medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Title

The relationship between genotype- and phenotype-based estimates of genetic liability to human psychiatric disorders, in practice and in theory.

Authors (Working)

Morten Dybdahl Krebs[1+], Vivek Appadurai[1], Kajsa-Lotta Georgii Hellberg[1], Henrik Ohlsson[2], Jette Steinbach[3], Emil Petersen[3], iPSYCH Study Consortium[#], Thomas Werge[1,4], Jan Sundquist[2], Kristina Sundquist[2], Na Cai[5], Noah Zaitlen[6,7], Andy Dahl[8], Bjarni Vilhjalmsson[3,9], Jonathan Flint[10], Silviu-Alin Bacanu[11,12], Andrew J. Schork*[1,4+], Kenneth S. Kendler*[11,12+]

Affiliations

1. Institute of Biological Psychiatry, Mental Health Center - Sct Hans, Copenhagen University Hospital – Mental Health Services CPH, Copenhagen, Denmark
2. Center for Primary Health Care Research, Lund University, Malmö, Sweden
3. NCRR - National Centre for Register-Based Research, Business and Social Sciences, Aarhus University, Aarhus V, Denmark
4. Section for Geogenetics, GLOBE Institute, Faculty of Health and Medical Science , Copenhagen University
5. Helmholtz Pioneer Campus, Helmholtz Zentrum München, Neuherberg, Germany
6. Department of Neurology, University of California, Los Angeles, California 90024, USA
7. Department of Computational Medicine, University of California, Los Angeles, California, USA
8. Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA
9. Department of Biomedicine, Aarhus University, Aarhus, Denmark
10. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA
11. Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA
12. Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA

[#]iPSYCH Study Consortium authors can be found in the supplementary tables.

*These authors jointly supervised this work

+Please address correspondence to:

Morten Krebs, morten.dybdahl.krebs@regionh.dk

Andrew J. Schork, Andrew.Joseph.Schork@regionh.dk

Kenneth S. Kendler, kenneth.kendler@vcuhealth.org

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Genetics as a science has roots in studying phenotypes of relatives, but molecular approaches facilitate direct measurements of genomic variation within individuals. Agricultural and human biomedical research are both emphasizing genotype-based instruments, like polygenic scores, as the future of breeding programs or precision medicine and genetic epidemiology. However, unlike in agriculture, there is an emerging consensus that family variables act near independent of genotypes in models of human disease. To advance our understanding of this phenomenon, we use 2,066,057 family records of 99,645 genotyped probands from the iPSYCH2015 case-cohort study to show that state-of-the-field genotype- and phenotype-based genetic instruments explain essentially independent components of liability to psychiatric disorders. We support these empirical results with novel theoretical analysis and simulations to describe, in a human biomedical context, parameters affecting current and future performance of the two approaches, their expected interrelationships, relative sample size efficiencies, and the striking consistency of observed results with expectations under simple additive, polygenic liability models of disease. We conclude, at least for psychiatric disorders, that phenotype- and genotype-based genetic instruments are likely to be very noisy measures of the same underlying additive genetic liability, should be seen, in the near future, as complementary, and integrated to a greater extent going forward.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

In the first decades of genetics as a science, the only approach for estimating genetic contributions to traits was by assessing phenotypes of relatives[1,2], but now molecular approaches provide direct measures of individual genomes cheaply, efficiently, and at scale. In agricultural applications, e.g., estimating breeding values of dairy cattle[3,4], "genotype-based genetic instruments" derived from molecular data are proving so effective that "phenotype-based genetic instruments" obtained from pedigree records are becoming redundant[3,5,6]. At the same time, human biomedical research is emphasizing genotype-based instruments, polygenic scores (PGS)[7], in precision medicine[8–10] and genetic epidemiology[11]. Psychiatric research, in particular, with few well-validated biomarkers, may have special interest in genetic instruments, but these and other human biomedical applications take place in a very different context. Human biomedical studies tend to focus on binary disease outcomes, that mya be rarer, with more complex genetic architectures, and in much larger, more diverse, uncontrolled (i.e. human) populations[6]. Here, any subsequent focus on PGS to the exclusion of phenotype-based instruments may be based more on convenience and enthusiasm than the principles and evidence of agricultural sciences[5,6,12,13]. There is a need to advance our understanding of if, how, and why phenotype-based and genotype-based genetic instruments complement each other in human biomedical research, and we aim to do so by combining theoretical analysis, simulations, and applications of state-of-the-field genotype- and phenotype-based genetic instruments in a unique psychiatric genetics data resource.

Somewhat counterintuitively, human biomedical studies combining PGS and indicators of positive family history (FH) find them complementary: they have low correlations and contribute near independently to classification and prediction. As examples in psychiatry, PGS and parental history for depression[14], autism[15], and schizophrenia[16]are near-independent in epidemiological models. More recent studies make similar observations for multiple others diseases[17,18]. Similar trends are found by studies using phenotype-based genetic instruments that combine genealogical records from registers into continuous liability scores. Studies using family genetic risk scores (FGRS; shared environment adjusted, kinship weighted sums of diagnoses in proband genealogies, e.g., [19–21]) have reported a number of associations replicating or that have been replicated by studies using PGS (e.g.,[22–25]). Model-based estimates of liability from family records, liability threshold on family history (LT-FH)[17] and Pearson-Aitken family genetic risk scores (PA-FGRS)[22], have low correlations with PGS and contribute near-independently to classification. There is an emerging consensus that family variables act independently of genotypes in models of human disease. Many intuit this as evidence that family variables reflect effects of shared environment but it is

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

currently unclear if these reports can be consistent with evidence of large additive genetic components in complex trait etiology[26,27].

Studies that thoroughly explore expectations around relationships between genotype- and phenotype-based genetic instruments come more from agricultural sciences, while those focusing on human biomedicine are sparser or limited. Expectations and strategies for estimating breeding values of dairy cattle, for example, from pedigree records and genotypes are mature, well-studied, and well-described[5,12,13]. The focus on quantitative traits, controlled breeding, small effective populations, long haplotypes, reduced genetic diversity, and extensive pedigree records of many, often genotyped, relatives and multiple offspring may limit how well trends transfer to human studies[6]. In humans and agriculture alike, PGS can, in theory, explain all heritable disease risk when constructed from all causative variants and errorless effect sizes. In practice, limited sets of marker genotypes and estimated allelic effects impose constraints. The error in empirical PGS reduces with increasing heritability, coverage of marker genotypes, and size of the GWAS used to estimate allelic effects[7,28–30], the rate of gain, however, may be slow and require large investments of resources. On the other hand, epidemiological models often incorporate simple surveys of family history or sophisticated summaries of genealogical records from registeries[31,32] into risk models, and others have derived breeding value-like genetic instruments for human studies[33,34]. Few human studies, however, describe the two approaches together or focus on relative strengths and weaknesses[35]. Less have addressed if, how, or why genotype- and phenotype- based estimates are complementary or expectations for their interrelationships under etiological models.

In this work we describe the relationships among phenotype-based genetic instruments (FH and PA-FGRS), genotype-based genetic instruments (PGS), and liability for five major psychiatric disorders. We use data from two independent cohorts of the iPSYCH2015 case-cohort study that include diagnostic information for 99,645 genotyped probands (62,357 psychiatric cases, 37,288 controls) and 2,066,657 relatives. To do so, we show specific approaches are needed to avoid an unreported, potential bias when estimating liability-scale variance explained by FH and PA-FGRS. We support our results with derivations of novel theoretical expectations for accuracies and interrelationships of FH, PA-FGRS, and PGS and simulation experiments. We use both to provide interpretive context for current and future research combining such instruments, with a specific focus on settings representative of human biomedical research. We consider parameters that affect utility, expectations for redundancy, and measures to compare relative efficiency with respect to, e.g., sample size. We then assess the fit of our results to expectations under a simple, additive polygenic liability model for genetic inheritance. We ultimately propose that the lack of redundancy between PA-FGRS (or FH) and PGS in human biomedical research is consistent with the hypothesis that both are simply very noisy measures of a highly similar underlying genetic liability. Different

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

from trends in agriculture applications, we suggest that such instruments should be emphasized as complementary, at least over the near-term, and should be better-integrated going forward.

Results

Common equations for variance explained on the liability scale produce biased estimates for phenotype-based genetic instruments.

The variance explained by a genetic instrument in a regression model is often transformed to the liability scale to remove its dependence on case-control sampling and enable direct comparisons to estimates of heritability[36,37]. The most frequently applied transformations (i.e., Lee et al[37]) were derived assuming one-stage case-control sampling (i.e., *complete* ascertainment) and well-behaved (i.e., Gaussian, exogenous) genetic instruments like PGS, but have not been evaluated assuming the more typical two-stage case-control sampling applied in molecular genetics studies (i.e., sampling *unrelated* cases; akin to observing relatives conditional on their relationship to an affected proband as traditionally done in twin register studies[38]) and less well-behaved (i.e., non-Gaussian, endogenous) genetic instruments (e.g., FH, PA-FGRS). In simulations (Supplementary Information S1) we observe common transformations from variance explained by a linear model on the observed scale to the liability scale are biased for FH and PA-FGRS, but not for PGS (Figures 1A-C). This resulted in overestimating the variance explained by more than 400% in some scenarios and produced bias in all scenarios where the trait prevalence was less than 50% (Figures 1A-C). Directly estimating variance explained on the liability scale using a weighted probit regression as described by Lee et al[37] (Online Methods) reduced this bias under population sampling (Figure 1D) and complete case-control sampling (Figure 1E), but introduced a downward bias of as much as 50% when unrelated case-control sampling was applied (Figure 1F). A weighted probit regression with two-stage sampling weights (Online Methods, Supplementary Information S1) to account for this second selection, namely, removing all but one member of a relative cluster, produces the most reliable estimates of variance explained on the liability scale, regardless of scenario (Figure 1G-I). Caution should be used when interpreting scale-transformed variance components if sampling is complex (i.e., multi-stage) and instruments are not well-behaved (e.g., FH, PA-FGRS, or other liability scores); in these scenarios a direct approach with appropriate sampling weights should be preferred.

Genotype- and phenotype-based genetic instruments explain nearly independent components of liability to psychiatric disorders

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We estimated the variance in liability to five psychiatric disorders explained by combinations of polygenic scores (PGS), indicators of positive family-history (FH), and Pearson-Aitken Family Genetic Risk Scores (PA-FGRS) using weighted probit regression accounting for two-stage sampling in data from the iPSYCH2015 case-cohort[39] (Online Methods; Figure 2; Supplementary Table S1, S2). Pseudo-R^2 from logistic models are shown in Supplementary Figure S2. PGS, FH, and PA-FGRS variables each explain significant liability-scale variance for each disorder, a result that replicates across two independent samples (Figure 2A,B, orange, yellow, green bars). For all disorders, PA-FGRS explain more liability variance than FH (mean increase 30%, range 11% to 46%; Supplementary Table S2). The variance explained by PGS and PA-FGRS (or FH) is nearly the sum of variance explained by PGS and PA-FGRS (or FH) independently (Figure 2A,B, blue and purple; Supplementary Table S2). Including FH did not typically increase the variance explained of models including PA-FGRS and PGS (4th vs. 6th bars in each set, Figure 2A,B, Supplementary Table S2; exception: MDD in iPSYCH 2012). However, including PA-FGRS in models that with PGS and FH did generally result in small but significant increases in variance explained (5th vs. 6th bars in each set, Figure 2A,B, Supplementary Table S2; exceptions: BPD, SCZ). Similarly, estimated of odds ratios (OR) for PA-FGRS (or PGS) were not strongly reduced when adjusting for PGS (or PA-FGRS) (Figure 2C,D, Supplementary Table S3). Estimates of ORs for FH adjusting for PA-FGRS were reduced to a larger extent than when PA-FGRS adjusting for FH (Supplementary Figure S3-4, Supplementary Table S3). Taken together, this suggests PA-FGRS is a better instrument for capturing familial genetics than FH. Finally, we observe low correlations between PGS and PA-FGRS in random population samples (Pearson correlation coefficient (r), 0.03 to 0.09; Figure 2E,F) that were consistently, yet modestly, larger than between PGS and FH (r, 0.02 to 0.07), and much lower than between PA-FGRS and FH (r, 0.72 to 0.82, Supplementary Figures S5 and S6). Genotype-based and phenotype-based genetic instruments complement each other in predictive models of psychiatric disorders, which is both useful for constructing more powerful risk assessments, but also may be counterintuitive as both purport to estimate a similar construct of genetic liability.

