Validating Orchid's Inflammatory Bowel Disease Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Inflammatory bowel disease (IBD) is a term for two conditions, Crohn's disease and ulcerative colitis, that are characterized by chronic inflammation of the gastrointestinal tract. It can cause chronic abdominal pain, diarrhea, ulcers, fatigue, weight loss, and anemia. Crohn's disease can also cause perianal lesions. Surprisingly, smoking is associated with an increased risk of Crohn's disease but a reduced risk of ulcerative colitis. Diet may also play a role in the development and management of both diseases.[1] IBD affects approximately 1.3% of US adults per a 2015 survey.[2] Current treatments emphasize mucosal healing through the reduction of gut inflammation, with biologics associated with increased remission and mucosal healing rates in moderate to severe cases.[1]

Genetic Risk Score

IBD 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. Although not diagnostic, a GRS can indicate how likely an individual is to develop the disease.

Orchid's IBD GRS was trained following current industry standards.[3,4] The GRS was constructed using the SBayesRC algorithm trained on publicly available FinnGen and Million Veterans Program summary statistics.[5,6] The summary statistics include 20,764 cases and 1,104,431 controls.[7] The resulting GRS contains over a million variants.

Risk predictions are adjusted to each individual's ancestry, with predictive power decaying as genetic distance from the predominately European training data increases.[8] Orchid considers a GRS meaningfully predictive if individuals at roughly the 97.7th percentile have an odds ratio (OR) of at least 2. The IBD GRS meets this criterion for all common ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's IBD GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.[9] We restricted the analysis to participants of British ancestry and defined IBD using the K50.x (Crohn's disease) and K51.x (ulcerative colitis) ICD-10 codes, yielding 5,987 cases and 402,533 controls (1.5% 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.

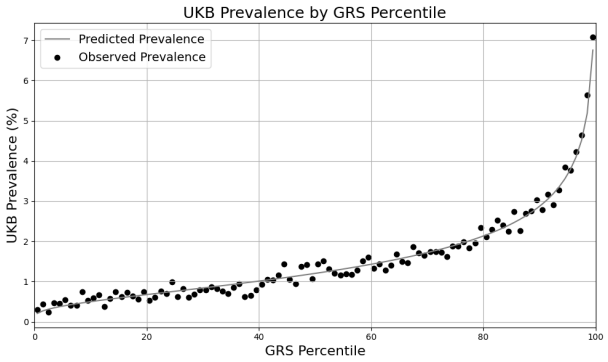

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

Table 1 shows the IBD 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 IBD, supporting the predictive accuracy of the GRS to identify individuals with elevated risk.

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	1.25%	1.00
Top 10%	4.04%	3.32
Top 5%	4.99%	4.13
Top 1%	7.08%	6.01

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

Estimating Lifetime Risk

Xu et al. estimate a 1.3% prevalence of IBD,[2] similar to the computed 1.5% prevalence in the UKB. We adjust our model so that its average prevalence aligns with the Xu et al. estimate (Figure 2).[10] 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).

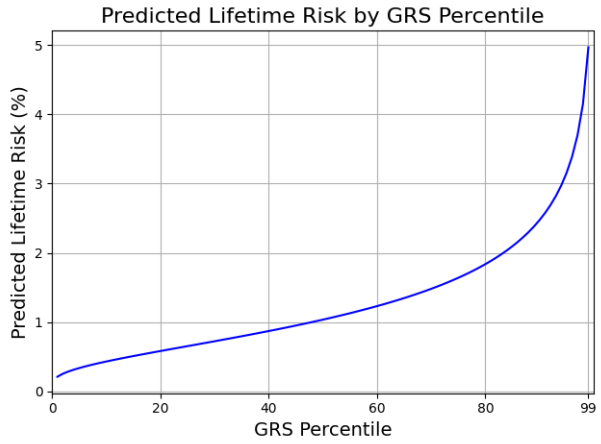

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 1.3% estimate.[2]

GRS Percentile	Predicted Lifetime Risk	Relative Risk
50th (baseline)	1.03%	1.00x
95th	3.16%	3.06x
97th	3.70%	3.58x
99th	4.97%	4.81x

