The American Journal of Human Genetics, Volume *106*

Supplemental Data

Assessing Digital Phenotyping to Enhance Genetic Studies of Human Diseases

Christopher DeBoever, Yosuke Tanigawa, Matthew Aguirre, Greg McInnes, Adam Lavertu, and Manuel A. Rivas

Supplemental Data

Supplemental Figures

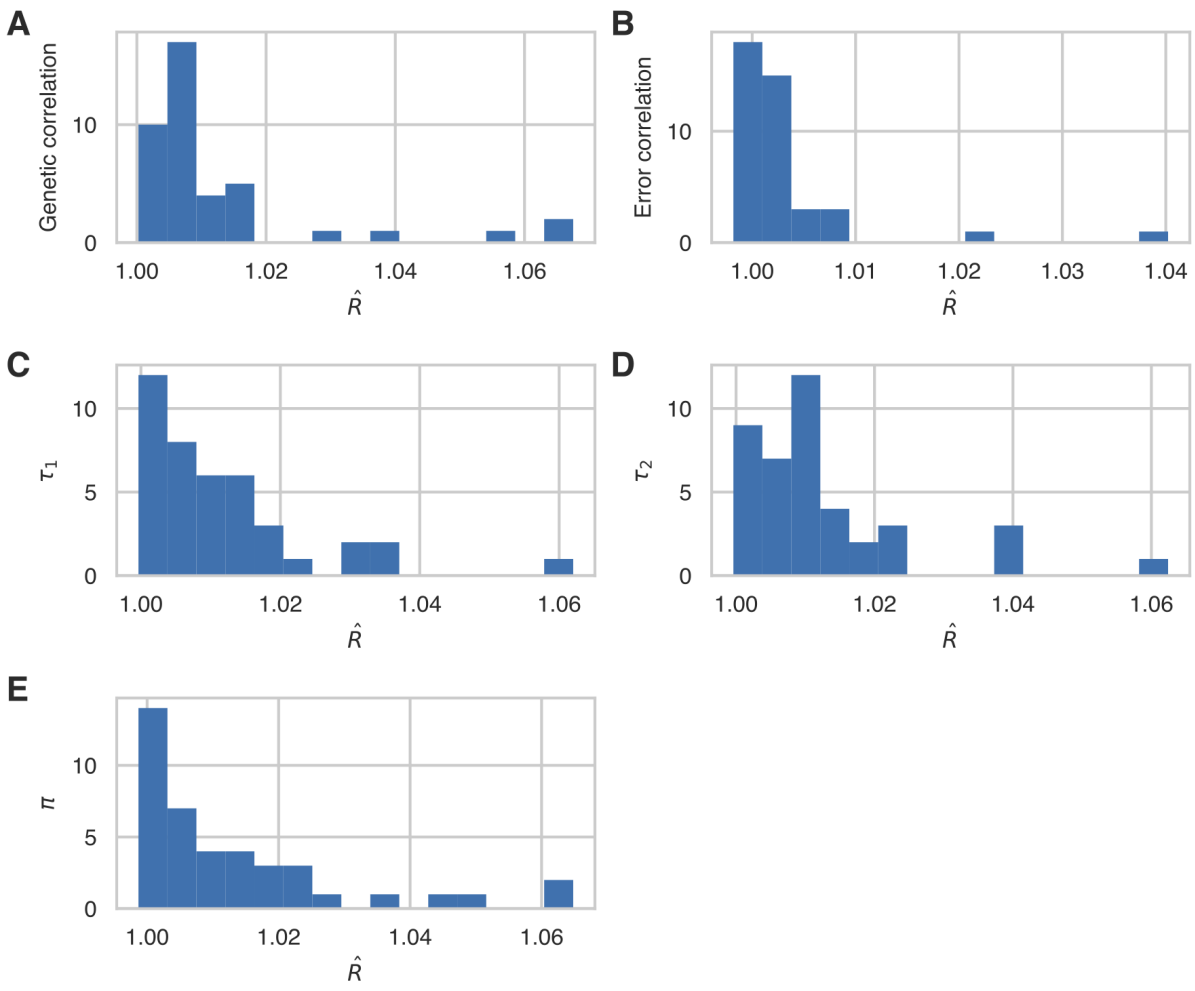


Figure S1. MVPMM $\hat{R}$ Values. Histogram of $\hat{R}$ values demonstrating MCMC convergence for parameters estimated by MVPMM using GWAS summary statistics for 41 phenotypes where cases were defined using hospital records or verbal questionnaire responses.