The current, future, and asymptotic performance of PA-FGRS depends on pedigree structure and trait architecture

We derive novel expectations for the accuracy of PA-FGRS (^_, the correlation between the instrument, ^_, and the true genetic value,), as well as for the performance of PA-FGRS (^{2}^_, the variance in liability explained or the squared correlation between the instrument and the total true liability value,) under a simple polygenic liability threshold model (Online Methods; Supplementary Information S2-6). We use these equations to describe how PA-FGRS relates to true liability, considering four hypothetical traits

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

(Figure 3A-F; Supplementary Figures S7-S15). First, PA-FGRS that do not incorporate phenotypes of the proband or offspring of the proband do not capture Mendelian segregation variance (i.e., between sibling genetic variance) as the input pedigree is identical for all unobserved full siblings without children[6]. In this scenario, the maximum theoretical accuracy and performance of PA-FGRS (Figures 3A-B, dashed lines) is bounded by the between family genetic variance (i.e., half of the total genetic variance), and equal to that of the true mid-parental genetic value (g_{MPV}; $_{,} = \sqrt{0.5}$, $^{2}_{,} = 0.5^{2}$)[6]. Full accuracy ($_{,\hat{}\_} = 1$) and performance ($^{2}_{,\hat{}\_} = {}^{2}$) can be achieved, in theory, when a proband has multiple offspring from different mates (i.e., *half-sibling offspring* or offspring that are half-siblings to each other; Figures 3C-D, dashed lines). Using simple, single-relative-class pedigrees we see that incorporating additional phenotypes of full siblings (Figure 3A,B) is less useful than phenotypes of such half-sibling offspring (Figure 3C,D). For plausible human pedigrees (<10) the accuracy and performance of PA-FGRS are well below theoretical maxima (Figure 3A-D, Supplementary Figures S7-S12). Second, the expected accuracy and performance of PA-FGRS that use complex, extended pedigrees asymptote practically with four generations of records (i.e., with great-grandparents and their descendants; Figure 3E-H). To compare the information in complex, arbitrarily sets of relatives, we define the number of full sibling equivalents (N_{sibe}; Online Methods; Figure 3I,J). For a simulated trait with a heritability of 0.7 and prevalence of 0.01, an N_{sibe} of 5 is achieved by ~1 monozygotic twin, ~4 half-sibling offspring, or approximately one entire great-grandparent founded pedigree with two children per generation (for other simulated trait architectures see, Supplementary Figures S16). While pedigree records are always more informative for common traits (e.g., Figure 3A-H, purple vs. orange lines), the trends in the N_{sibe} across traits (Supplementary Figures S16) show distant relatives (e.g., half-siblings or cousins, green or blue lines, respectively, Figure 3IJ) are *relatively* more predictive for rarer traits (i.e., correspond to higher N_{sibe}), while closer relatives (i.e., half-sibling offspring, yellow solid lines, Figure 3IJ) are relatively more important for common traits. Simulations with shared (or "familial") environmental contributions from siblings and parents show that PA-FGRS will incorporate confounded genetic and familial variance into predictions, slightly overperforming the level expected by genetics alone (Supplementary Figures S17-S20). PA-FGRS will perform best for highly heritable, common traits and with many close relatives, but even in optimistic scenarios for human pedigrees (N_{sibe}=5) it can only explain a modest proportion of heritability.

PA-FGRS is expected to outperform PGS when GWAS samples are small

Next, we compared the expected performance of PGS trained using GWAS of various sizes to PA-FGRS with N_{sibe} = 1, 3, or 5 (Online Methods; Supplementary Information S7-8; Figure 4). PGS can, in theory,

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

achieve full accuracy (i.e., $_{,}$ =1) and performance (i.e., $^{2}_{,} = {}^{2}$) if incorporating all causative variants and errorless effect sizes, but training GWAS impose limits in practice. We estimate (Online Methods) that a PGS tagging a human common variant genome (i.e., capturing m=60,000 independent genetic factors[6,29]) accounting for half of the narrow sense heritability (i.e., $^{2}_{,} = 0.5^{2}$) is expected to perform as well as PA-FGRS with N_{sibe}=1 if a population training GWAS has roughly N=65,000 (range of sample size estimates across simulated traits: 61,546 to 71,356; Figure 4A-D, Supplementary Table S4). The same PGS is expected to perform as well as PA-FGRS with N_{sibe}=3 and N_{sibe}=5, with population training GWAS of N=190,000 (range: 176,047 to 214,070) and N=300,000 (range: 280,368 to 356,784), respectively (Figure 4A-D, Supplementary Table S4). Trait architecture (^{2} and prevalence) have smaller impacts on these relationships than features of the PGS instrument (e.g., m, the number of independent genetic factors covered, $h^{2}_{l,SNP}/h^{2}$, the proportion of heritability captured by the markers, and p, an effective polygenicity parameter; Supplementary Figures S21-S23; Supplementary Table S4). For rarer traits, case-control sampling (c.c.; here, a maximally powered study of 50% cases and 50% controls) can increase performance of PGS substantially relative to PA-FGRS (e.g., Figure 4A). For more common traits, the difference between the PGS and PA-FGRS is less, even under case-control sampling (e.g., Figure 4D). Simulations including shared environmental contributions from siblings and parents suggest PGS do not incorporate simple models of confounded genetic and familial variance (Supplementary Figures S17-S20). Minimal (N_{sibe}=1), realistic (N_{sibe}=3), and optimistic (N_{sibe}=5) PA-FGRS can outperform within population PGS when training GWAS have modest sample sizes, with the relative performance of PA-FGRS being larger for more common traits.

PA-FGRS is expected to complement PGS for at least the near future

Increasing performance of one instrument does not necessarily make other instruments redundant, as observed in Figure 1, even if they measure the same construct. Under a simple polygenic liability model (Online Methods), PGS and PA-FGRS can be viewed as estimators of the same latent true genetic value, *G*, and the expected correlation, as well as joint and conditional performances, are simple functions of their accuracies (Figure 5, Supplementary Figure S24, Supplementary Information S9). When the individual accuracies of PGS and PA-FGRS are modest, the correlation between the instruments is expected to be modest, even under models assuming no shared environment (Figure 5A-D). Because PA-FGRS has low accuracy in all considered scenarios, it will always show modest correlations with PGS (Supplementary Figures S25-S27). Modest correlations between PGS and PA-FGRS imply that, even as PGS training GWAS sample sizes grow, PA-FGRS should remain complementary to PGS (Figures 5E-H). This trend is stronger when trait prevalence is higher (e.g., Figure 5E vs Figure 5F) and when the PGS is limited due to, e.g., small

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

training GWAS or large contributions from non-tagged variants (i.e., when h^2_{SNP}/h^2 is low or m is high, Figure 5, Supplementary Figures S28-30). Simulating shared environmental contributions from siblings and parents demonstrates that confounded familial and genetic variance in PA-FGRS reduces its empirical correlation with PGS relative to expectations given its performance (Supplementary Figures S17-S20). Taken together, low accuracies of PGS and PA-FGRS suggest currently observed low to modest correlations between genotype- and phenotype-based genetic instruments are expected, even under an optimistically simple polygenic liability threshold model. This trend should persist even as GWAS sample sizes grow.

Expected and observed performance of genetic instruments are in alignment.

Finally, we compared the empirical performance of our instruments (Figure 1) to expectations under a simple polygenic liability model, given our input data (Online Methods; Figure 6; Supplementary Tables S5,S6). The iPSYCH proband pedigrees have roughly N_{sibe} = 1.63, 1.4, 1.61, 1.3, 1.76, for predicting ADHD, ASD, BPD, MDD, and SCZ, respectively (Online Methods; Supplementary Table S1). The modest empirical performances of PA-FGRS in our data (Figure 2AB) are broadly consistent with theoretical expectations, with the exception SCZ (Figure 6A; Supplementary Figure S31). Similarly, the empirical performances of PGS in our data (Figure 2AB) are also broadly consistent with expectations, with the exception of ADHD (Figure 6B, Online Methods, Supplementary Table S5). The modest empirical correlations between PGS and FGRS (Figure 2EF) are larger than expected under our simplified generative model (Figure 6C). Our PA-FGRS (Observed N_{sibe}, 1.3 to 1.8, Supplementary Table S1) perform empirically as well in we estimate a PGS trained on population samples of 190,000 to 1,140,000 individuals from the same target population (Figure 6D, Supplementary Table S6). Our PGS (Observed $N_{Pop.\ GWAS}$, 100,000 to 2,100,000, Supplementary Table S1) perform empirically as well as we expect PA-FGRS constructed from pedigrees with N_{sibe} from 0.3 to 4.7 (Figure 6E, Supplementary Table S6). In contrast with the expectations above, PGS requiring much larger than 65,000 sample training GWAS to achieve the performance of PA-FGRS with N_{sibe}=1 due to empirical h^2_{SNP}/h^2 trending much less than 0.5 and genetic correlations between training and target populations trending less than 1 (Supplementary Table S5). These results do not suggest large confounding by shared environment, as that would predict higher than expected performance of PA-FGRS and lower than expected correlation between PA-FGRS and PGS (Supplementary Figures S17-20). As such, low empirical correlations can be consistent with simple polygenic liability models of inheritance and PA-FGS with only a few relatives can be as useful for predicting disease in practice as PGS constructed using extremely large GWAS, in practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

For nearly a century, genetics relied almost exclusively on phenotypic records in related individuals to predict and describe the etiology of traits, but now there is a focus on direct assessment of molecular genotypes. Despite the successes of molecular genetics, our work suggests we should not discard as old fashioned, the potential for familial phenotypic records in modern genetic medicine and genetic epidemiology. Even as molecular focused GWAS grow to sample sizes in the millions, we expect that phenotype-based genetic instruments will be important, powerful, and useful tools. They can complement PGS and other genotype-based instruments in prediction, classification, and description of disease etiology. Also, we emphasize that the lack of redundancy observed between PA-FGRS (or FH) and PGS here, and in similar studies, is not strong evidence for the presence of shared environmental effects. In fact, our results fit very closely to expectations where PA-FGRS, FH, and PGS are simply highly noisy estimators of a highly similar underlying genetic liability. We also stress that, in scenarios where financial resources, trait (genetic) architecture, or obtainable population sizes (e.g., isolated or marginalized populations with fewer than a million members) are limiting, opting to collect detailed familial records could, in fact, remain a more efficient approach for building best-in-practice genetic instruments. Continuing to develop methods, models, and data resources that allow us integration of phenotype- and genotype-based genetic instruments will remain relevant moving forward.