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

Conclusion

In this study, we evaluated our IBD 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 IBD 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. G. Roda, S. C. Ng, P. G. Kotze, et al. Crohn's disease. *Nat Rev Dis Primers*, 6(1):22, 2020. doi:10.1038/s41572-020-0156-2.
2. F. Xu, J. M. Dahlhamer, E. P. Zammitti, et al. Health-Risk Behaviors and Chronic Conditions Among Adults with Inflammatory Bowel Disease — United States, 2015 and 2016. *MMWR Morb Mortal Wkly Rep*, 67(6):190–195, 2018. doi:10.15585/mmwr.mm6706a4.
3. S. Moore, I. Davidson, J. Anomaly, et al. Development and validation of polygenic scores for within-family prediction of disease risks. *medRxiv*, 2025. doi:10.1101/2025.08.06.25333145.
4. S. Cordogan, D. B. Starr, N. R. Treff, et al. Within- and between-family validation of nine polygenic risk scores developed in 1.5 million individuals: implications for IVF, embryo selection, and reduction in lifetime disease risk. *medRxiv*, 2025. doi:10.1101/2025.10.24.25338613.
5. Z. Zheng, S. Liu, J. Sidorenko, et al. Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nat Genet*, 56:767–777, 2024. doi:10.1038/s41588-024-01704-y.
6. FinnGen. FinnGen+MVP+UKBB Summary Statistics. https://mvp-ukbb.finngen.fi/about, 2025. Accessed 2025-12-05.
7. FinnGen. FinnGen+MVP+UKBB Phenotypes. https://mvp-ukbb.finngen.fi, 2025. Accessed 2025-12-15.
8. Florian Privé 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.
9. 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.
10. 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.

Supplementary Information

Baseline Risk	OR per SD	OR per 2 SD
1.03%	2.00	4.00

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
1.5%	1.3%	6.39%

Table 4. Liability R^2. The estimated liability R^2 using a population prevalence of 1.3%.

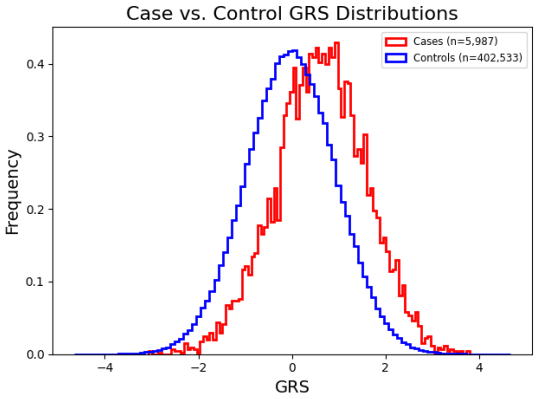

Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.

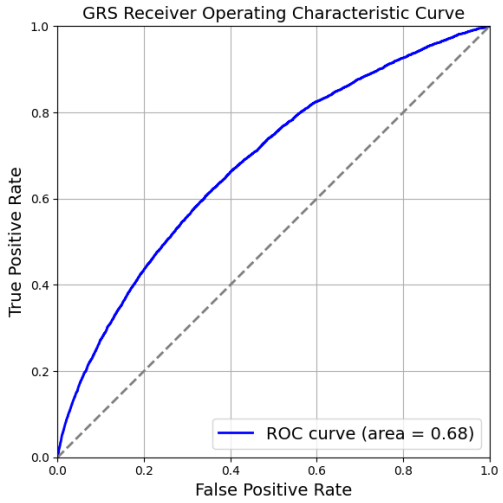

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds.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

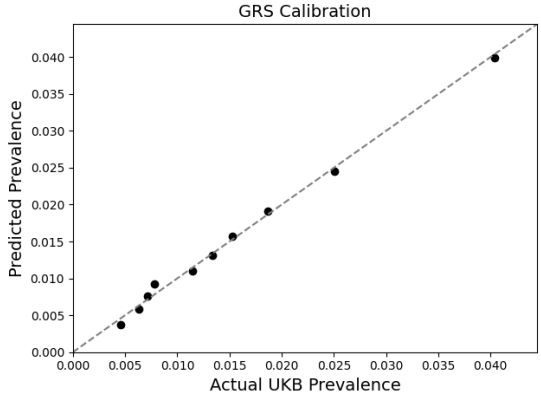

Figure 5. Calibration Curve. Calibration plot showing observed disease prevalence versus predicted risk across GRS deciles.