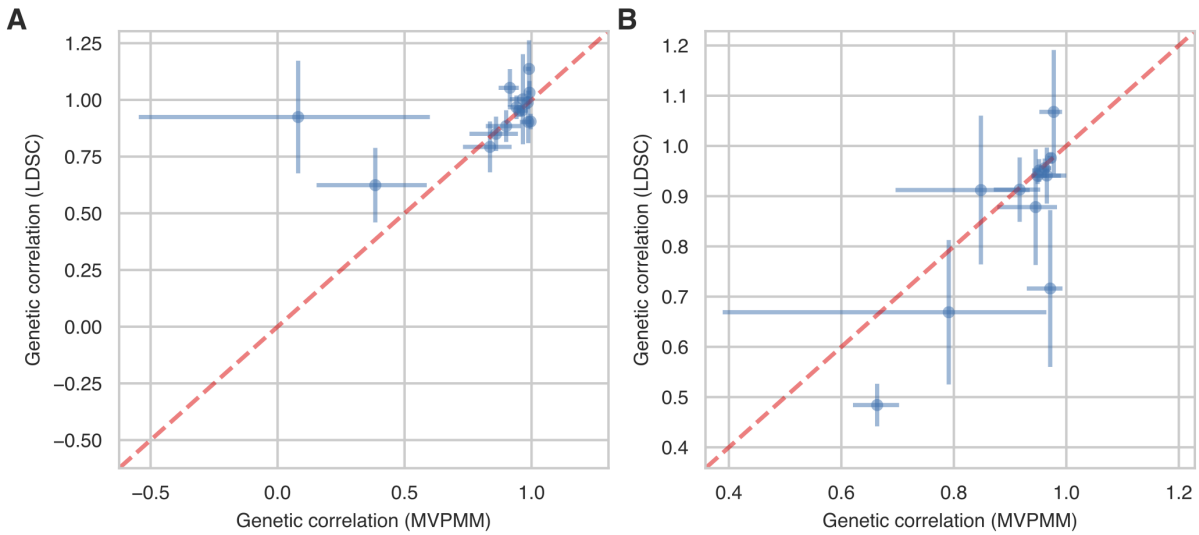


Figure S2. Comparison of Genetic Correlation Estimates from MVPMM and LD Score Regression. (A,B) Genetic correlation estimates from MVPMM (x-axis) and LD score regression (y-axis) using (A) GWAS summary statistics generated using disease definitions from hospital records or verbal questionnaire responses (minimum 1,500 cases for each) or (B) GWAS summary statistics from disease diagnosis or family history of disease (minimum 1,500 cases for each). X-axis error bars are 95% highest posterior densities and y-axis error bars are standard errors.

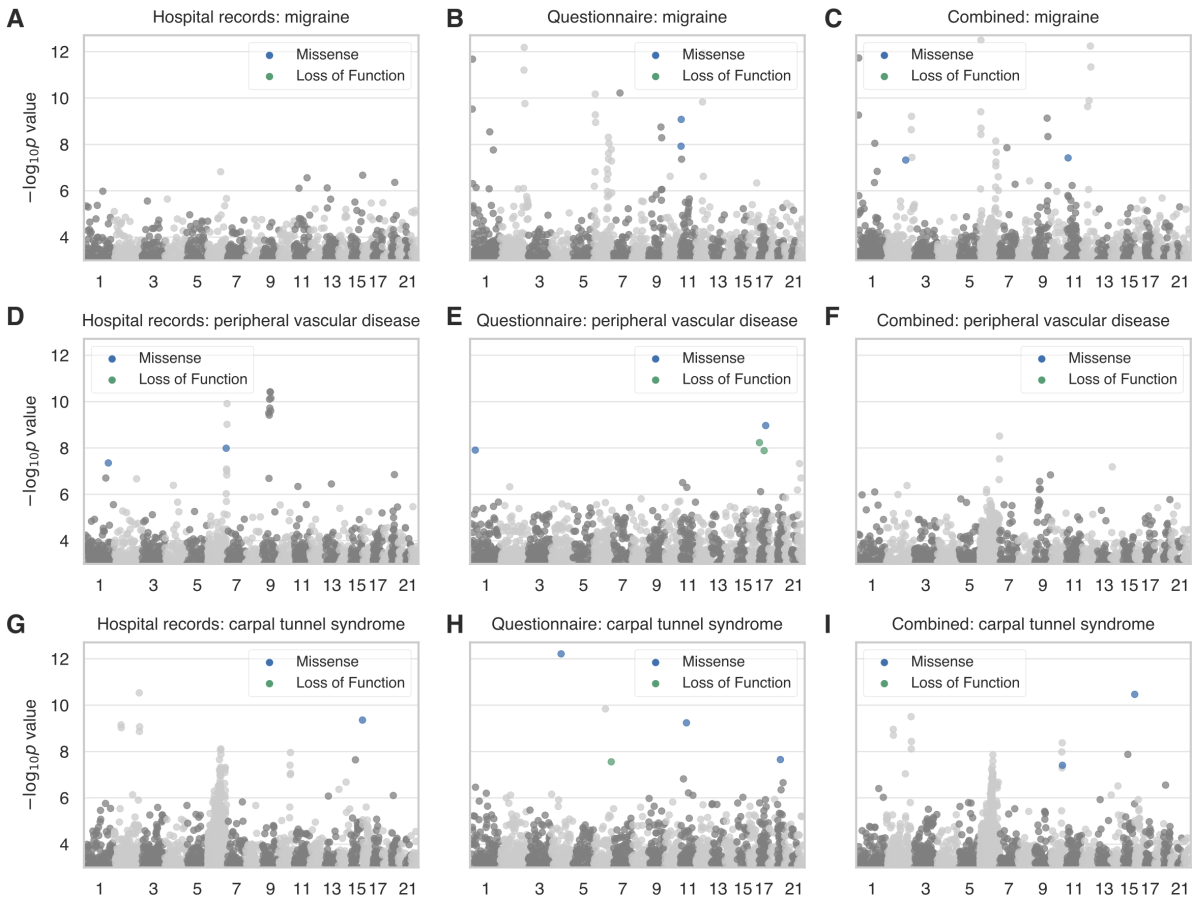


Figure S3. Hospital Record, Verbal Questionnaire, and Combined GWAS Manhattan Plots for Three Phenotypes. (A-C) Manhattan plots for migraine where cases were ascertained from hospital records (A), questionnaire responses (B), or both methods combined (C). (D-F) Manhattan plots for peripheral vascular disease where cases were ascertained from hospital records (D), questionnaire responses (E), or both methods combined (F). (G-I) Manhattan plots for carpal tunnel syndrome where cases were ascertained from hospital records (G), questionnaire responses (H), or both methods combined (I). For all panels, loss of function and missense variants with p<5e-8 are colored blue and green, respectively. Grey dots indicate all other variants.