Our study highlights that, at the present moment, we can obtain comparably accurate and performant estimates of genetic liability based on phenotypes of relatives and direct genomic variation for state-of-the-field PGS, albeit with some variability. For prototypical successes of psychiatric GWAS, bipolar disorder[40] and schizophrenia[41], which are rarer, highly heritable disorders with massive investments in powerful case-control GWAS (i.e., sample sizes equivalent to population sampling of over 1,500,000 participants), we saw PGS outperform PA-FGRS in our data by ~70% on average. In absolute terms, however, this was less than 1% of variance in liability. This 1% of variance in liability (in our iPSYCH data) could be accounted for on sample size scales, by 1,000,000 individuals in a population GWAS, or, alternatively, one full sibling worth of phenotype records in a PA-FGRS. For ADHD, on the other hand, a moderately large but homogeneously ascertained GWAS was performed in the same population as the evaluation (i.e., restricted genetic ancestry population served by the Danish national healthcare system and ascertained via a single national register based research program) the implied sample size differences were less dramatic. We similarly observed differences of 1% variance explained for ADHD instruments, but here this difference in sample size scales corresponds to ~100,000 individuals in GWAS or ~0.5 full sibling equivalents in PA-FGRS. As we demonstrate, many features of the trait architecture (the heritability,

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

population prevalence) and trait genetic architecture (proportion of the genome contributing risk) interact with aspects of the instruments (amount of pedigree information, size of training GWAS, number of included markers, proportion of trait heritability accounted for by the markers) in non-linear, and complex ways to determine the performance of both instruments. Here, for SCZ and ADHD in particular, PGS performance may be especially affected by test-training cohort heterogeneity. SCZ is a disorder where the large training GWAS was least representative of our iPSYCH test cohort (test-training genetic correlation of ~0.45), whereas the GWAS for ADHD was performed within-population, eliminating this challenge (test-training genetic correlation of ~1). Cross cohort differences in, at least, genetic ancestry[42], ascertainment schema[43], and phenotype quality[44] are known to reduce genetic correlation among potential test-training pairs, and for PGS performance to improve efficiently as sample sizes grow, these challenges will need careful attention, whereas the within family nature of PA-FGRS mitigates some of this loss of efficiency.

PA-FGRS and other similarly developed phenotype-based instruments are less sensitive to some known PGS pitfalls, as others have shown[17], creating advantages with some open challenges. Incorporation of multi-generational pedigree information has unique challenges with room for methodological developments. For example, there are implicit ascertainment biases of individuals enrolled into any birth cohort - namely they were born and from two specific mates. It is known that issues of differential fertility[45,46], survivorship to and following reproduction[47], and mate choice[48,49] complicate the estimation and interpretation of parameters core to our model, such as heritability. While the Danish civil registration system[50] allows high quality follow-up on the Danish population, its initiation in 1968 means large portions of grand-parents of our probands were only followed for the end of their lives. Many great grand-parents are not observed at all, meaning we cannot identify relationships among grandparents to join their descendant lineages (i.e., more distant cousins can not be recovered). PA-FGRS is built on a mixture model to account for this early life censoring that performs well in simulations and practice, but could be further optimized. As we showed, constructing accurate genetic instruments from pedigree information is difficult when pedigrees are small, censored, and disorders are rare - we just do not observe enough affected relatives. Methods that successfully augment the imputation of liability from other traits and measures[51,52] could be valuable. In addition to methodological advances, obtaining detailed records for pedigrees for at least three generations, especially in isolated, marginalized, or small populations, could be powerful.

The near independence of our genotype- and phenotype-based genetic instruments is consistent with observations made in prior studies[22]. We show that at current sample sizes, even when both instruments estimate the same, simple additive genetic liability, low correlations and nearly independent contributions to classification are expected. This is important because, e.g., little difference between marginal and mutually adjusted odds ratios, nearly independent contributions to classification, etc., can not

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

be taken as evidence that these instruments represent different underlying constructs. This is supported by the general consistency of empirical results with expectations, and where inconsistencies were observed, they were different than what would be expected under confounding of shared environment. Still, our results should not be taken as claims that shared environmental effects are identically 0. Rather, we suggest that a simpler or first-order explanation is that these are two very, very noisy instruments that probe a highly similar underlying construct of additive genetic liability. For phenotypes such as educational attainment, where larger and especially complex familial effects exist[53,54], more complicated inferences may be required. In previous work, we have proposed corrections for scenarios with large contributions from shared environment, namely, constructing PA-FGRS after removing close (e.g., first degree) relatives[22] or applying a post-hoc shrinkage to resulting FGRS to correct for variance inflation[21,22,55]. Future work could estimate PA-FGRS summing more complicated covariances for pedigree liability values to explicitly model shared environment, indirect genetic effects, or misspecified population assumptions due to, e.g., assortative mating.

We also highlight two important challenges with current applications of genotype- and phenotype-based genetic instruments. Firstly, the most commonly used scale-transformation, fitting a linear model to binary outcomes and transforming the observed scale variance explained to a liability scale[37], was highly biased for phenotype-based instruments in simulations (both for family history indicators and PA-FGRS) and would have resulted in qualitatively different empirical results. Caution should be used when interpreting scale-transformed variance components if sampling is complex (i.e., multi-stage) and instruments are not well-behaved (e.g., FH, PA-FGRS, or other liability scores). In these scenarios a direct approach, such as weighted probit models proposed by Lee et al[37], with appropriate multi-stage sampling weights should be preferred. Second, the expectation that PGS and PA-FGRS (or FH) have low correlations arising from poor marginal accuracies (i.e., are highly noisy measures of their underlying constructs) implies that adjusting, e.g., PGS for FH or vice versa, does not provide meaningful control for the contribution of additive genetics. Substantial residual genetic confounding should be expected after such adjustments. Models that consider or make adjustments for the expected amount of noise in an instrument[56], a quantity that could be estimated with reasonable accuracy given our results, could be considered.

Our study should be viewed in light of some important limitations. First, even though our familial records are incomplete, we may still be optimistic in terms of how powerful our phenotype-based-instruments could be in the clinic. Our prediction experiments leverage all familial information available at the end of follow-up, but that may be different than what is available at first hospital contact (i.e., an older sibling will not have access to a future onset of disease in a younger sibling; childrens' diagnoses may not be observed for parental onset). In iPSYCH, which is relatively young, this may be less of an issue than for

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

an older cohort such as the UKBiobank. Predictive studies should carefully attend to such "conditioning on the future" to better mimic timing of real clinical assessments or risk-screening. Second, the phenotype information available through national registers may be qualitatively different or limited relative to a clinical setting where different instruments are administered to patients directly. Third, even though we have large genealogical records in iPSYCH, heterogeneity in diagnosis quality over time or across regions and missing relatives could limit the informativeness of the familial diagnoses. Fourth, and related, PA-FGRS is derived under a simple, additive liability model, which has been the predominant model for dichotomous traits in genetics for nearly a century, but future work could extend both sensitivity analysis of current instruments or the instruments themselves to better incorporate secondary components of , e.g., non-additive or shared environmental variance.

Genetics as a field is rooted in resemblance among relatives and even as molecular techniques provide increasingly deep and specific access to variability across the genomes of millions of individuals, incorporating these foundational sources of data is expected to play a continued, critical role in describing genetic liability for complex traits and disorders.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Falconer, D. S. *Introduction to quantitative genetics.* (Prentice Hall, 1996).
2. Lynch, M. & Walsh, B. Genetics and analysis of quantitative traits. https://www.invemar.org.co/redcostera1/invemar/docs/RinconLiterario/2011/febrero/AG_8.pdf.
3. Goddard, M. E. & Hayes, B. J. Genomic selection. *Journal of Animal Breeding and Genetics* vol. 124 323–330 Preprint at https://doi.org/10.1111/j.1439-0388.2007.00702.x (2007).
4. Meuwissen, T., Hayes, B. & Goddard, M. Genomic selection: A paradigm shift in animal breeding. *Animal Frontiers* vol. 6 6–14 Preprint at https://doi.org/10.2527/af.2016-0002 (2016).
5. Meuwissen, T. H., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001).
6. Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E. & Visscher, P. M. Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans: Genomic Prediction. *Genetics* **211**, 1131–1141 (2019).
7. Wray, N. R., Goddard, M. E. & Visscher, P. M. Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res.* **17**, 1520–1528 (2007).
8. Torkamani, A., Wineinger, N. E. & Topol, E. J. The personal and clinical utility of polygenic risk scores. *Nat. Rev. Genet.* **19**, 581–590 (2018).
9. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
10. Schork, A. J., Schork, M. A. & Schork, N. J. Genetic risks and clinical rewards. *Nature genetics* vol. 50 1210–1211 (2018).
11. Dudbridge, F. Polygenic Epidemiology. *Genet. Epidemiol.* **40**, 268–272 (2016).
12. Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E. & Fernando, R. Additive genetic variability and the Bayesian alphabet. *Genetics* **183**, 347–363 (2009).
13. Hazel, L. N., Dickerson, G. E. & Freeman, A. E. The selection index--then, now, and for the future. *J.*

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Dairy Sci. **77**, 3236–3251 (1994).

14. Agerbo, E. *et al.* Risk of Early-Onset Depression Associated With Polygenic Liability, Parental Psychiatric History, and Socioeconomic Status. *JAMA Psychiatry* **78**, 387–397 (2021).
15. Schendel, D. *et al.* Evaluating the interrelations between the autism polygenic score and psychiatric family history in risk for autism. *Autism Res.* **15**, 171–182 (2022).
16. Agerbo, E. *et al.* Polygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the Risk for Schizophrenia: A Danish Population-Based Study and Meta-analysis. *JAMA Psychiatry* **72**, 635–641 (2015).
17. Hujoel, M. L. A., Loh, P.-R., Neale, B. M. & Price, A. L. Incorporating family history of disease improves polygenic risk scores in diverse populations. *Cell Genom* **2**, (2022).
18. Mars, N. *et al.* Systematic comparison of family history and polygenic risk across 24 common diseases. *bioRxiv* (2022) doi:10.1101/2022.07.06.22277333.
19. Kendler, K. S., Ohlsson, H., Bacanu, S., Sundquist, J. & Sundquist, K. Differences in genetic risk score profiles for drug use disorder, major depression, and ADHD as a function of sex, age at onset, recurrence, mode of ascertainment, and treatment. *Psychol. Med.* **53**, 3448–3460 (2023).
20. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. Impact of comorbidity on family genetic risk profiles for psychiatric and substance use disorders: a descriptive analysis. *Psychol. Med.* **53**, 2389–2398 (2023).
21. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. Family Genetic Risk Scores and the Genetic Architecture of Major Affective and Psychotic Disorders in a Swedish National Sample. *JAMA Psychiatry* **78**, 735–743 (2021).
22. Dybdahl Krebs, M. *et al.* A novel method for estimating familial genetic risk complements whole genome genotyping in improving our understanding of the genetic architecture of major depressive disorder. *In Preparation* (2022).
23. Musliner, K. L. *et al.* Polygenic Risk and Progression to Bipolar or Psychotic Disorders Among

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Individuals Diagnosed With Unipolar Depression in Early Life. *Am. J. Psychiatry* **177**, 936–943 (2020).

