Validating Orchid's Multiple Sclerosis Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Multiple sclerosis (MS) is a disease of the central nervous system. It is often characterized by temporary vision loss, weakness, and sensory/autonomic dysfunction. The initial symptoms typically manifest in patients between 20 and 40 years old as reversible episodes of neurological dysfunction lasting several days or weeks. Irreversible damage occurs as the disease progresses. MS-related complications account for over half of deaths among patients with MS, with patients living 7-14 fewer years than average. Common risk factors include low vitamin D, smoking, and obesity.[1]

MS affects over 2 million people worldwide and approximately 0.3% of adults in the US.[2] While there are a number of disease-modifying treatments that may slow down the development of MS, there is no cure. However, there are a number of environmental risk factors that one can minimize, most notably: smoking, adolescent obesity, and vitamin D deficiency. Some of these risk factors have shown gene-environment interactions, suggesting that these interventions may be more effective if one has a high genetic risk.[1]

Genetic Risk Score

MS is shaped by both environmental and genetic factors. Monogenic testing is not available because no single gene causes the condition. Genetic risk scores (GRS), which combine the small effects of many variants into a single score, are currently the only way to estimate genetic risk.[3] Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease.

Risk predictions are adjusted to each individual's ancestry, with predictive power decaying as genetic distance from the predominantly European training data increases.[4] Orchid considers a GRS meaningfully predictive if individuals at approximately the 97.7th percentile have an odds ratio (OR) of at least 2. The MS GRS meets this criterion for all common ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's MS GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.[5] We restricted the analysis to participants of British ancestry and defined MS using the G35 ICD-10 code, yielding 1,729 cases and 406,791 controls (0.4% prevalence). We then grouped individuals by GRS percentile and compared the observed disease prevalence within each group to our model's predictions (Figure 1). For additional technical details, see the Supplementary Information.

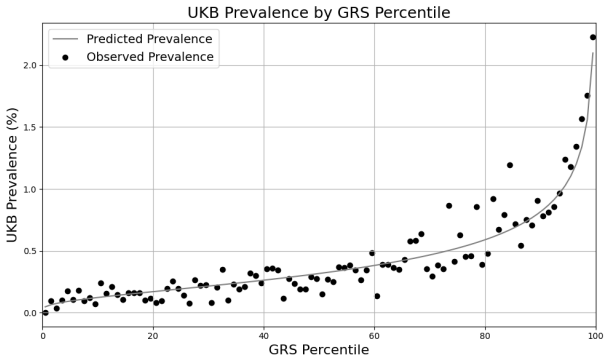

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

Table 1 shows the MS observed prevalence for individuals in the UKB grouped by GRS percentile range (top 10%, 5%, and 1%), as well as how their risk compares to the baseline risk at the 50th GRS percentile. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of MS, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	0.21%	1.00
Top 10%	1.36%	6.46
Top 5%	1.69%	8.06
Top 1%	2.23%	10.70

Table 1. Observed prevalence of MS in the UKB by GRS percentile range. Those with higher GRS relative to the population baseline also had substantially higher observed prevalence of MS.

Estimating Lifetime Risk

Wallin et al. estimate a 0.3% prevalence of MS in the US,[2] similar to the computed 0.4% prevalence in the UKB. We adjust our model so that its average prevalence aligns with the Wallin et al. estimate (Figure 2).[3] People at the high end of the GRS distribution are predicted to have an elevated lifetime risk of the disease relative to the population (Table 2).

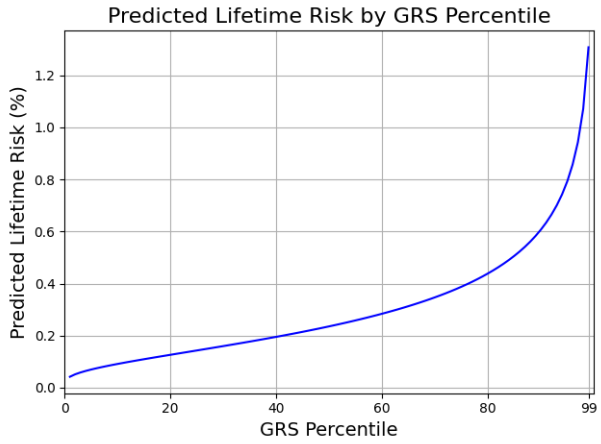

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 0.3% estimate.[2]

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	0.24%	1.00x
95th	0.79%	3.37x
97th	0.94%	4.00x
99th	1.31%	5.55x

Table 2. Predicted lifetime prevalence of MS at different GRS percentiles. Individuals with the highest GRS percentiles are predicted to have an increased risk of MS relative to those at the 50th percentile.