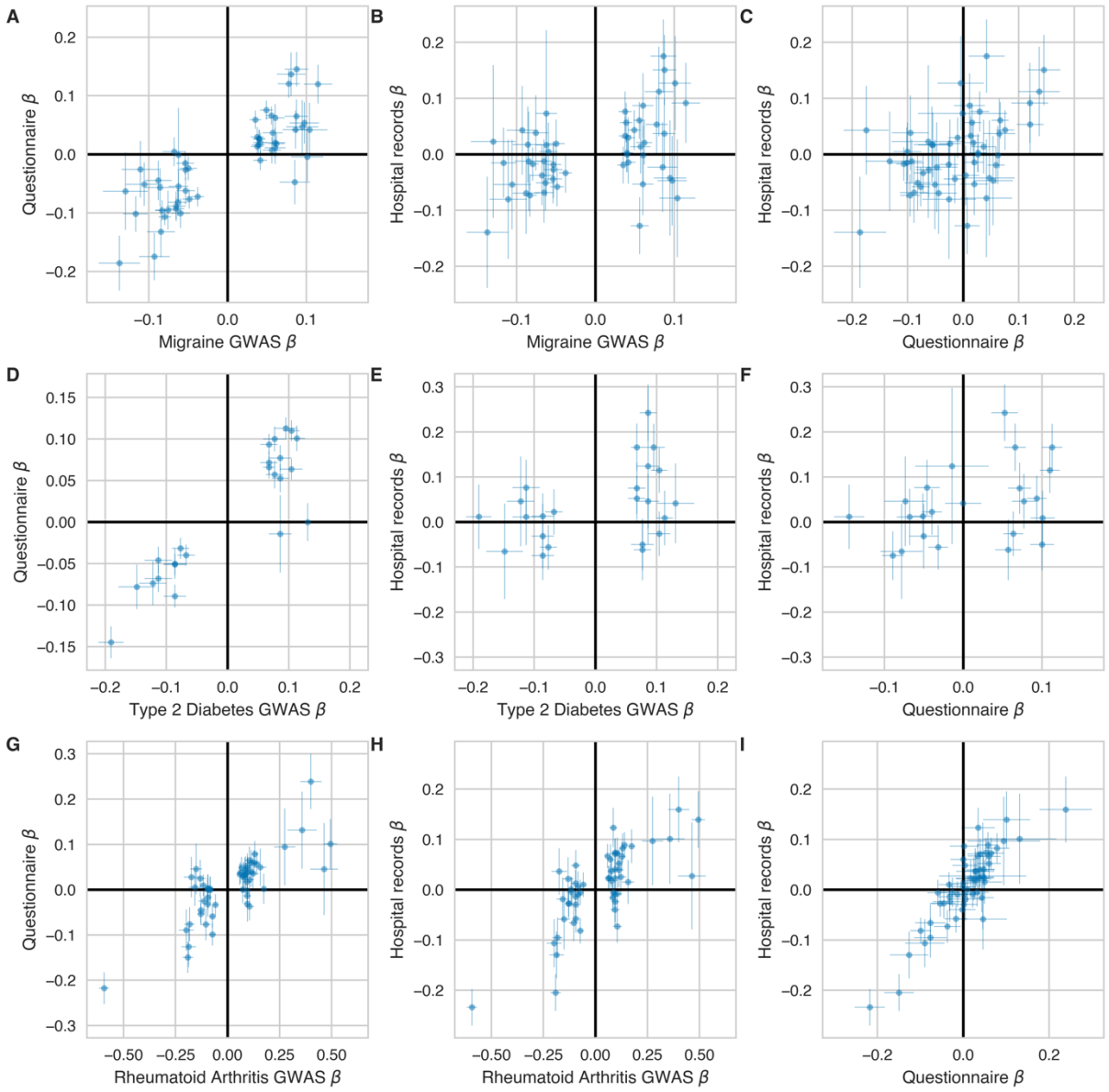
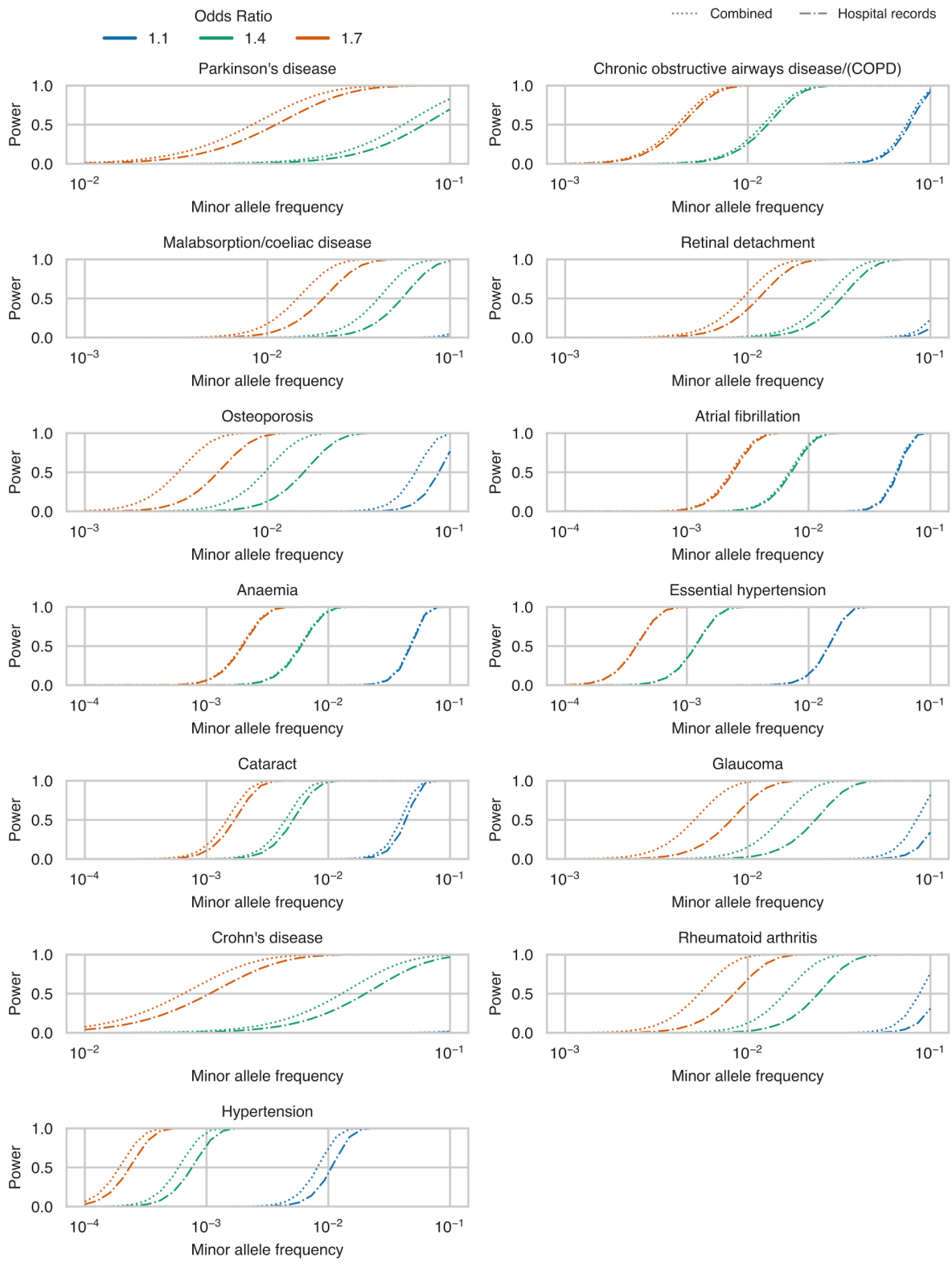