24. Musliner, K. L. *et al.* Polygenic Liability and Recurrence of Depression in Patients With First-Onset Depression Treated in Hospital-Based Settings. *JAMA Psychiatry* **78**, 792–795 (2021).
25. Musliner, K. L. *et al.* Association of Polygenic Liabilities for Major Depression, Bipolar Disorder, and Schizophrenia With Risk for Depression in the Danish Population. *JAMA Psychiatry* **76**, 516–525 (2019).
26. Hill, W. G., Goddard, M. E. & Visscher, P. M. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet.* **4**, e1000008 (2008).
27. Polderman, T. J. C. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
28. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3**, e3395 (2008).
29. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**, 507–515 (2013).
30. Wray, N. R. *et al.* From Basic Science to Clinical Application of Polygenic Risk Scores: A Primer. *JAMA Psychiatry* **78**, 101–109 (2021).
31. Lange, K. *et al.* Mendel: the Swiss army knife of genetic analysis programs. *Bioinformatics* **29**, 1568–1570 (2013).
32. So, H.-C., Kwan, J. S. H., Cherny, S. S. & Sham, P. C. Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *Am. J. Hum. Genet.* **88**, 548–565 (2011).
33. Truong, B. *et al.* Efficient polygenic risk scores for biobank scale data by exploiting phenotypes from inferred relatives. *Nat. Commun.* **11**, 3074 (2020).
34. de Los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C. & Sorensen, D. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet.* **9**, e1003608 (2013).
35. Do, C. B., Hinds, D. A., Francke, U. & Eriksson, N. Comparison of family history and SNPs for predicting risk of complex disease. *PLoS Genet.* **8**, e1002973 (2012).

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

36. Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.* **15**, 765–776 (2014).

37. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

38. McGue, M. When assessing twin concordance, use the probandwise not the pairwise rate. *Schizophr. Bull.* **18**, 171–176 (1992).

39. Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. Preprint at https://doi.org/10.1101/2020.11.30.20237768.

40. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

41. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

42. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nat. Commun.* **10**, 3328 (2019).

43. Schork, A. *et al.* Exploring contributors to variability in estimates of SNP-heritability and genetic correlations from the iPSYCH case-cohort and published meta-studies of major psychiatric disorders. *bioRxiv* 487116 (2019) doi:10.1101/487116.

44. Cai, N. *et al.* Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat. Genet.* **52**, 437–447 (2020).

45. Risch, N. Estimating morbidity risks in relatives: the effect of reduced fertility. *Behav. Genet.* **13**, 441–451 (1983).

46. Kendler, K. S. Fitness and the risk of illness and ?spectrum disorder? in offspring, parents, and siblings. *Behav. Genet.* **16**, 417–431 (1986).

47. Van Vleck, L. D. Selection bias in estimation of the genetic correlation. *Biometrics* **24**, 951 (1968).

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

48. Border, R. *et al.* Assortative mating biases marker-based heritability estimators. *Nat. Commun.* **13**, 660 (2022).

49. Border, R. *et al.* Cross-trait assortative mating is widespread and inflates genetic correlation estimates. *Science* **378**, 754–761 (2022).

50. Pedersen, C. B. The Danish Civil Registration System. *Scand. J. Public Health* **39**, 22–25 (2011).

51. An, U. *et al.* Deep learning-based phenotype imputation on population-scale biobank data increases genetic discoveries. *bioRxiv* (2022) doi:10.1101/2022.08.15.503991.

52. Dahl, A. *et al.* Phenotype integration improves power and preserves specificity in biobank-based genetic studies of MDD. *bioRxiv* (2022) doi:10.1101/2022.08.15.503980.

53. Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science* **359**, 424–428 (2018).

54. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat. Genet.* **54**, 437–449 (2022).

55. Kendler, K. S., Ohlsson, H., Sundquist, J. & Sundquist, K. The patterns of family genetic risk scores for eleven major psychiatric and substance use disorders in a Swedish national sample. *Transl. Psychiatry* **11**, 326 (2021).

56. Pingault, J.-B. *et al.* Genetic sensitivity analysis: Adjusting for genetic confounding in epidemiological associations. *PLoS Genet.* **17**, e1009590 (2021).

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Online Methods

iPSYCH2015 Case-Cohort Study

The Lundbeck Foundation initiative for Integrative Psychiatric Research (iPSYCH)[1,2] samples singleton births between 1981 and 2008, alive on their first birthday, with mothers residing in Denmark (N=1,657,449). The iPSYCH 2012 case-cohort includes 86,189 individuals (30,000 random population controls; 57,377 psychiatric cases)[1], while the the iPSYCH 2015i case-cohort includes an independent 56,233 individuals (19,982 random population controls; 36,741 psychiatric cases)[1,2]. Taken together, these two components make up the iPSYCH 2015 case-cohort study. DNA was taken from dried blood spots housed at the Danish Neonatal Screening Biobank[3]. The Infinium PsychChip v1.0 array (2012) or the Global Screening Array v2 (2015i) were used for genotyping. Register based diagnoses are made by licensed practitioners during in- or out- patient specialty care (diagnoses or treatments assigned in primary care are not included) and are obtained from the Danish Psychiatric Central Research Register (PCRR)[4] and the Danish National Patient Register (DNPR)[5]. Unique identifiers of the Danish Civil Registration System[6] enables linkage across population registers, to parental identifiers, and to the neonatal blood spots. The use of this data follows standards of the Danish Scientific Ethics Committee, the Danish Health Data Authority, the Danish Data Protection Agency, and the Danish Neonatal Screening Biobank Steering Committee. Data access was via secure portals in accordance with Danish data protection guidelines set by the Danish Data Protection Agency, the Danish Health Data Authority, and Statistics Denmark.

Imputation and quality control of genotypes followed custom, mirrored protocols [7]. BEAGLEv5.1[8,9] was used to phase and impute the data in conjunction with reference haplotypes from the Haplotype Reference Consortium v1.1 (HRC)[10]. SNPs were checked for missing data across SNPs and individuals, deviations from Hardy-Weinberg equilibrium in controls, abnormal heterozygosity, minor allele frequency (MAF), batch artifacts, and imputation quality. 7,649,999 imputed allele dosages were retained for analysis. Samples were checked for genotype-phenotype sex discordance and abnormal heterozygosity. KING[11] was used to estimate kinship and ensure no second degree or higher relatives remained within either cohort or across cohorts. The smartpca module of EIGENSOFT[12] was used to estimate PCs and a classifier was used to remove individuals distant from those with parents and four grandparents born in Denmark. For this study, in particular, we defined the following disorders according to ICD-10 codes associated with at least one registered treatment: major depressive disorder (MDD; F32-3), bipolar disorder (BPD; F31), schizophrenia (SCZ; F20), autism spectrum disorders (ASD; F84, F84.1, F84.5, or F84.9), and attention deficit hyperactivity disorder (ADHD; F90.0). In total, we retain, in iPSYCH 2012 24,266

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

random probands, 14,970 ADHD probands, 12,101 ASD probands, 1581 BPD probands, 18238 MDD probands, 2624 SCZ probands, and in iPSYCH 2015i 15,381 random probands, 7499 ADHD probands, 5600 ASD probands, 1141 BPD probands, 8668 MDD probands, 2721 SCZ probands.

iPSYCH 2015 case-cohort genealogies

Details of the iPSYCH 2015 case-cohort genealogy are described elsewhere [13] and were constructed following previous work [14]. Briefly, 2,066,657 unique relatives were obtained using mother-father-offspring linkages for the 141,265[15] (before QC) genotyped probands. A population graph was constructed using *kinship2*[16] and *FamAgg*[15] packages for R with edges defined by recorded trios. Relatedness coefficients are estimated by summing unique paths between two individuals, weighted by (0.5)^(number of edges in the path)[17]. Same-sex twins were given kinship of 0.375 and maternal siblings with missing paternal records were given kinship of 0.25 according to prior analysis [13].

Diagnoses made between 1968 and 1994 are limited to in-patient contacts and ICD-8 codes and were assigned to disorder labels to match ICD-10 groups as follows: major depressive disorder (MDD; 296.09, 296.29, 298.09, or 300.49, but not 296.19, 296.39, or 298.19), bipolar disorder (BPD; 296.19, 296.39, or 298.19), schizophrenia (SCZ; 295.x9, excluding 295.79), autism spectrum disorders (ASD; 299.01-299.04), and attention deficit hyperactivity disorder (ADHD; 308.01)[1]. FH variables were defined as having at least one affected parent, half sibling, or full sibling.

Narrow sense heritability (h^2) for each psychiatric disorder was estimated using relatives in the genealogies of the random population sample. To account for cohort effects, we selected 250,057 sibling pairs in these genealogies such that both siblings were born in the same decade (either, 1960-1970, 1970-1980, 1980-1990, 1990-2000, 2000-2010). We then estimated the sib-pair tetrachoric correlation for each disorder within each decade, combining the estimates in an inverse variance weighted meta-analysis. The meta-analyzed tetrachoric correlation and it's standard deviation were divided by the sibling kinship (0.5) to give an estimate of h^2 and compute 95% confidence intervals.

A simple, additive polygenic liability threshold model

Our approach is based on a simple additive polygenic liability threshold model, which is tractable, useful, and commonly employed. Briefly, We assume that for a fully observed individual their binary disease status, , is the result of a latent, normally distributed liability, . is expressed when is above a critical threshold, , that is the upper tail standard normal quantile corresponding to the lifetime prevalence

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

of the disease, K_{pop}.

$$D_i = \begin{cases} 1, & L_i \geq t \\ 0, & L_i < t \end{cases}, \quad t = \Phi^{-1}\left(1 - K_{pop}\right)$$

The liability is the sum of a genetic value, , and an environmental deviation, , and we assume no G-E interactions and no contributions from familial environment. is an additive function of genetic factors, , with true allelic average effects, , and explains h^2 proportion of population variance in *L*. The genetic value may itself have two components, a GWAS related SNP component, , explaining h^2_{SNP} proportion of variance in liability, and genetic residual, explaining $h^2 - h^2_{SNP}$ that may represent incomplete linkage or non-GWAS SNP genetic factors.

Computing Pearson-Aitken Family Genetic Risk Scores (PA-FGRS)

PA-FGRS estimates the genetic component of latent disease liability under assumptions of a modified polygenic liability threshold model that relaxes the assumption above that individuals are fully observed by defining

$$Y_i = \begin{cases} Bernoulli\left(\dfrac{K_i}{K_{Pop}}\right), & D_i = 1 \\ 0, & D_i = 0 \end{cases}$$

where is the cumulative incidence of the disease in the age window during which individual *i* was observed. The key difference from above, is that, here, the probability of a true disease status being observed (i.e., that = 1) is related to the proportion of the lifetime incidence for which an individual was observed. As a result, we model the expected liabilities of cases and controls as *mixtures* of truncated normals such that observed controls are treated as a probabilistic mixture of a true case and a true control.

PA-FGRS estimates the genetic component of latent disease liability under this model for a proband, *p*, from the observed disease statues of their relatives, , a modified relatedness matrix, , describing the relationship between trait liability in relatives and genetic liability of the proband, and

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

known parameters, , such as the age of each relative at the end of follow up, and the age-specific incidence of the trait.

$$\Sigma = \begin{bmatrix} h_l^2 & h_l^2 r_{p,1} & \cdots & h_l^2 r_{p,n_{rel}} \\ h_l^2 r_{p,1} & 1 & \cdots & h_l^2 r_{p,n_{rel}} \\ \vdots & \vdots & \ddots & \vdots \\ h_l^2 r_{p,n_{rel}} & h_l^2 r_{p,n_{rel}} & \cdots & 1 \end{bmatrix}$$

where is the relatedness coefficient between the proband and their i-th relative. Briefly, we estimate according to the Pearson-Aitken (PA) selection formula to provide a direct, efficient, and highly accurate solution for an estimator that otherwise requires expensive resampling or numerical integration techniques [13].

