Ph.D.)
3200 Vine Street, Cincinnati, OH 45220
- Clement J. Zablocki VA Medical Center (Jeffrey Whittle, M.D., M.P.H.)
5000 West National Avenue, Milwaukee, WI 53295
- VA Northeast Ohio Healthcare System (Frank Jacono, M.D.)
10701 East Boulevard, Cleveland, OH 44106
- Durham VA Medical Center (Jean Beckham, Ph.D.)
508 Fulton Street, Durham, NC 27705
- Edith Nourse Rogers Memorial Veterans Hospital (John Wells., Ph.D.)
200 Springs Road, Bedford, MA 01730
- Edward Hines, Jr. VA Medical Center (Salvador Gutierrez, M.D.)
5000 South 5th Avenue, Hines, IL 60141
- Veterans Health Care System of the Ozarks (Kathrina Alexander, M.D.)
1100 North College Avenue, Fayetteville, AR 72703
- Fargo VA Health Care System (Kimberly Hammer, Ph.D.)
2101 N. Elm, Fargo, ND 58102
- VA Health Care Upstate New York (James Norton, Ph.D.)
113 Holland Avenue, Albany, NY 12208
- New Mexico VA Health Care System (Gerardo Villareal, M.D.)
1501 San Pedro Drive, S.E. Albuquerque, NM 87108
- VA Boston Healthcare System (Scott Kinlay, M.B.B.S., Ph.D.)
150 S. Huntington Avenue, Boston, MA 02130
- VA Western New York Healthcare System (Junzhe Xu, M.D.)
3495 Bailey Avenue, Buffalo, NY 14215-1199
- Ralph H. Johnson VA Medical Center (Mark Hamner, M.D.)
109 Bee Street, Mental Health Research, Charleston, SC 29401
- Columbia VA Health Care System (Roy Mathew, M.D.)
6439 Garners Ferry Road, Columbia, SC 29209
- VA North Texas Health Care System (Sujata Bhushan, M.D.)
4500 S. Lancaster Road, Dallas, TX 75216
- Hampton VA Medical Center (Pran Iruvanti, D.O., Ph.D.)
100 Emancipation Drive, Hampton, VA 23667
- Richmond VA Medical Center (Michael Godschalk, M.D.)

1201 Broad Rock Blvd., Richmond, VA 23249
- Iowa City VA Health Care System (Zuhair Ballas, M.D.)
601 Highway 6 West, Iowa City, IA 52246-2208
- Eastern Oklahoma VA Health Care System (River Smith, Ph.D.)
1011 Honor Heights Drive, Muskogee, OK 74401
- James A. Haley Veterans' Hospital (Stephen Mastorides, M.D.)
13000 Bruce B. Downs Blvd, Tampa, FL 33612
- James H. Quillen VA Medical Center (Jonathan Moorman, M.D., Ph.D.)
Corner of Lamont & Veterans Way, Mountain Home, TN 37684
- John D. Dingell VA Medical Center (Saib Gappy, M.D.)
4646 John R Street, Detroit, MI 48201
- Louisville VA Medical Center (Jon Klein, M.D., Ph.D.)
800 Zorn Avenue, Louisville, KY 40206
- Manchester VA Medical Center (Nora Ratcliffe, M.D.)
718 Smyth Road, Manchester, NH 03104
- Miami VA Health Care System (Ana Palacio, M.D., M.P.H.)
1201 NW 16th Street, 11 GRC, Miami FL 33125
- Michael E. DeBakey VA Medical Center (Olaoluwa Okusaga, M.D.)
2002 Holcombe Blvd, Houston, TX 77030
- Minneapolis VA Health Care System (Maureen Murdoch, M.D., M.P.H.)
One Veterans Drive, Minneapolis, MN 55417
- N. FL/S. GA Veterans Health System (Peruvemba Sriram, M.D.)
1601 SW Archer Road, Gainesville, FL 32608
- Northport VA Medical Center (Shing Shing Yeh, Ph.D., M.D.)
79 Middleville Road, Northport, NY 11768
- Overton Brooks VA Medical Center (Neeraj Tandon, M.D.)
510 East Stoner Ave, Shreveport, LA 71101
- Philadelphia VA Medical Center (Darshana Jhala, M.D.)
3900 Woodland Avenue, Philadelphia, PA 19104
- Phoenix VA Health Care System (Samuel Aguayo, M.D.)
650 E. Indian School Road, Phoenix, AZ 85012
- Portland VA Medical Center (David Cohen, M.D.)
3710 SW U.S. Veterans Hospital Road, Portland, OR 97239
- Providence VA Medical Center (Satish Sharma, M.D.)
830 Chalkstone Avenue, Providence, RI 02908
- Richard Roudebush VA Medical Center (Suthat Liangpunsakul, M.D., M.P.H.)
1481 West 10th Street, Indianapolis, IN 46202
- Salem VA Medical Center (Kris Ann Oursler, M.D.)
1970 Roanoke Blvd, Salem, VA 24153
- San Francisco VA Health Care System (Mary Whooley, M.D.)
4150 Clement Street, San Francisco, CA 94121
- South Texas Veterans Health Care System (Sunil Ahuja, M.D.)
7400 Merton Minter Boulevard, San Antonio, TX 78229
- Southeast Louisiana Veterans Health Care System (Joseph Constans, Ph.D.)