Figure S4. Comparison of Estimated Effect Sizes from UK Biobank Verbal Questionnaire and Hospital Record GWAS to Published GWAS Effect Sizes. Comparison of estimated effect sizes for migraine (A-C), diabetes (D-F), and rheumatoid arthritis (G-I) associations from GWAS using cases defined by UK Biobank questionnaire responses, GWAS using cases defined by hospital records, or published GWAS results. Effect sizes are plotted for variants that were used in MVPMM genetic correlation estimates for each phenotype and had p<1e-5 in the summary statistics for the published GWAS study. Chromosome 6 variants were not plotted for rheumatoid arthritis to remove associations driven by the major histocompatibility complex. Error bars are standard errors for estimated effect sizes.


Odds Ratio
1.1
1.4
1.7
Combined
Hospital records
Parkinson's disease
Chronic obstructive airways disease/(COPD)
Malabsorption/coeliac disease
Retinal detachment
Osteoporosis
Atrial fibrillation
Anaemia
Essential hypertension
Cataract
Glaucoma
Crohn's disease
Rheumatoid arthritis
Hypertension
Power
Minor allele frequency

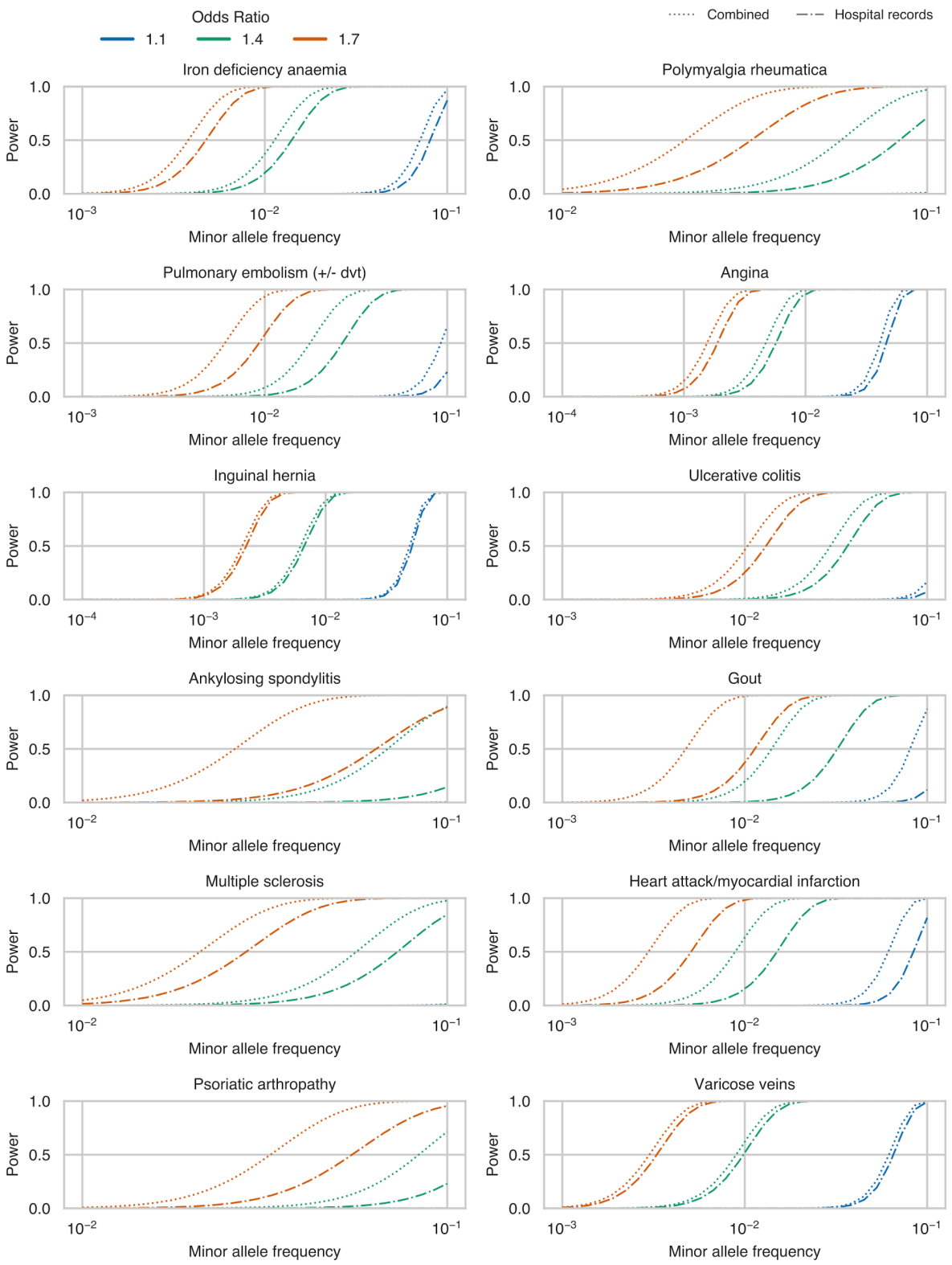
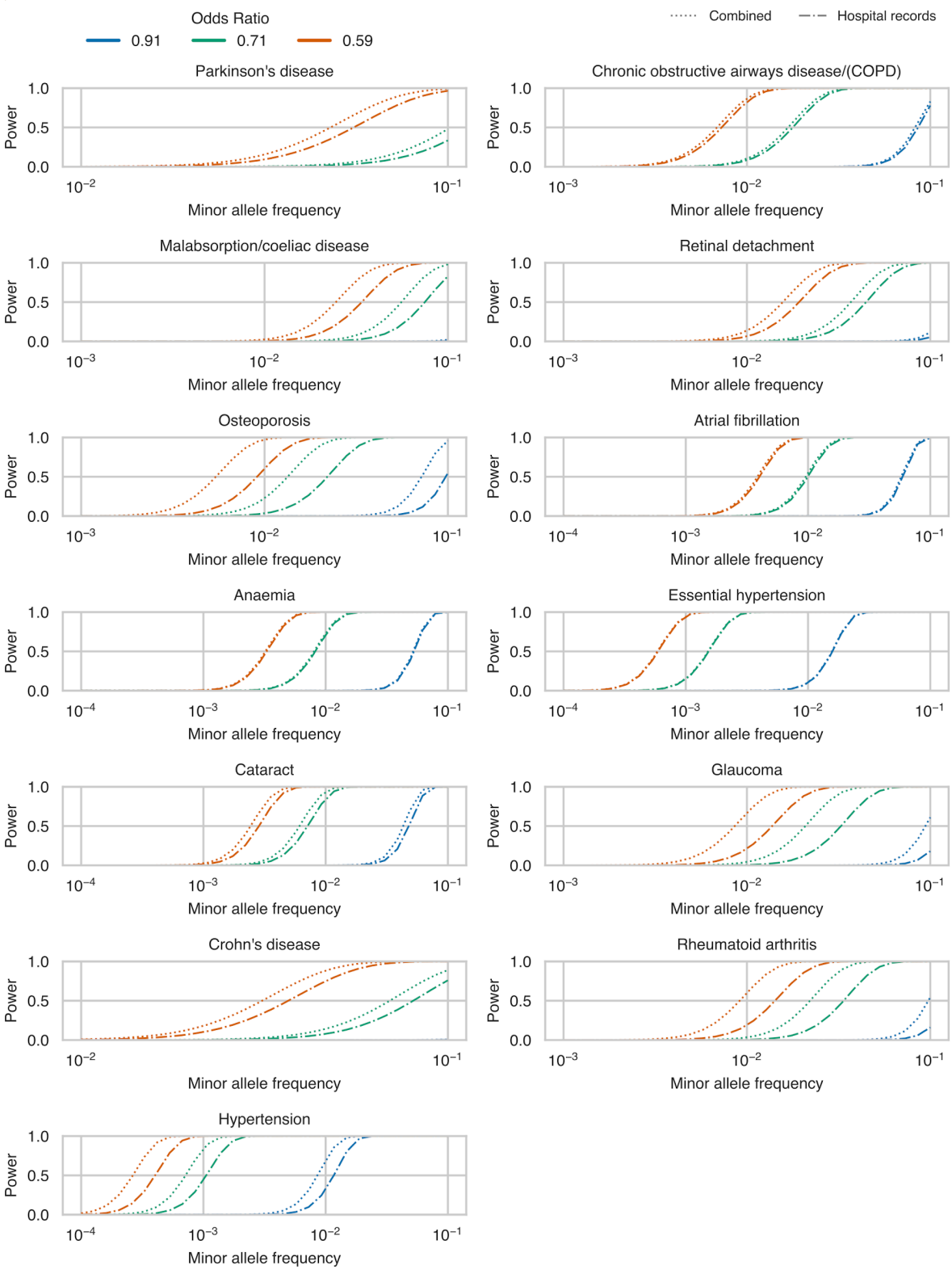


Figure S5. Statistical Power to Detect Risk Associations for Rare Variants. Power to detect rare risk associations among white British subjects in the UK Biobank using cases ascertained using only hospital records (dash-dot lines) or ascertained using hospital records and

questionnaire responses (dotted lines). All phenotypes plotted had a mean posterior genetic correlation of at least 0.8. The only parameters that differ between the dot-dash lines and dotted lines of a given color are the number of cases and controls; the dotted lines include cases that were identified from verbal questionnaire data that are otherwise classified as controls for the the dot-dash lines.