Our PA approach proceeds as follows. First, the vector of mean expected liabilities for all relatives, , and the expected covariance among liabilities of relatives, , are initialized to expectations in the population assuming no observations have been made: and . Then, and are updated iteratively, conditioning on each relative in sequence. At each step, the current expected liability of the selected relative, , is updated to a conditional expectation, , based on the current estimate, the observed disease status of the relative, and other known parameters.

with , the PDF of the standard normal distribution, , the CDF of the standard normal distribution, , and if or 1 if . All remaining (i.e., unselected) liabilities are then updated conditional on the updated liability of the selected relative.

This can be thought of as propagating a non-linear (i.e., liability threshold model appropriate) weighting of the observed disease status of the selected relative into the expected liabilities of their genetic relatives, including the proband.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Similarly, the liability covariance, Ω, is updated at each step by first updating the variance of the mean expected liability for the selected relative, and then the rest of the covariance conditional on this updated variance for the selected relative.

$$
\begin{aligned}
{}^{*}_{,} = {} & \left(\left(\quad - \sqrt{,}\,\frac{0}{0} \right)^{2} + {}_{,}\left(1 - \frac{0}{0} - \left(\frac{0}{0}\right)^{2}\right)\right) + \\
& (1 -)\left(\left(\quad + \sqrt{,}\,\frac{0}{1-0} \right)^{2} + {}_{,}\left(1 + \frac{0}{1-0} - \left(\frac{0}{1-0}\right)^{2}\right)\right) - (\mu^{*})^{2}
\end{aligned}
$$

$$
{}^{*} = - {}_{,}\left({}^{-1}_{,} - {}^{-1}_{,}\,{}^{*}_{,}\,{}^{-1}_{,}\right)_{,}
$$

The procedure is repeated until only the proband remains unselected (i.e., proband liability has been conditioned on all relatives). For further details see Krebs et al [13] and Supplementary Information. We computed these PA-FGRS for individuals from the iPSYCH 2012 and iPSYCH 2015i cohorts separately (Supplementary Table S1) using trait parameters described in Supplementary Table S5.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Computing polygenic scores (PGS)

PGS approximate the genetic value under the polygenic liability threshold model by substituting a set of observed SNP marker genotypes (') for an individual, i, for the unknown generative genetic factors (X), and using transformations of per allele effect estimates from a training GWAS ($\hat{\beta}$) as a surrogate for true allelic effects (β).

$$\hat{}_{,} = \sum_{=1} \quad \hat{}$$

We computed PGS combining imputed dosages with $\hat{\beta}$ estimated using SBayesR to update results provided by large GWAS. For SBayesR we followed the developer recommendations using provided UKBiobank LD matrix estimated for 2.8 million SNPs and with parameters: gamma=0.0,0.01,01,1, pi=0.95,0.02,0.02,0.01, burn-in=5000, chain-length=25000, exclude-mhc, impute-n. For BPD [18], MDD [19], and SCZ [20] we use large published GWAS that are independent of iPSYCH. For ADHD and ASD we use the results of mixed effects model GWAS [21] from the complementary iPSYCH cohort (e.g., beta from GWAS in iPSYCH 2015i for PGS in iPSYCH 2012).

Statistical model fitting to iPSYCH data

Variance in liability explained by combinations of genetic instruments was estimated with weighted multivariate generalized linear models incorporating a probit link function. A baseline model included age, gender, and the first ten principal components of an estimated genetic relatedness matrix with subsequent models including combinations of PGS, FH, and PA-FGRS. The liability scale variance for each model was estimated according to Lee et al [22] as,

$$^{2} = \frac{(^{*}\hat{}\_)}{(^{*}\hat{}\_)+1}$$

where $\hat{b}_{IPW-probit}$ is a vector of coefficients from the inverse probability weighted probit regression of the disease status on the matrix of covariates X. For the weighting, two-stage sampling weights were computed as the product of the case-control sampling weights suggest by Lee et al [22] (unit weights in controls and $\frac{(1-w)K_{Pop}}{w(1-K_{Pop})}$ in cases) and a second, study specific,

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

probability weight to reflect selection out of the sample due to relatedness pruning (Supplementary Note S1). X^*, a random subset of X such that $w *= K_{Pop}$.

Reported liability variance explained is after subtracting that of a baseline covariates only model. Empirical p-values were used to test for significance of changes in liability variance explained of nested models as the proportion of times out of 100,000 that the empirical change associated with adding a variable to a model was larger than adding a randomly permutation of the variable. Significance adjusted to p < 0.05/110 model contrasts = 4.55 x 10-4. Unadjusted and adjusted odds ratios (OR) were estimated using multivariate generalized linear models with a logit link function. Pearson's correlations among genetic instruments were estimated in the random subcohort of each iPSYCH cohort and not adjusted for covariates.

Estimating the expected Accuracy of Pearson-Aitken Family Genetic Risk Scores (PA-FGRS)

Exact Solution. The expected accuracy of our PA-FGRS estimator was derived fully in the Supplementary Information and is defined as the product of the inverse of the heritability and the variance of the unbiased estimator ^_,

$$(,\hat{\_}) = \sqrt{(^2)^{-1} \sum_{=1} \quad ((\quad = \ | ,) (| \ = , ,)^2)}$$

where D is -variate thresholded gaussian, and the conditional expectation of G is the PA estimator above. Where $\{_1, _2, \ldots, \}$ is the set of $= ()^2$possible configurations of disease in the selected pedigree. For pedigrees with relatively few numbers of relatives we can solve this by a explicit calculation of the -variate integral over the configurations (see code availability for our function *fgrs_accuracy(…, method="pa" ,estimate= "theory")*), but for larger pedigrees it is intractable. This full solution is used for the FGRS expectation results presented in Figures 2-4.

Linearly Approximated Solution. Alternatively, one may use more tractable expectations of accuracy from a linear estimator (i.e., expected breeding value) in place of the full liability scale estimator PA-FGRS, similarly defined as the product of the inverse of the heritability and the variance of^,

$$(,\hat{}) = \sqrt{(^2)^{-1} \quad ^{-1}}$$

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with being a 1 x n_{rel} vector describing the expected covariance between the proband genetic liability and the true disease status of the i-th relative such that $= (,) = {}^{2}()$. **P** is an n_{rel} x n_{rel} matrix describing the expected covariance in disease status among the relatives, which can be approximated such that $_{,\neq} = (,) \approx ()^{2\,2}\left(1 + \frac{22}{2}\right)$ and $_{,=} = () = (1 -)$, following Golan et al [23]. Here, is the relatedness coefficient between the proband, p, and their i-th relative, is relatedness coefficient between the i-th and j-th relatives, and is the liability threshold value, $= {}^{-1}(1 -)$. We show these approximation to be a good approximation of the full solution in smaller pedigrees (Supplementary Figures S11 and S12) and use these formulas to compute the expected accuracy of PA-FGRS for iPSYCH pedigrees as presented in Figure 6. When individuals are not fully observed, we use $\tilde{} = (,) = -{}^{2}()$ and $\tilde{}$ where and $\tilde{}_{,=} = (1 -)$ and $\tilde{}_{,\neq} \approx ()^{2\,2}\left(1 + \frac{22}{2}\right)\frac{}{2}$ to estimate this accuracy as,

$$(,\hat{}) = \sqrt{(^{2})^{-1\tilde{}}\ \ {}^{-1\tilde{}}}$$

Linearly Approximated Number of Full Sibling Equivalents. In the special case of pedigree containing one relative type (e.g., Full siblings of the proband), this linear approximation reduces to,

$$(,\hat{}) \approx \sqrt{\frac{2}{\frac{(1 -)}{()^{22}} + (- 1)\left(1 + \frac{22}{2}\right)}}$$

Here, is the relatedness coefficient between the relatives and the proband, is the relatedness coefficient among the relatives which may be different than (e.g., as for half-sibling offspring where =0.5, =0.25). To derive the for any predictor, we set this equation equation to its observed or expected accuracy, which for an empirical predictor can be recovered its performance as $\sqrt{^{2}/^{2}}$, and solve this equation for the value of when $= = 0.5$. This approximation was used to estimate for iPSYCH pedigrees, simulated pedigrees, and to equate observed PGS sample size to equivalent PA-FGRS .

Estimating the accuracy of polygenic scores (PGS)

The expected accuracy of polygenic scores has been described elsewhere by, e.g., Wu et al [24], Daetwyleer et al [25], and Dudbrige [26] under related, but not identical assumptions (see

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Information). We use the formulas from Wu et al [24] to estimate the expected squared accuracy of a PGS as

$$N_q = \begin{cases} N, & \textit{Quantitative trait} \\ \left(N_{case} + N_{control}\right) w(1-w) \left(\dfrac{\phi(t)}{K_{Pop}\left(1-K_{Pop}\right)}\right)^2, & \textit{Binary trait} \end{cases}$$

where M is the effective number of independent markers. Shrinkage based estimators can gain power in the context of non-infinitesimal models (e.g., ldpred [27]) and we can evaluate,

$$\left(\frac{N_q h^2}{N_q h^2 + Mp}\right)\left(\frac{1}{\sigma_1^2}\right)\int_{-\infty}^{\infty} x^2 \frac{pf\left(x \mid \mu, \sigma_1^2\right)^2}{pf\left(x \mid \mu, \sigma_1^2\right) + (1-p) f\left(x \mid \mu, \sigma_2^2\right)} \mathrm{d}x$$

where a non-standardized Gaussian – – , and with , – –, and –.

Expected correlation between PA-FGRS and PGS

Under a simple liability framework as we propose, the and are statistically independent conditional on the true genetic value (G)[28]. In this scenario, the expected correlation is

Where and are the liability variance explained by the estimated genetic values and is the liability scale heritability.

Expected joint performance of PA-FGRS and PGS

The expected semi-partial accuracy in liability of the PA-FGRS, or expected correlation between the PA-FGRS and liability after accounting for the PGS, can be written as,

which allows us to predict the expected joint performance of a PGS and PA-FGRS fit in the same model as,

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

$$^{2}_{,+-} = ^{2}_{,} + ^{2}_{,-|}$$

Comparisons of expected and observed performance

Expected performance and relationships of genetic instruments depends on both assumed and estimated trait parameters that are described in Supplementary Table S5. The expected performance of PA-FGRS in iPYSCH were computed using the mean of the linear approximation for the expected accuracy of PA-FGRS for each iPSYCH pedigree using known values for ^{2} and .

The expected performance of the PGS in iPSYCH were calculated by first estimating, using the non-shrinkage estimator, the expected squared accuracy of the PGS in the training sample following the expectations of the non-shrinkage estimator above assuming M=60,000, and using training sample parameters N_{case}, $N_{controls}$, $^{2}_{,}$, . The expected out of population test sample performance (i.e., in iPSYCH) was then estimated as,

$$^{2}_{,} = (\hat{,})^{2} \quad * ^{2}_{,,} * ^{2}_{,,}$$

where $^{2}_{,,}$ is the squared genetic correlation between the phenotype as defined in the training GWAS and in iPSYCH, and $^{2}_{,,}$ is the SNP-based heritability for the iPSYCH trait.

The expected correlation between the observed PGS and PA-FGRS were estimated as above using provided values for ^{2}.

The population GWAS sample size expected to result in a PGS with performance equal to that of the empirical PA-FGRS instruments, was calculated by setting the non-shrinkage based estimator above equal to the observed value for squared accuracy of the PA-FGRs, (i.e., $^{2}_{,-} / ^{2}$) and solving for assuming = . The calculations assumed the training GWAS was performed in the same population that the PGS was applied within (i.e., $^{2}_{,,} = ^{2}_{,,}$ and $_{,,}$ = 1), that m=60,000 and known .

