scientific reports

OPEN

MATTERS ARISING

Genome-wide association studies of polygenic risk score-derived phenotypes may lead to inflated false positive rates

Emil Uffelmann [1]✉, Danielle Posthuma [1,2] & Wouter J. Peyrot[1,3]

Gouveia and colleagues (2022)[1] conducted a genome-wide association study (GWAS) of a polygenic risk score (PRS)-derived phenotype (N = 37,784), in which they identified 246 independent loci and 473 lead SNPs. This is an enormous increase compared to the most recent and largest GWAS of AD[2] (N = 1,126,563), which identified 38 loci. Here we show that the applied approach by Gouveia and colleagues may lead to an inflated false positive rate.

In this approach, beta-estimates from a recent GWAS of Alzheimer's disease (AD)[3] were used to construct PRSs in the European UK Biobank[4] sample, using pruning and thresholding[5] with a p-value threshold of 5%. Next, a new case–control phenotype was constructed based on the bottom and top 5% of the PRS distribution, removing 90% of their initial sample. Lastly, a GWAS was conducted on this new PRS-derived phenotype. The authors reasoned that by enriching the sample for individuals with known AD-associated variants, you may also enrich for unknown AD-associated variants. Our major concern is that the applied approach used the same single-nucleotide polymorphisms (SNPs) to construct, as well as to predict the phenotype. In other words, the phenotype was partly regressed on itself, which can inflate test statistics.

We performed simulations roughly emulating the approach (see Methods). In short, we simulated individual phenotypes under a liability threshold model and genotypes that loosely reflect the genetic architecture of AD[2,3,6] (excluding the APOE locus) including 170,000 independent SNPs of which 1200 were causal and 168,800 were non-causal (null-SNPs). We then simulated a discovery sample such that the PRS explains approximately 5% of the phenotypic variance on the liability scale (N = 366,771). We ran a GWAS of AD in this discovery sample and used the estimated betas to construct a PRS in a target sample (N = 300,000). We then selected individuals in the top and bottom 5% of the PRS distribution (N = 30,000) and ran a second GWAS on this new PRS-derived case–control phenotype. The target cohort overlapped to varying degrees with the discovery cohort (i.e. 0%, 50%, and 100%), noting the AD GWAS summary statistics used by Gouveia and colleagues (2022)[1] also contained the UK Biobank.

Our results show highly inflated false positive rates in the GWAS of the PRS-derived phenotype (see Fig. 1 and Supplementary Table). Across all null-SNPs and when there is no overlap between discovery and target cohort, the false positive rate was 0.0024 (s.e.m. = 1×10^{-5}), which constitutes a 48,000-fold increase compared to a well-controlled false positive rate of 5×10^{-8} (see Supplementary Fig. 1 for α = 0.05). This inflation is driven by null-SNPs that were used to construct the PRS-derived phenotype. The false positive rate of these null-SNPs was equal to 0.05 (s.e.m. = 0.0002, a 1×10^{6}-fold increase) when there was no overlap, while null-SNPs which were not used to construct the PRS-derived phenotype did not show any inflation. We also looked at the number of false positive associations per study (i.e. false positive rate times the number of null-SNPs considered), which was 402 on average when there was no overlap and was fully driven by SNPs used to construct the PRS-derived phenotype. Decreasing the significance threshold does not protect from inflation in false positive rates. At a significance threshold of 1×10^{-15}, we observe a mean false positive rate of 9.48×10^{-6} (s.e.m. = 8.7×10^{-7}), a 9.5×10^{9}-fold increase.

Overlap between the discovery and target cohort exacerbated false positive rate inflation, increasing the false positive rate to 0.004 (s.e.m. = 1.5×10^{-5}) across all null-SNPs when there was complete overlap. Similarly, the

1Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. 2Department of Child and Adolescent Psychiatry and Pediatric Psychology, Section Complex Trait Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam University Medical Center, Amsterdam, The Netherlands. 3Department of Psychiatry, Amsterdam UMC, Amsterdam, The Netherlands. ✉email: e.uffelmann@vu.nl
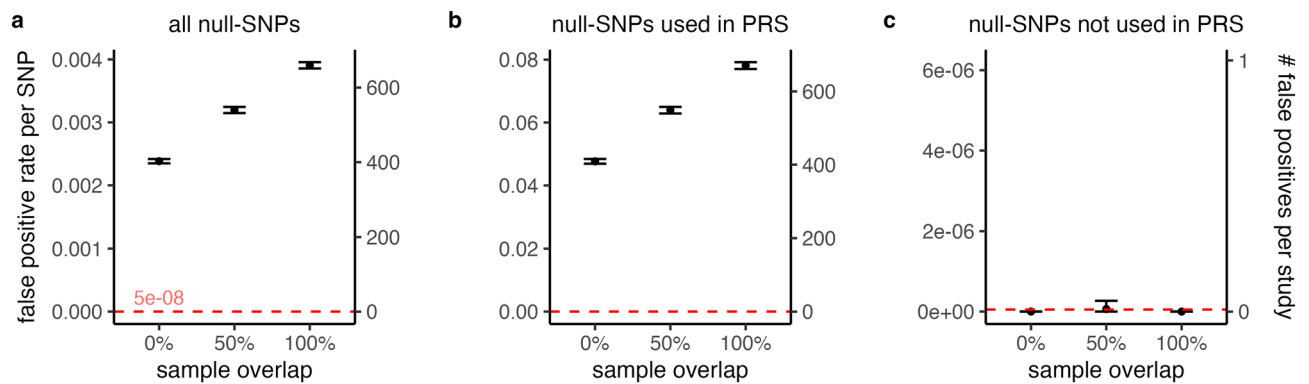
Figure 1. Inflated false positive rates in GWAS of PRS-derived phenotype. The false positive rates **(a–c)** are displayed for varying degrees of sample overlap between discovery and target cohort in a GWAS of a PRS-derived phenotype. Across 100 simulation runs, we observe highly inflated false positive rates. For all null-SNPs, the mean false positive rate ranges between 0.0024 (0% overlap) and 0.0039 (100% overlap) at a significance threshold of 5×10^{-8} **(a)**. Null-SNPs used to construct the PRS-derived phenotype show the highest inflation **(b)**, while all other null-SNPs do not show any inflation **(c)**. As such, the inflation in all null-SNPs is driven by SNPs used to construct the PRS-derived phenotype. Increasing overlap between the target and discovery cohort exacerbates the inflation. We additionally plot the number of false positive associations per study (i.e. false positive rate per SNP times the number of SNPs considered). The mean number of false positives ranged between 402 and 659, and is driven by null-SNPs used to construct the PRS-derived phenotype. The error-bars show the 99.9%-confidence interval of the mean. See Supplementary Table for descriptive statistics.

number of false positive associations increased to 659 (s.e.m. = 2.6). Interestingly, overlap between the discovery and target cohort inflated the false positive rate for null-SNPs used to construct the PRS-derived phenotype but deflated it for all other null-SNPs (see Supplementary Fig. 1). The reason for this is that p-values for null-SNPs will be correlated between the GWAS for AD and the GWAS for the PRS-derived phenotype when there is sample overlap (because AD and the PRS-derived phenotype are correlated and the same individuals are used). Selecting SNPs with p-values smaller than 0.05 for the PRS similarly selects SNPs not part of the PRS with p-values larger than 0.05. As a consequence, the GWAS of the PRS-derived phenotype will have deflated test statistics at null-SNPs not included in the PRS.

Next, we varied the p-value threshold for inclusion in the PRS (i.e. varying the threshold from 0.05 to 1 and 5×10^{-8}, thus including either all SNPs or only genome-wide significant SNPs, respectively). We found that using all SNPs in constructing the PRS-derived phenotype reduced the inflation of false positive rates (as well as the number of false positives, see Supplementary Fig. 2). This reduction is observed because the bias is diluted across all null-SNPs and so the mean false positive rate decreases. Reducing the p-value threshold to 5×10^{-8} resulted in false positive rates that are not inflated. This is because almost no null-SNP had such a low p-value for AD, and thus almost no null-SNPs were used to construct the PRS-derived phenotype.

Lastly, we evaluated a potential power gain for causal SNPs that were not included in the PRS. We calculated the difference in test statistics between the two GWAS (i.e. $Z_{\text{PRS-derived phenotype}} - Z_{\text{AD}}$) and found a strong power decrease (mean difference = -0.14, $p < 2.2 * 10^{-16}$) in the GWAS of the PRS-derived phenotype. This can be explained by the reduction in sample size and only a partial phenotypic correlation between AD and the PRS-derived phenotype. Thus, an increase in power can only be observed for causal SNPs included in the PRS. But because it is not known which SNPs are causal, true associations cannot be distinguished reliably from false positives.