Odds Ratio
0.91
0.71
0.59
Combined
Hospital records
Parkinson's disease
Chronic obstructive airways disease/(COPD)
Malabsorption/coeliac disease
Retinal detachment
Osteoporosis
Atrial fibrillation
Anaemia
Essential hypertension
Cataract
Glaucoma
Crohn's disease
Rheumatoid arthritis
Hypertension
Power
Minor allele frequency

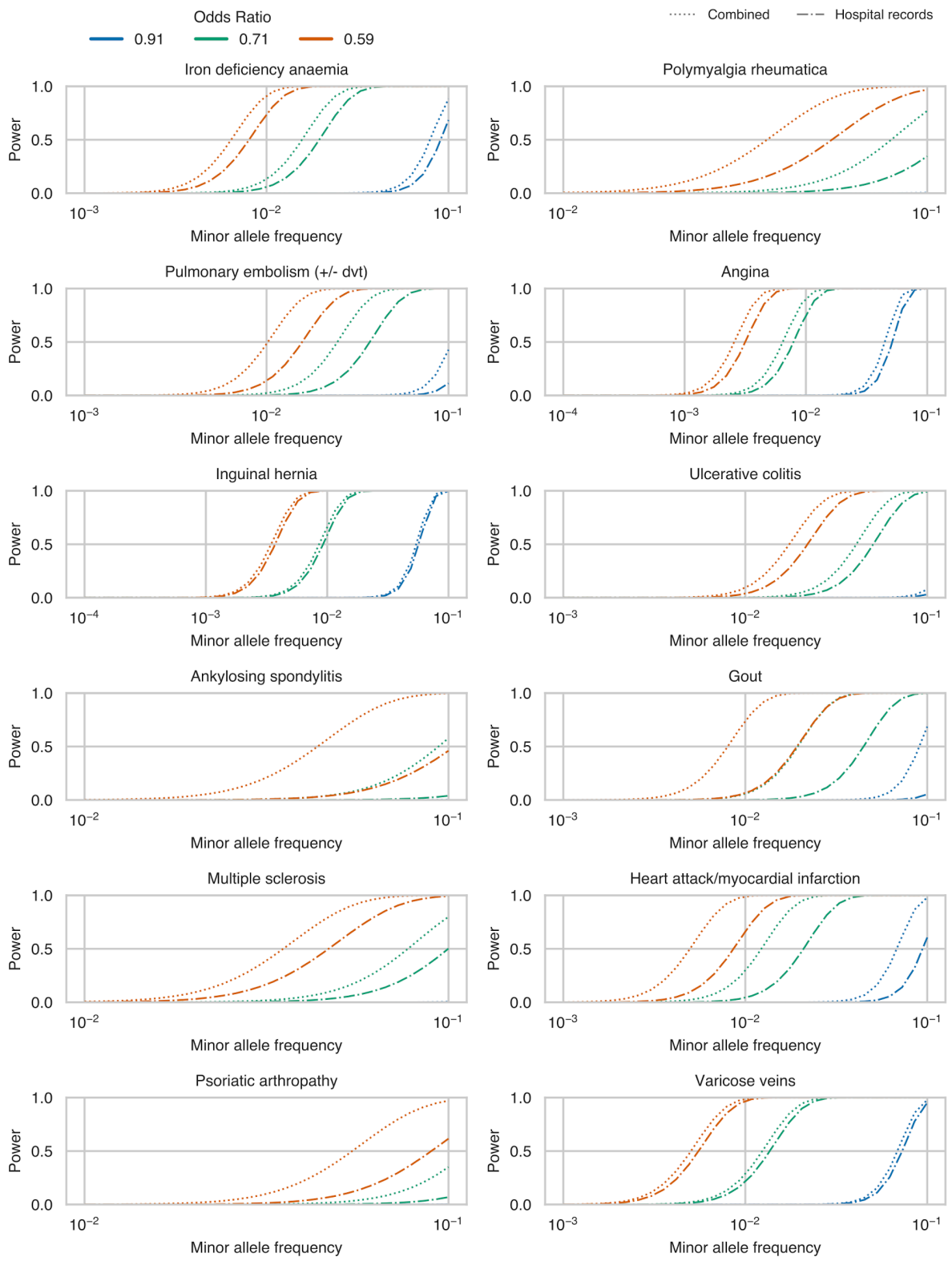


Figure S6. Statistical Power to Detect Protective Associations for Rare Variants. Power to detect rare protective associations among white British subjects in the UK Biobank using cases ascertained using only hospital records (dash-dot lines) or ascertained using hospital

records and questionnaire responses (dotted lines). All phenotypes plotted had a mean posterior genetic correlation of at least 0.8. The only parameters that differ between the dot-dash lines and dotted lines of a given color are the number of cases and controls; the dotted lines include cases that were identified from verbal questionnaire data that are otherwise classified as controls for the the dot-dash lines.