The pedigree expected to result in a PA-FGRS with performance equal to that of the empirical PGS instruments, was calculated by setting the linear approximation estimator for full sibling pedigrees above equal to the observed value for squared accuracy of the PGS, (i.e., $^{2}_{,} / ^{2}_{,}$) and solving for assuming and ^{2}.

Code availability

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

PA-FGRS, predicted performance, and simulations are implemented in R code available at https://github.com/BioPsyk/PAFGRS.

Online Methods References

1. Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).
2. Bybjerg-Grauholm, J. *et al.* The iPSYCH2015 Case-Cohort sample: updated directions for unravelling genetic and environmental architectures of severe mental disorders. *bioRxiv* (2020) doi:10.1101/2020.11.30.20237768.
3. Nørgaard-Pedersen, B. & Hougaard, D. M. Storage policies and use of the Danish Newborn Screening Biobank. *J. Inherit. Metab. Dis.* **30**, 530–536 (2007).
4. Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand. J. Public Health* **39**, 54–57 (2011).
5. Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish National Patient Register. *Scand. J. Public Health* **39**, 30–33 (2011).
6. Pedersen, C. B. The Danish Civil Registration System. *Scand. J. Public Health* **39**, 22–25 (2011).
7. Appadurai, V. *et al.* Accuracy of haplotype estimation and whole genome imputation affects complex trait analyses in complex biobanks. *bioRxiv* 2022.06.27.497703 (2022) doi:10.1101/2022.06.27.497703.
8. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
9. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
10. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

11. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

12. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

13. Dybdahl Krebs, M. *et al.* A novel method for estimating familial genetic risk complements whole genome genotyping in improving our understanding of the genetic architecture of major depressive disorder. *In Preparation* (2022).

14. Athanasiadis, G. *et al.* A comprehensive map of genetic relationships among diagnostic categories based on 48.6 million relative pairs from the Danish genealogy. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).

15. Rainer, J. *et al.* FamAgg: an R package to evaluate familial aggregation of traits in large pedigrees. *Bioinformatics* **32**, 1583–1585 (2016).

16. Sinnwell, J. P., Therneau, T. M. & Schaid, D. J. The kinship2 R package for pedigree data. *Hum. Hered.* **78**, 91–93 (2014).

17. Wright, S. Coefficients of Inbreeding and Relationship. *Am. Nat.* **56**, 330–338 (1922).

18. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).

19. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

20. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).

21. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

22. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

23. Golan, D., Lander, E. S. & Rosset, S. Measuring missing heritability: inferring the

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

contribution of common variants. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5272–81 (2014).

24. Wu, T., Liu, Z., Mak, T. S. H. & Sham, P. C. Polygenic power calculator: Statistical power and polygenic prediction accuracy of genome-wide association studies of complex traits. *Front. Genet.* **13**, (2022).
25. Daetwyler, H. D., Villanueva, B. & Woolliams, J. A. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* **3**, e3395 (2008).
26. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).
27. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
28. Wright, S. The Method of Path Coefficients. *aoms* **5**, 161–215 (1934).

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

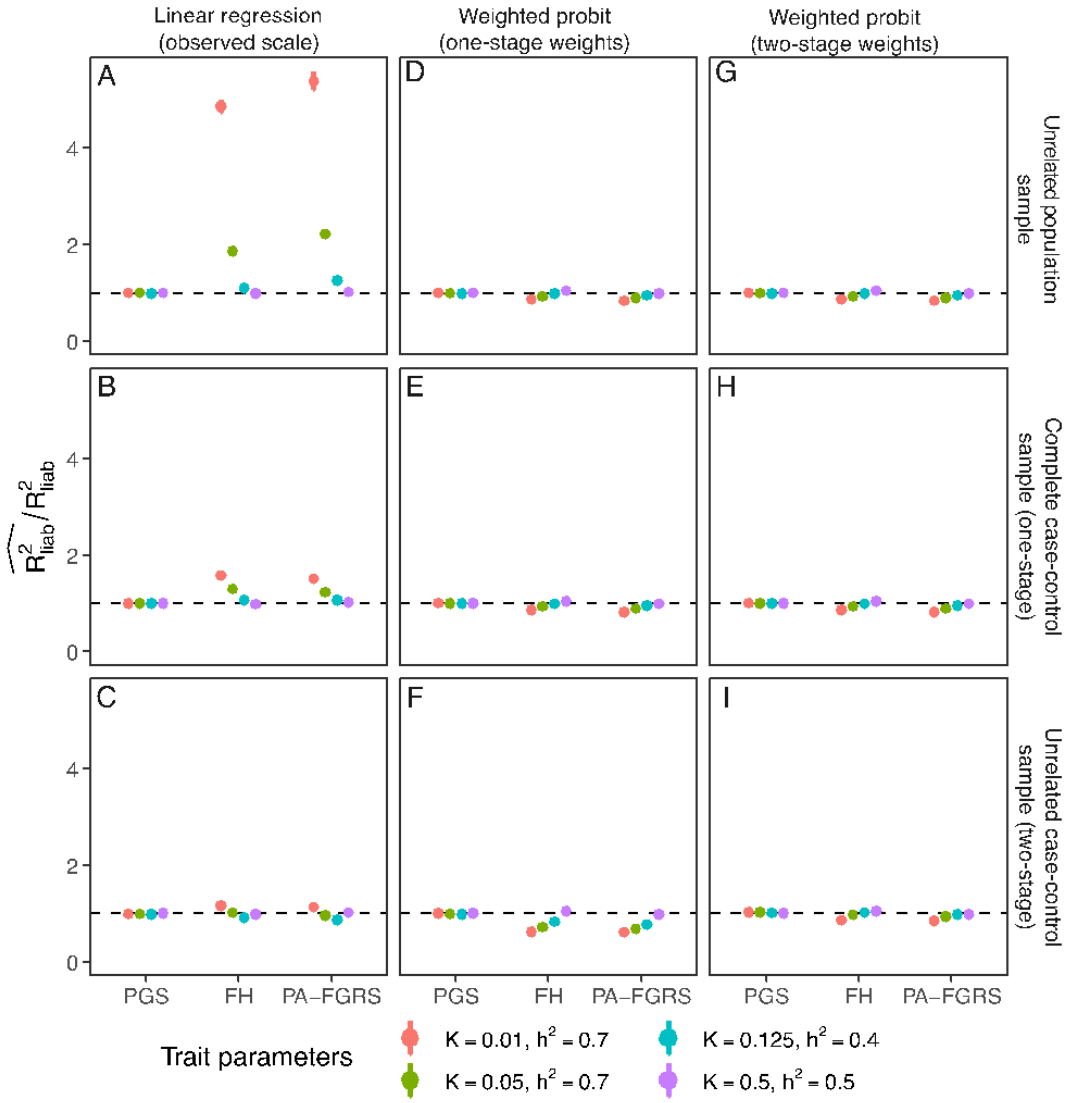
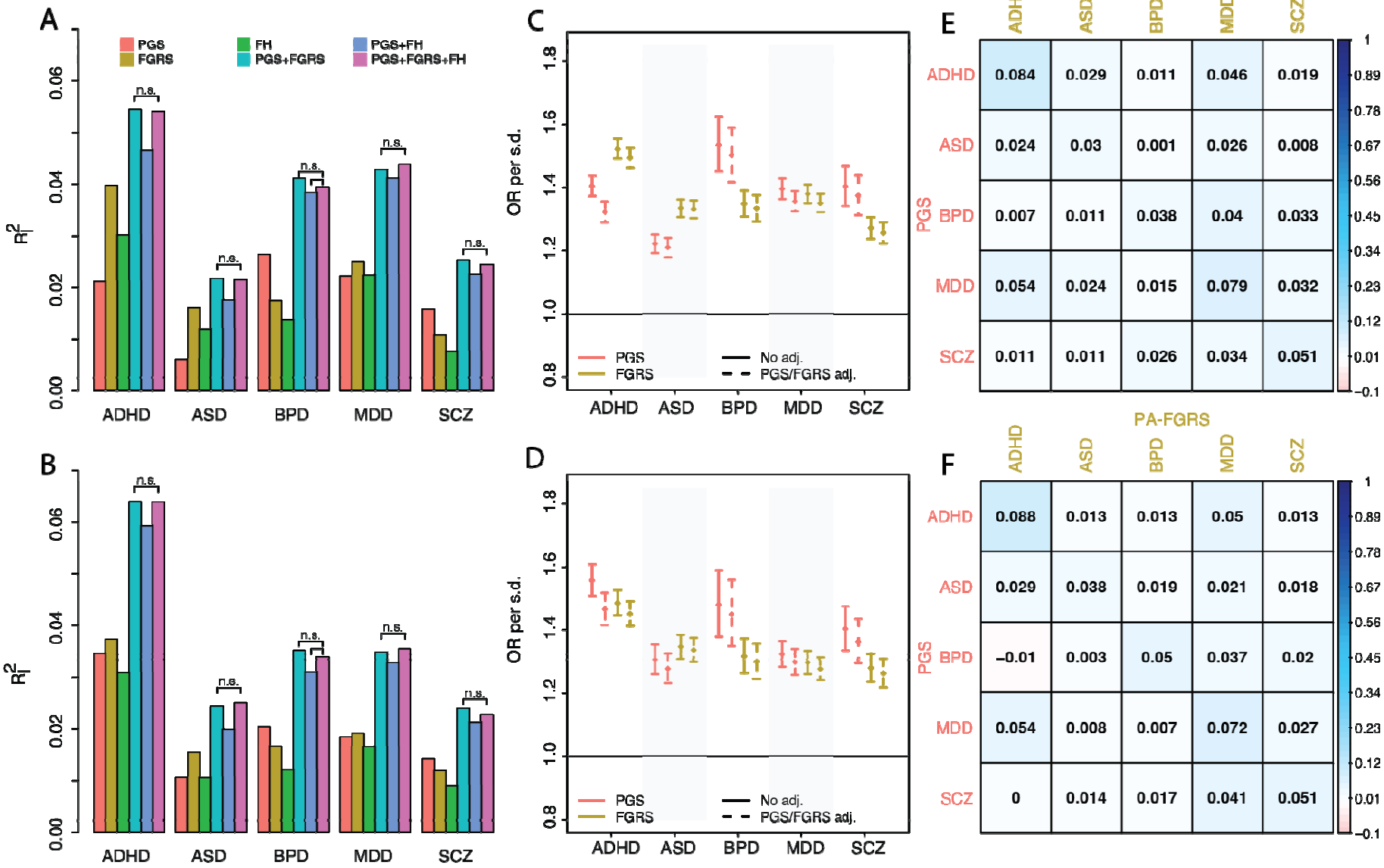


Figure 1. Common estimates of variance explained on the liability scale are biased for phenotype-based genetic instruments. In a simulated population of families, we compare different strategies for estimating liability scale variance explained by PGS, FH, and PA-FGRS, under different sampling schemes. The dashed line in each plot represents the ratio of estimated to true liability scale variance explained where values deviating from 1 indicate biased estimates. Transforming variance explained from the observed scale (A,B,C) is only appropriate for PGS, but produces biased estimates for FH and PA-FGRS across nearly all sampling and trait configurations tested. Directly estimating liability scale variance using a weighted probit regression to account for one-stage case-control sampling (D,E,F) corrects for this bias, but overcorrects when two-stage sampling (e.g., pruning of relatives) is applied (F). The most robust estimates were found by directly estimating liability scale variance with a weighted probit regression that accounts for the two-stage sampling associated with case oversampling followed by relatedness pruning (G,H,I). When estimating variance explained on the liability scale for phenotype-based instruments, accounting for sampling beyond case-control proportion may be critical. Confidence intervals are generally narrower than and thus contained within each circle. PGS, polygenic score; FH, indicator of first-degree family history; PA-FGRS, Pearson-Aitken Family Genetic Risk Score; R^2_{liab}, Liability scale variance explained; K, lifetime prevalence; h^2, narrow-sense heritability.