Conclusion

In this study, we evaluated our MS GRS on data from the UKB. We found that it performed well, particularly for identifying individuals with elevated risk of the disease relative to the population. In our embryo and couple reports, we adjust the model to predict risk consistent with the estimated prevalence in the US general population. The MS GRS model is available to individuals of all ancestry groups.

Acknowledgments

This research was conducted using the UK Biobank Resource under Application Number 80545.

References

1. M. Filippi, A. Bar-Or, F. Piehl, et al. Multiple sclerosis. *Nat Rev Dis Primers*, 4:43, 2018. doi:10.1038/s41572-018-0041-4.
2. M. T. Wallin, W. J. Culpepper, J. D. Campbell, et al. The prevalence of MS in the United States: A population-based estimate using health claims data. *Neurology*, 92(10):e1029–e1040, 2019. doi:10.1212/WNL.0000000000007035.
3. N. Chatterjee, J. Shi, M. García-Closas, et al. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*, 17:392–406, 2016. doi:10.1038/nrg.2016.27.
4. Florian Privé, Hugues Aschard, Shai Carmi, et al. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *American Journal of Human Genetics*, 109(1):12–23, 2022. doi:10.1016/j.ajhg.2021.11.008.
5. C. Sudlow, J. Gallacher, N. Allen, 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 Medicine*, 12(3):e1001779, 2015. doi:10.1371/journal.pmed.1001779.

Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
0.24%	2.10	4.40

Table 3. OR per SD. The baseline risk for an individual with a median GRS, and the predicted OR at one and two SDs, respectively. A GRS must have a predicted OR >2 at 2 SD to be included in Orchid's clinical reports.

UKB Prevalence	Population Prevalence	Liability R^2
0.4%	0.3%	7.84%

Table 4. Liability R^2. The estimated liability R^2 using a population prevalence of 0.3%.

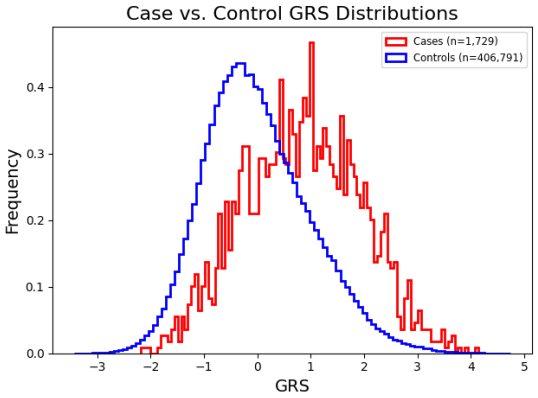

Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.

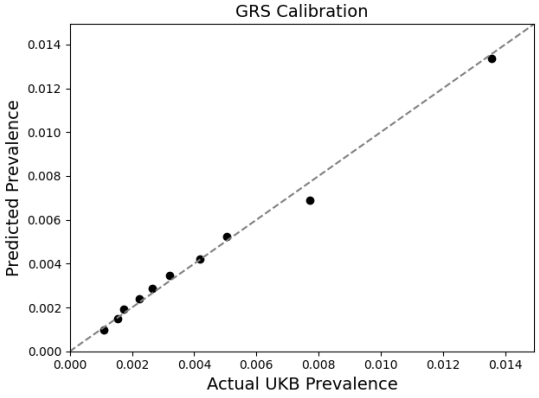

Figure 5. Calibration Curve. Calibration plot showing observed disease prevalence versus predicted risk across GRS deciles.

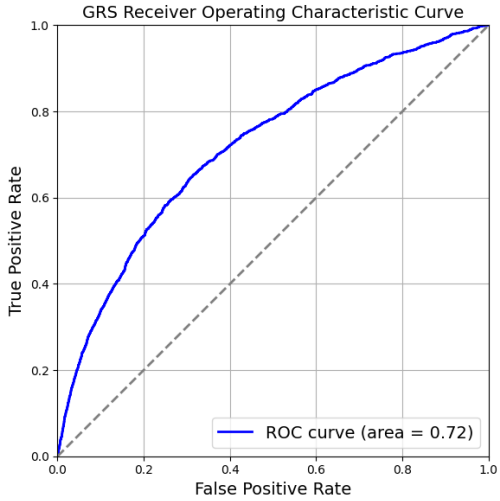

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds. A curve closer to the top-left indicates a better model, while a diagonal line (AUC = 0.5) represents random guessing.