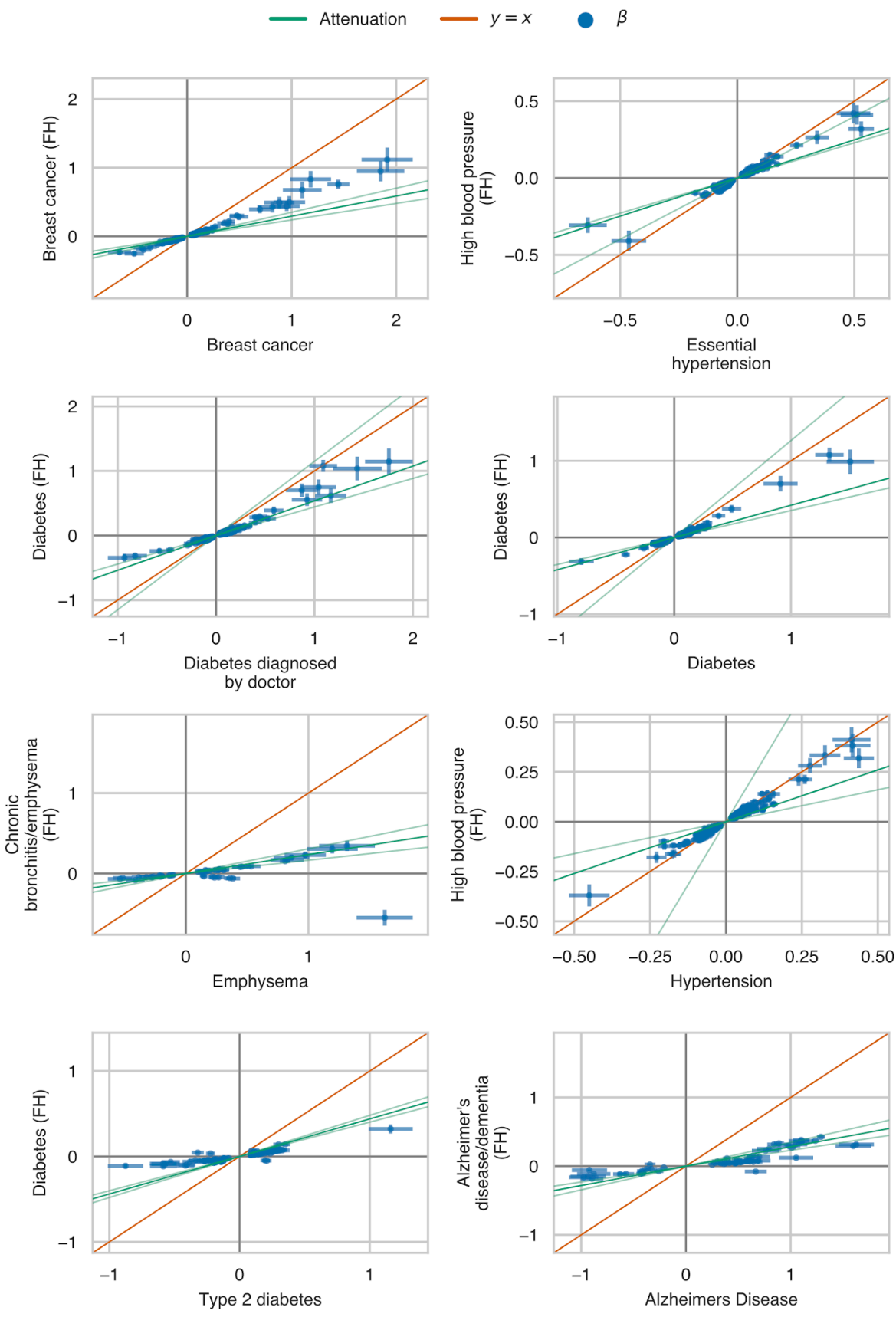

Attenuation
y = x
β
Breast cancer (FH)
Breast cancer
High blood pressure (FH)
Essential hypertension
Diabetes (FH)
Diabetes diagnosed by doctor
Diabetes
Chronic bronchitis/emphysema (FH)
Emphysema
High blood pressure (FH)
Hypertension
Diabetes (FH)
Type 2 diabetes
Alzheimer's disease/dementia (FH)
Alzheimers Disease

Figure S7. Estimated Effect Sizes and Effect Size Attenuation for Family History GWAX and Combined Hospital Record/Verbal Questionnaire GWAS. Attenuation estimates (green line, 95% highest posterior density indicated by light green lines) and estimated effect sizes (error bars are standard errors) for GWAS summary statistics from eight traits where cases were defined by either combined hospital record/verbal questionnaire data (x-axis) or family history of disease (y-axis).