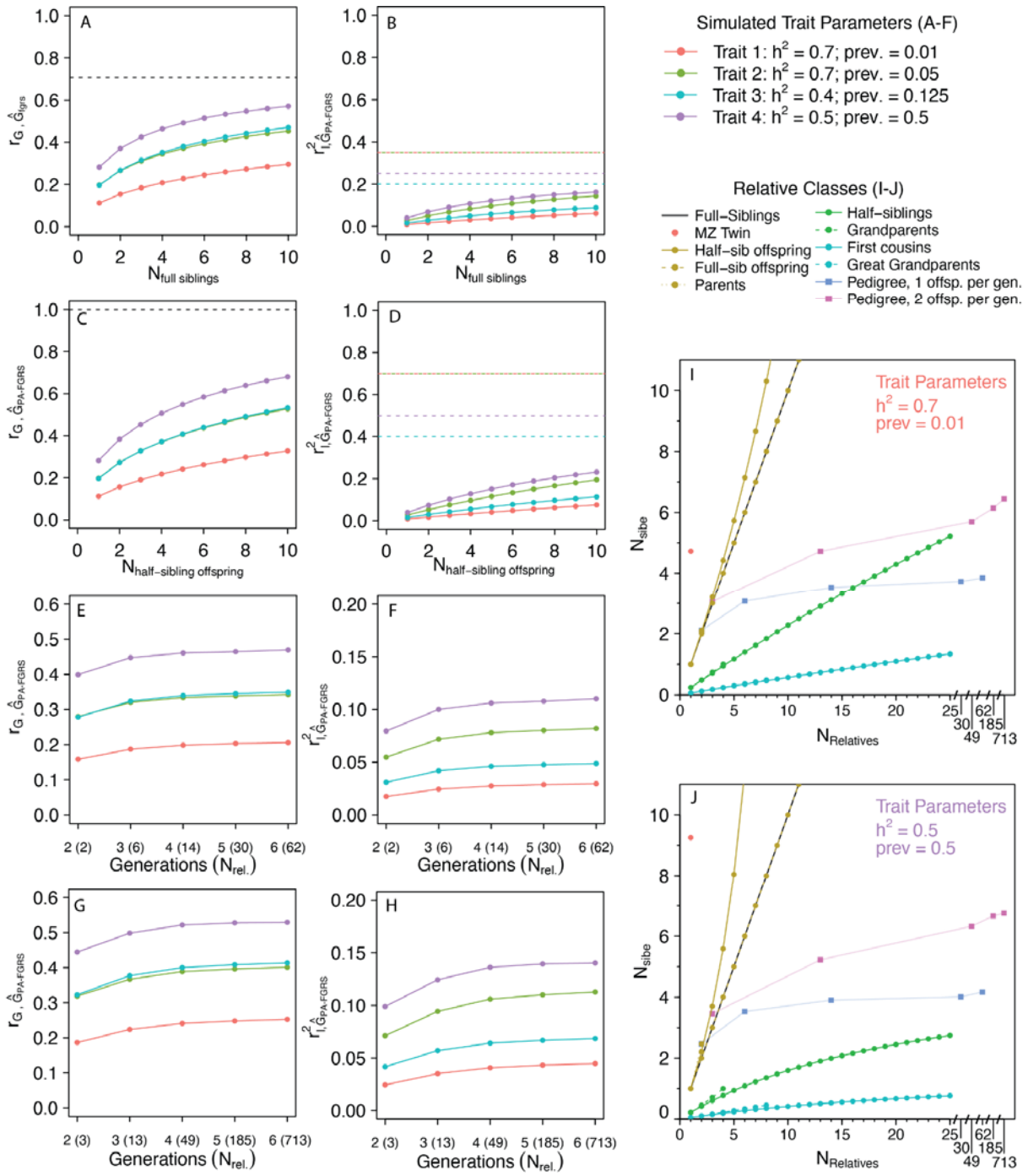
medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2. Genotype- and phenotype-based genetic instruments explain nearly independent components of liability to psychiatric disorders.

Multiple genetic instruments, alone and in combination, explain variance on the liability scale for five major psychiatric disorders after accounting for covariates. The liability variance explained by each instrument alone and in combination are shown for each of five disorders in two independent cohorts, (A) the iPSYCH2012 cohort and (B) the iPSYCH2015i cohort. Increases in variance explained associated with adding each variable to each model was tested via permutations and all were significant after Bonferroni correction ($p < 0.05/110 = 4.55 \times 10^{-4}$) unless annotated with n.s.. The unadjusted (solid line) and adjusted (dashed line) OR for PGS and PA-FGRS in logistic models are similar for all disorders in both the (C) iPSYCH2012 and (D) iPSYCH2015i cohorts. Pearson correlations among PGS and PA-FGRS instruments are modest, in general, but slightly larger within disorder for both the (E) iPSYCH2012 and (F) iPSYCH2015i random population sub-cohorts. ADHD, Attention deficit hyperactivity disorder; ASD, autism spectrum disorder; BPD, bipolar disorder; MDD, major depressive disorder; SCZ, schizophrenia; PGS, polygenic scores; FH, family history indicators; PA-FGRS, Pearson-Aitken Family Genetic Risk Scores; n.s., no significant difference; OR, odds ratio; adj., adjustment; s.d., standard deviation.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3. The expected and asymptotic accuracy and performance of PA-FGRS depend on pedigree structure, trait architecture, and realistic bounds on numbers of observed relatives.

The expected accuracy of PA-FGRS ($\hat{}_{,A-}$) from records of increasing numbers of full siblings (A) varies by trait h^2 and prev. with a theoretical asymptote at $\sqrt{0.5} \approx 0.71$ (dashed line), however, for realistic bounds on the number of full siblings ($N_{Full\ sibliings} < 10$) the observed accuracy is expected to be well below this asymptotic value. There is a corresponding variability in expected performance of full sibling PA-FGRS ($^2{}_{,\hat{}}$) (B) which can reach a theoretical asymptote of $^2(^2_{\hat{},-}) = 0.5^2$ but in realistic scenarios ($N_{Full\ sibliings} < 10$) the observed performance is correspondingly lower than the theoretical asymptote. PA-FGRS using records of half-sibling offspring show the same trends in terms of variability in accuracy (C) and performance (D) by h^2 and prev., but these, uniquely, have much higher theoretical asymptotes at 1 and ^{2}, respectively. With realistic numbers of relatives, the accuracy and performance, however, are similarly expected to be below the asymptote. Complex, multi-generation pedigrees of different depth that assume one (E,F) or two (G,H) children per mate pair are bounded in accuracy (E,G) and performance (F,H) by the same principles. Here, practical asymptotes occur at approximately four generations of pedigree depth and these levels are well below theoretical maxima (the same as depicted in A,B). The expected accuracy of a PA-FGRS based on records from any arbitrary set of relatives can be equated to that of a PA-FGRS incorporating only full sibling records to provide a per-pedigree metric, the number of full sibling equivalents (N_{sibe}), that can be used to compare pedigrees and provide benchmark estimates of accuracy, (I,J). Relative to full siblings, distant relatives are more important or rare traits (I) and close relatives are more important for common traits (J). PA-FGRS, Pearson-Aitken Family Genetic Risk Scores; h^2, narrow-sense heritability; $\hat{}_{,PA-}$, PA-FGRS accuracy; $^2{}_{,\hat{}_{A-}}$, PA-FGRS performance; prev, lifetime prevalence; l, Liability; N_{sibe}, Number of full sibling equivalents; Rel., Relatives; Sib., Siblings; Equiv., Equivalents; MZ, Monozygotic; Offsp., Offspring; gen., generation