2400 Canal Street, New Orleans, LA 70119
- Southern Arizona VA Health Care System (Paul Meyer, M.D., Ph.D.)
3601 S 6th Avenue, Tucson, AZ 85723
- Sioux Falls VA Health Care System (Jennifer Greco, M.D.)
2501 W 22nd Street, Sioux Falls, SD 57105
- St. Louis VA Health Care System (Michael Rauchman, M.D.)
915 North Grand Blvd, St. Louis, MO 63106
- Syracuse VA Medical Center (Richard Servatius, Ph.D.)
800 Irving Avenue, Syracuse, NY 13210
- VA Eastern Kansas Health Care System (Melinda Gaddy, Ph.D.)
4101 S 4th Street Trafficway, Leavenworth, KS 66048
- VA Greater Los Angeles Health Care System (Agnes Wallbom, M.D., M.S.)
11301 Wilshire Blvd, Los Angeles, CA 90073
- VA Long Beach Healthcare System (Timothy Morgan, M.D.)
5901 East 7th Street Long Beach, CA 90822
- VA Maine Healthcare System (Todd Stapley, D.O.)
1 VA Center, Augusta, ME 04330
- VA New York Harbor Healthcare System (Peter Liang, M.D., M.P.H.)
423 East 23rd Street, New York, NY 10010
- VA Pacific Islands Health Care System (Daryl Fujii, Ph.D.)
459 Patterson Rd, Honolulu, HI 96819
- VA Palo Alto Health Care System (Philip Tsao, Ph.D.)
3801 Miranda Avenue, Palo Alto, CA 94304-1290
- VA Pittsburgh Health Care System (Patrick Strollo, Jr., M.D.)
University Drive, Pittsburgh, PA 15240
- VA Puget Sound Health Care System (Edward Boyko, M.D.)
1660 S. Columbian Way, Seattle, WA 98108-1597
- VA Salt Lake City Health Care System (Jessica Walsh, M.D.)
500 Foothill Drive, Salt Lake City, UT 84148
- VA San Diego Healthcare System (Samir Gupta, M.D., M.S.C.S.)
3350 La Jolla Village Drive, San Diego, CA 92161
- VA Sierra Nevada Health Care System (Mostaqul Huq, Pharm.D., Ph.D.)
975 Kirman Avenue, Reno, NV 89502
- VA Southern Nevada Healthcare System (Joseph Fayad, M.D.)
6900 North Pecos Road, North Las Vegas, NV 89086
- VA Tennessee Valley Healthcare System (Adriana Hung, M.D., M.P.H.)
1310 24th Avenue, South Nashville, TN 37212
- Washington DC VA Medical Center (Jack Lichy, M.D., Ph.D.)
50 Irving St, Washington, D. C. 20422
- W.G. (Bill) Hefner VA Medical Center (Robin Hurley, M.D.)
1601 Brenner Ave, Salisbury, NC 28144
- White River Junction VA Medical Center (Brooks Robey, M.D.)
163 Veterans Drive, White River Junction, VT 05009
- William S. Middleton Memorial Veterans Hospital (Prakash Balasubramanian,

M.D.)
2500 Overlook Terrace, Madison, WI 53705