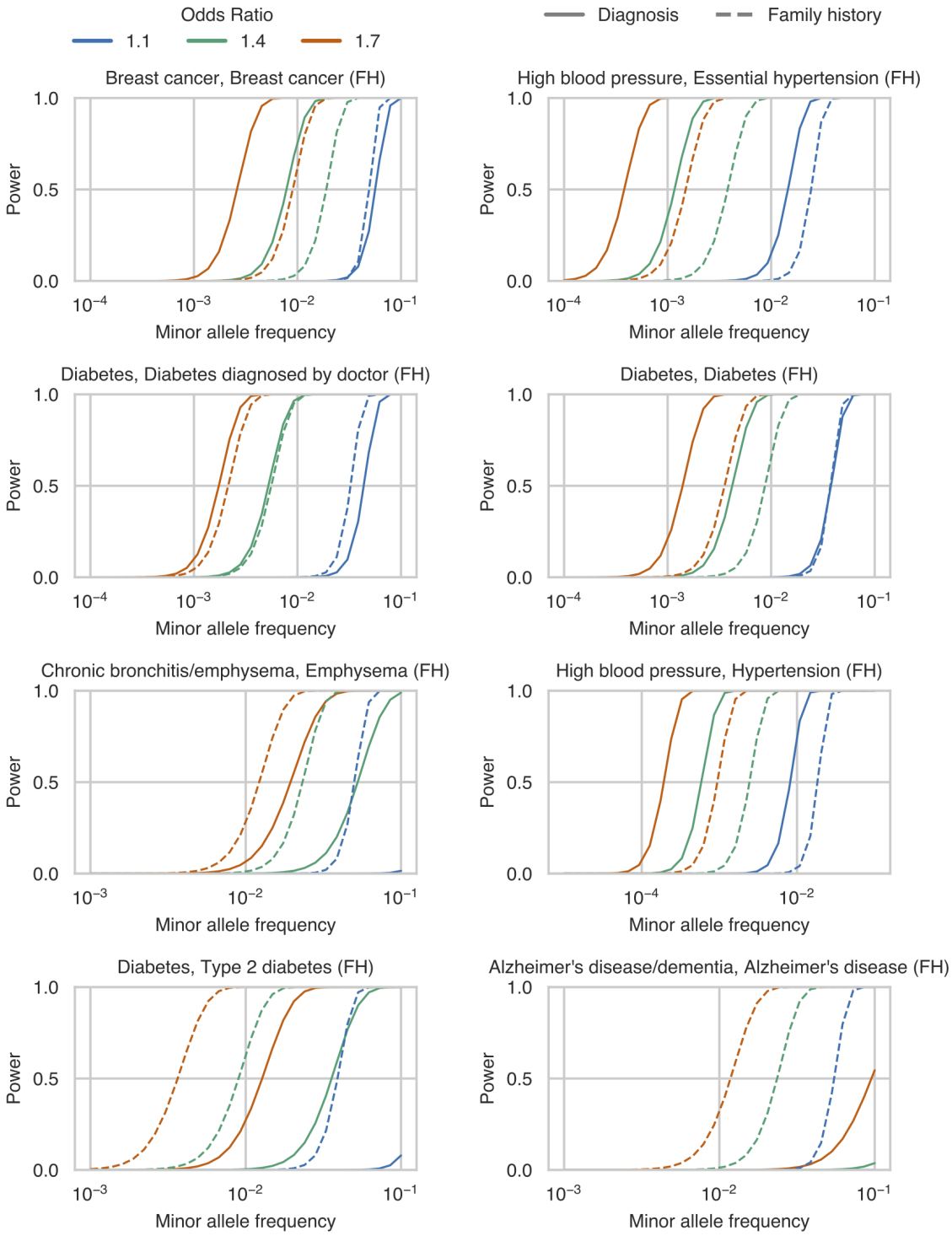


Figure S8. Statistical Power to Detect Risk Associations for Rare Variants using Family History of Disease. Power to detect rare risk associations among white British subjects in the UK Biobank using cases ascertained using hospital records and questionnaire responses (solid line) or family history of disease (dashed). The only parameters that differ between the solid lines and dashed lines of a given color are the number of cases and controls.

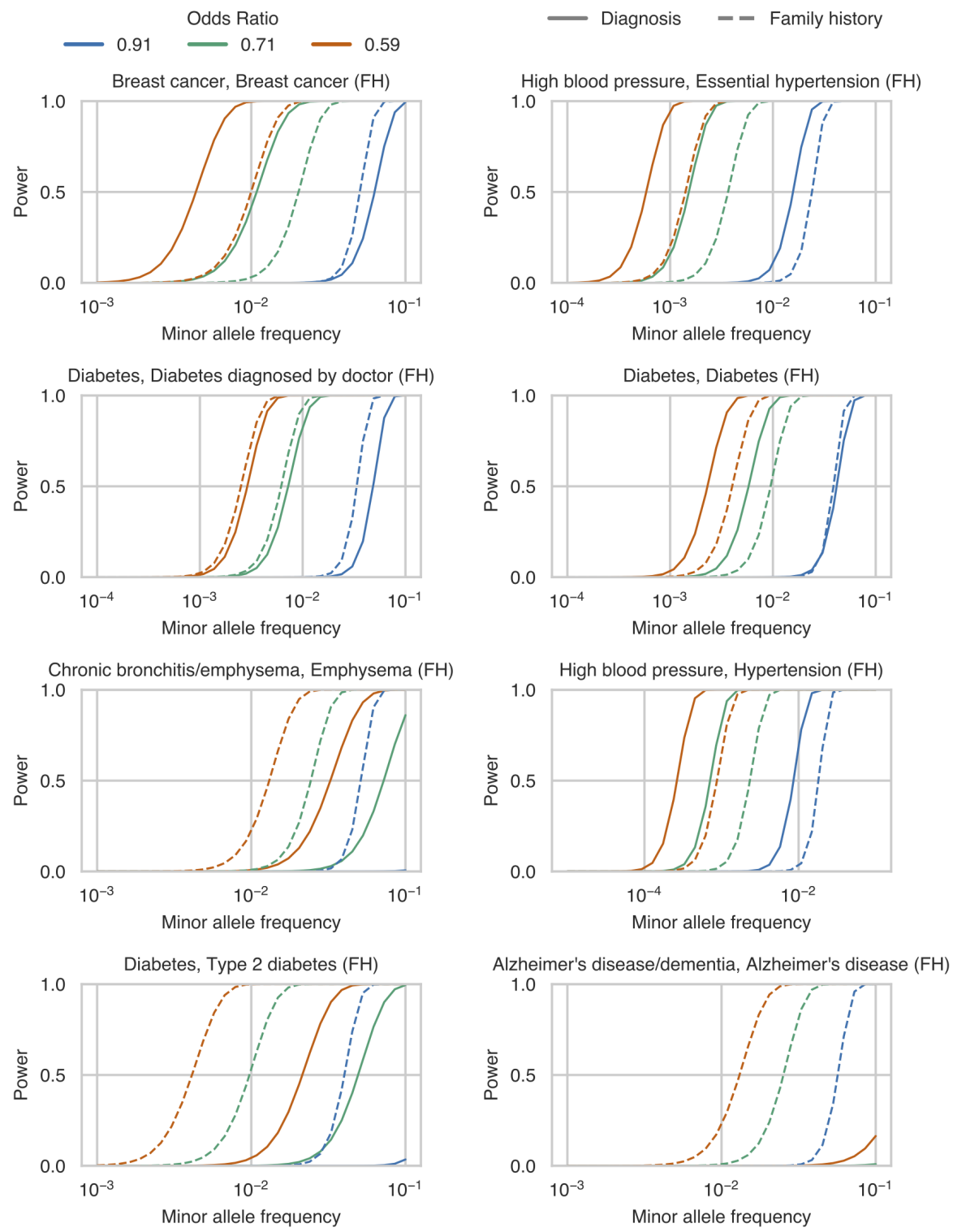


Figure S9. Statistical Power to Detect Protective Associations for Rare Variants using Family History of Disease. Power to detect rare protective associations among white British subjects in the UK Biobank using cases ascertained using hospital records and questionnaire

responses (solid line) or family history of disease (dashed). The only parameters that differ between the solid lines and dashed lines of a given color are the number of cases and controls.

Supplemental Tables

Table S1. MVPMM genetic parameter estimates using different priors for for 12 phenotypes where cases were defined using either family history of disease or diagnosis from hospital records and verbal questionnaire responses. The parameter estimates are point estimates obtained by maximizing the joint posterior using Stan's "optimizing" function.

Table S2. Number of cases ascertained by hospital records, verbal questionnaire responses, and family history of disease.

Table S3. MVPMM genetic parameter estimates for comparisons of GWAS using hospital records versus questionnaire data, combined hospital records and questionnaire data versus either hospital records or questionnaire data, and family history GWAX versus combined hospital records and questionnaire data.