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

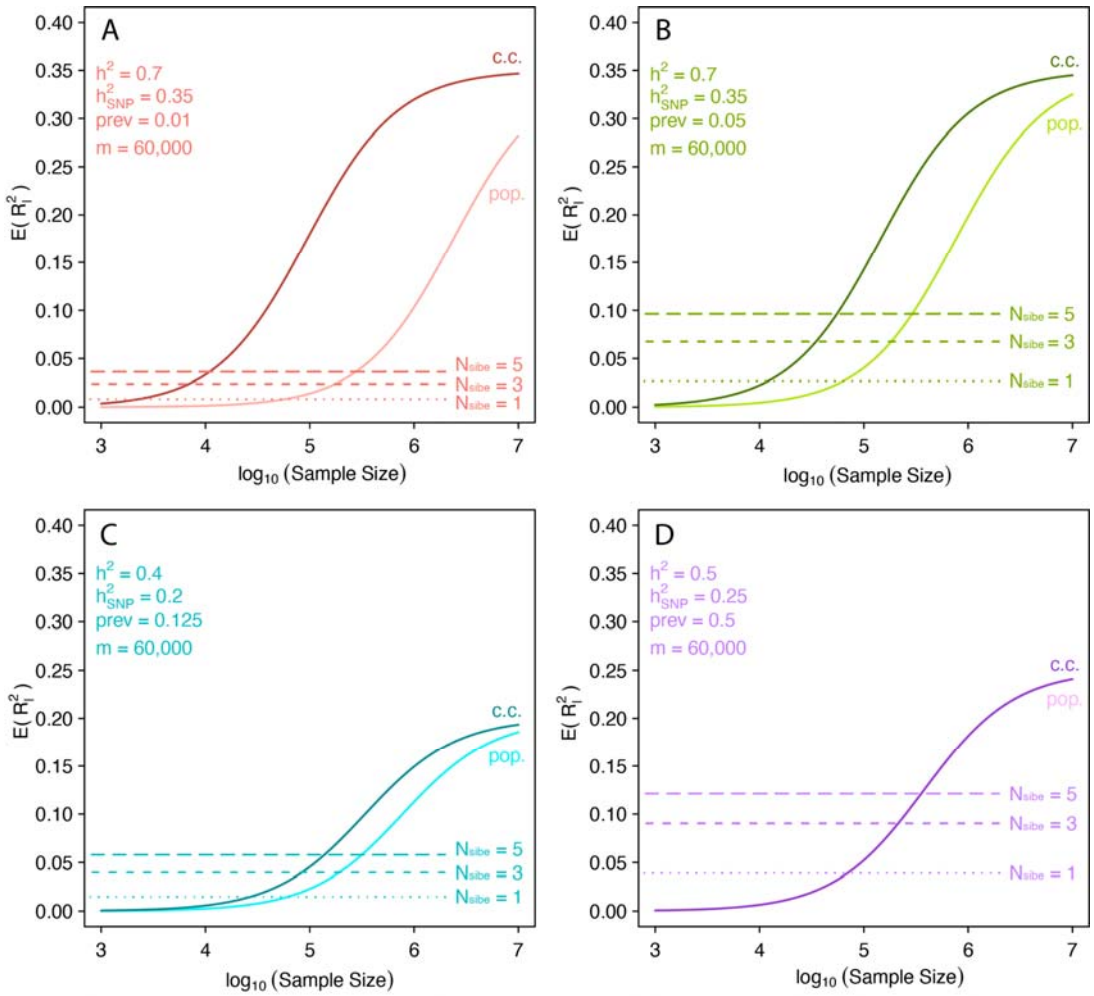
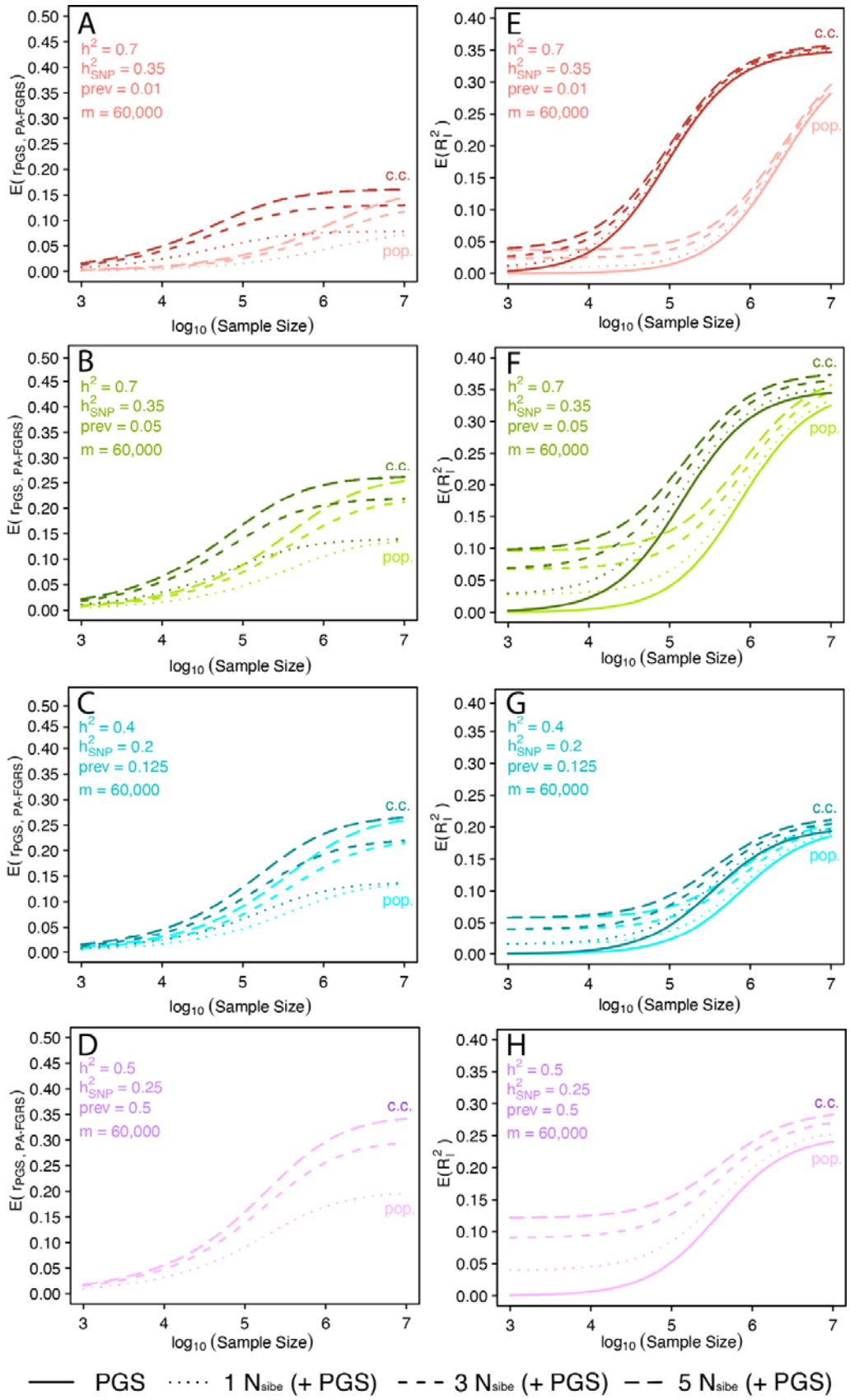


Figure 4. PA-FGRS are expected to outperform PGS unless training GWAS are large.

The expected performance of PGS will vary as a function of , prevalence, and training GWAS sample size (A,B,C,D). Here, we assume trait in the test sample is equal to in the training GWAS, the between the test and training sample is 1, the number of markers in the PGS is m=60,000, and the effective proportion of causal markers is p=1. The expected performance of PA-FGRS with N_{sibe}=1,3, or 5 (dashed lines) is not dependent on a training sample and therefore outperform PGS when large GWAS are unavailable. Across these simulated traits, the size of a population sampled training GWAS needed to achieve performance equal to PA-FGRS with N_{sibe}=1,3, or 5 are qualitatively similar (x-axis value where solid and dashed lines intersect), but PA-FGRS do relatively better for common traits (dashed lines vs. in A,B vs C,D). PA-FGRS, Pearson-Aitken Family Genetic Risk Scores; PGS, polygenic score; h^2, narrow-sense heritability; , SNP heritability; , genetic correlation; prev, lifetime prevalence; N_{sibe}, number of full sibling equivalents; c.c., case-control sampling; pop., population sampling.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

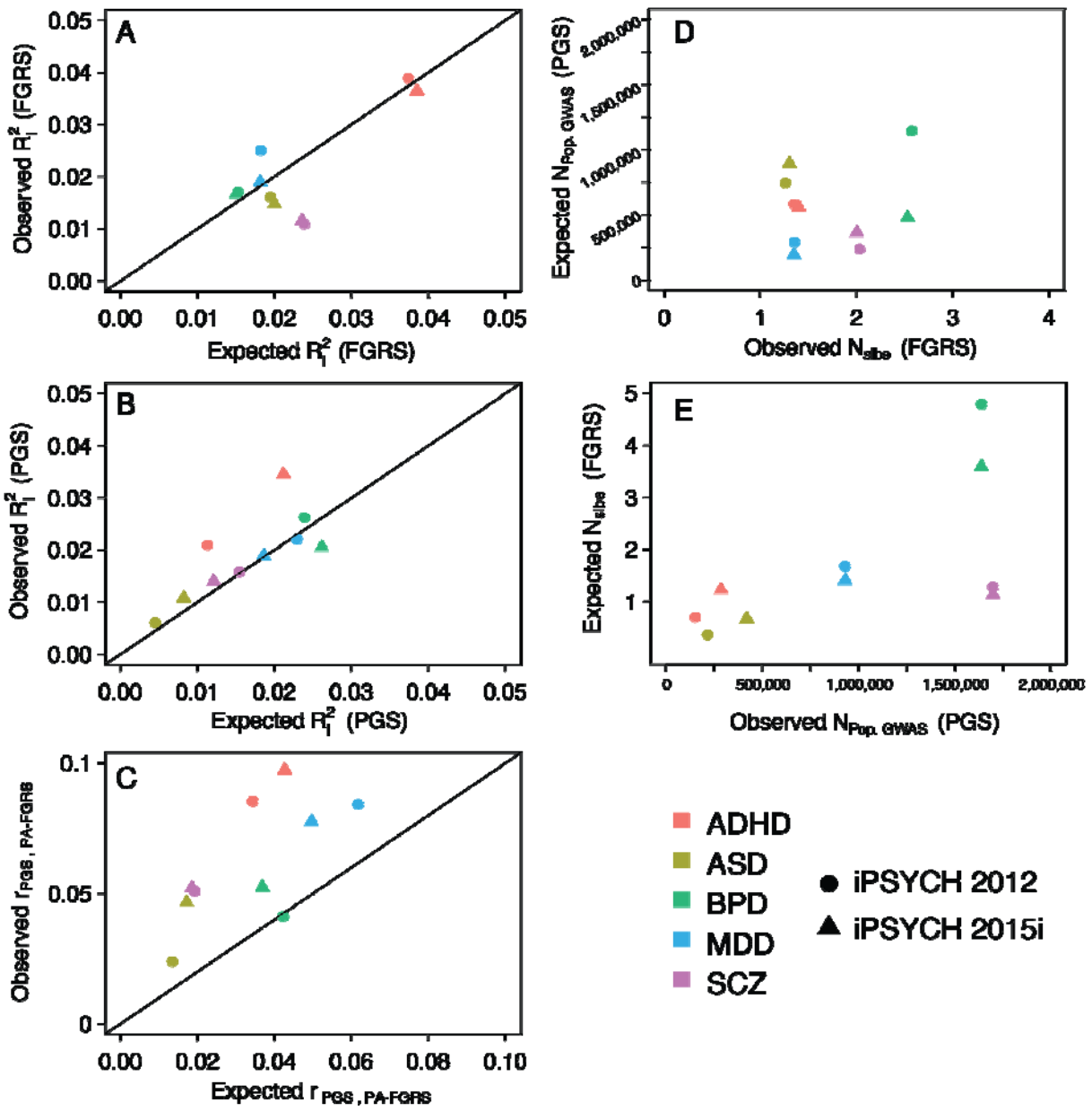


medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 5. PA-FGRS are expected to complement PGS at current sample sizes and in the near future

The expected correlation between PGS and PA-FGRS (A-D) and the expected joint performance of PGS and PA-FGRS (E-H) varies as a function of ^{2}, prevalence, and training GWAS sample size. The expected correlation between the two instruments (A-D) will remain modest, even once PGS approach maximum performance, especially for rare traits (A,B). These modest correlations imply the expected joint performance of PGS and PA-FGRS (E-H, dashed lines) is higher than PGS alone (E-H, solid lines). This effect is larger when PGS training GWAS are smaller and when traits are more common, where higher expected correlations are offset by the higher performance of PA-FGRS (EF vs. GH). Here, we assume that the trait ^{2} in the test sample is equal to ^{2} in the training GWAS, the between the test and training sample is one, the number of markers in the PGS is m=60,000, and the effective proportion of causal markers is p=1. PA-FGRS, Pearson-Aitken Family Genetic Risk Scores; PGS, polygenic score; ^{2}, SNP heritability; , genetic correlation; prev, lifetime prevalence; c.c., case-control sampling; pop., population sampling.

medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.06.19.23291606; this version posted June 20, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 6. Observed performance of genetic instruments is generally in alignment with expectations assuming simple polygenic liability models. (A) The observed performance of PA-FGRS is broadly consistent with expectations from theory under a simple additive polygenic model, given the of the observed pedigrees, ^2_l, prev., and modeling choices for handling censoring and generational cohort effects. (B) The observed performance of PGS is likewise broadly consistent with expectations from theory, given estimated parameters for ^{2} in the test sample, ^{2} in the training GWAS, between the test and training sample, and assumed parameters for prev., the number of independent factors comprising the PGS (m=60,000), and the effective proportion of causal markers (p=1). (C) The observed correlations between PGS and PA-FGRS are modest, but in fact larger than predicted by a simple additive polygenic liability model with chosen ^{2}. (D) The empirical performance of the 10 PA-FGRS (with ≈ 1.3 to 1.77) is estimated to equal the performance of a PGS trained in very large samples from the same population (N ≈ 190,000 to 1,140,000; assuming ^{2} test = ^{2} train, test-trait = 1). (E) The empirical performance of the 10 PGS (training GWAS with ≈ 150,000 to 2,100,000) is expected to equal the performance of PA-FGRS with ≈ 0.3 to 4.7. PA-FGRS, Pearson-Aitken family genetic risk scores; PGS, polygenic score; Liab., liability; Sib., sibling; Equiv., Equivalence; ADHD, Attention deficit hyperactivity disorder; ASD, autism spectrum disorder; BPD, bipolar disorder; MDD, major depressive disorder; SCZ, schizophrenia; ^2_l, narrow-sense heritability; ^{2}, SNP heritability; prev., lifetime prevalence; , genetic correlation; $_{sibe}$, Number of siblings equivalent to pedigree; , equivalent N assuming a population sampling GWAS.