To summarize, Gouveia and colleagues (2022)[1] used a new study design with the aim to improve the power for a GWAS of Alzheimer's disease. Based on simulations, we showed that this approach may lead to inflated false positive rates of 80,000-fold increases at a genome-wide significance threshold of 5×10^{-8}. The reason for this is that the same SNPs used to construct the PRS-derived phenotype were subsequently tested for association with this newly constructed phenotype. We found the false positive rate inflation was more pronounced in the case of sample overlap between the discovery and target cohort. Our results show that false positive rates are not inflated when the GWAS of the PRS-derived phenotype is performed on SNPs that were not also used to construct the PRS. However, we note that when there is linkage disequilibrium between SNPs included in the PRS and null-SNPs not included in the PRS this could still result in an inflated false positive rate. An appealing approach may be to use a leave-one-chromosome-out approach, where the PRS is constructed using 21 chromosomes, and the GWAS of the PRS-derived phenotype only uses the 22nd left-out chromosome (repeated 22 times so that all chromosomes are left out once). However, in our simulations we found a power decrease for causal SNPs that were not included in the PRS. Moreover, we note SNPs can also be correlated across chromosomes due to e.g. non-random mating[7] which could in theory also lead to inflated false positive rates for this approach, but we are not certain about the extent of this inflation which could well be negligible. See the Supplementary Note for a short discussion of some other approaches analyzing (partly) PRS-derived phenotypes, including an approach to improve power[8,9].

To conclude, phenotype definitions based on PRSs require careful consideration in subsequent GWAS. While excluding any SNP (and those in linkage disequilibrium) from the GWAS that was used to construct the PRS-derived phenotype prevents inflation of false positive rates, it also leads to a loss of power for causal SNPs.

Methods

Simulation. We simulated individual genotype and phenotype data based on the liability threshold model. Our chosen parameters were loosely based on Alzheimer's disease[2,3,6], with a population and sample prevalence of 5%, SNP-heritability (h^2_{SNP}) of 10% on the liability scale, and a PRS that explains 5% of the variance (R^2) on the liability scale. We simulated a total of 170,000 SNPs in linkage equilibrium with a minimum minor allele frequency of 0.1%, as this was the number of pruned SNPs used by Gouveia et al. (2022)[1]. Out of these, 1200 SNPs were causal, as previously estimated for Alzheimer's disease[6], and 168,800 were non-causal. We used the avengeme R package to calculate the number of individuals required for the discovery cohort to produce a PRS that explains the desired R^2 value on the liability scale[10]. We simulated individuals and their liabilities, such that individuals with liabilities larger than the liability-threshold are designated cases, and otherwise controls. We repeatedly simulated individuals until we reached the desired number of individuals (N = 366,771 discovery, N = 300,000 target). We repeated the simulation for three target cohorts. That is, within the same simulation run, one target cohort was fully independent of the discovery cohort (0% sample overlap), in the other 50% (and 100%) of individuals were also present in the discovery cohort. Next, we ran a GWAS in the discovery cohort using plink version 1.9[11]. Using the estimated betas, we calculated PRS in the target cohorts to determine the top and bottom 5% of the PRS distribution to define the PRS extremes (i.e. the PRS-derived phenotype), and thus removed 90% of the sample. Lastly, we ran a second GWAS of the PRS-derived phenotype (N = 30,000) and recorded the false positive rate and the variance of test statistics. We repeated the simulation 100 times. We performed several model checks to ensure our simulations have the desired characteristics; specifically, we verified that the false positive rate and test statistics are not inflated for the primary GWAS of Alzheimer's disease.

Data availability

All code used for this manuscript is available at https://doi.org/10.5281/zenodo.7501520. Simulation results can be downloaded from https://doi.org/10.5281/zenodo.7330490.

References

1. Gouveia, C. *et al.* Genome-wide association of polygenic risk extremes for Alzheimer's disease in the UK Biobank. *Sci. Rep.* **12**, 8404 (2022).
2. Wightman, D. P. *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).
3. Jansen, I. E. *et al.* Genome-wide association meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
4. Sudlow, C. *et al.* UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS Med.* **12**, e1001779 (2015).
5. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468 (2015).
6. Holland, D. *et al.* Beyond SNP heritability: Polygenicity and discoverability of phenotypes estimated with a univariate Gaussian mixture model. *PLOS Genet.* **16**, e1008612 (2020).
7. Yengo, L. *et al.* Imprint of assortative mating on the human genome. *Nat. Hum. Behav.* **2**, 948–954 (2018).
8. the Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. A polygenic resilience score moderates the genetic risk for schizophrenia. *Mol. Psychiatry* https://doi.org/10.1038/s41380-019-0463-8 (2019).
9. Zaitlen, N. *et al.* Analysis of case-control association studies with known risk variants. *Bioinform. Oxf. Engl.* **28**, 1729–1737 (2012).
10. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLOS Genet.* **9**, e1003348 (2013).
11. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

Acknowledgements

Author contributions

E.U. conceived the project, wrote the manuscript text and prepared the figures. W.J.P. and E.U. wrote the analysis code. W.J.P. and D.P. supervised the project. All authors discussed and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-29428-9.

Correspondence and requests for materials should be addressed to E.U.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.