Validating Orchid's Childhood Asthma Genetic Risk Score

ORCHID

Written by Orchid Bioinformatics Staff

Introduction

Asthma is a common disease of the lungs characterized by inflammation and structural changes to the airways that lead to wheezing, reduced lung function, and heightened sensitivity to allergens. It can develop in either childhood or adulthood, but childhood-onset asthma is more strongly influenced by genetic factors.[1] In addition to genetics, risk factors include prenatal stress and preterm birth, as well as exposure to tobacco smoke both before and after birth.[2]

Asthma affects over 300 million people worldwide and is far more prevalent in high-GDP countries, with prevalence ranging from under 1% in low-GDP regions to around 10% in high-GDP ones. It also much more common near urban centers.[2] In the United States, approximately 6.5% of children under 18 are affected.[3] Treatments typically include inhaled corticosteroids for long-term control and short-acting bronchodilators for fast-acting symptom relief. For children, there is a greater focus on growth monitoring and removal of triggers.[2]

Genetic Risk Score

Childhood asthma 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 childhood asthma GRS was trained following current industry standards.[4,5] The GRS was constructed using the SBayesRC algorithm trained on publicly available FinnGen and Million Veterans Program summary statistics.[6,7] The summary statistics include 113,450 cases and 813,781 controls.[8] 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 predominantly European training data increases.[9] Orchid considers a GRS meaningfully predictive if individuals at roughly the 97.7th percentile have an odds ratio (OR) of at least 2. The childhood asthma GRS meets this criterion for all common ancestry groups.

Evaluation on UK Biobank Data

We evaluated the predictive accuracy of Orchid's childhood asthma GRS using the UK Biobank (UKB), a research database of roughly 500,000 genotyped individuals from the United Kingdom.[10] We restricted the analysis to participants of British ancestry and excluded those who reported developing asthma after age 18. Cases (12,266) were defined as participants who self-reported asthma onset by age 18, and controls (356,923) as participants who self-reported never having asthma. 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.

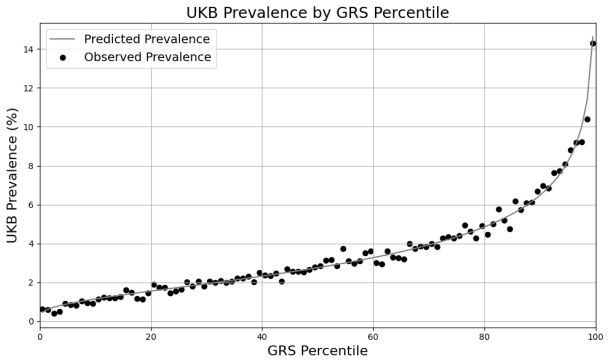

Figure 1. Risk Stratification. Predicted and observed prevalence in the UKB for individuals grouped by GRS percentile.

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

GRS Group	Observed UKB Prevalence	Odds Ratio
Baseline (50th percentile)	2.81%	1.00
Top 10%	8.60%	3.25
Top 5%	10.07%	3.87
Top 1%	14.26%	5.75

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

Estimating Childhood-Onset Risk

The average observed prevalence of childhood asthma in the UKB was 3.3%. This is considerably lower than the prevalence in the US general population, which has been estimated to be approximately 6.5%.[3] This is likely due in part to the fact that UKB participants tend to be healthier than the general population, which leads to lower observed disease prevalence.[11]

Orchid's clinical reports include predicted childhood-onset disease risk for asthma, which we calculate by first estimating how disease risk varies across GRS in the UKB and then rescaling that pattern so the average matches the known population risk (Figure 2).[12] People at the high end of the GRS distribution are predicted to have an elevated risk of the disease relative to the population (Table 2).

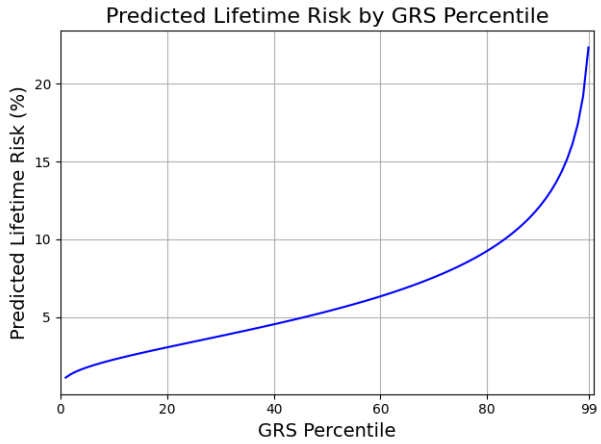

Figure 2. Adjusted Risk Stratification. Predicted risk estimates adjusted so that overall prevalence matches the 6.5% estimate.[3]

GRS Percentile	Predicted Risk	Relative Risk
50th (baseline)	5.35%	1.00x
95th	15.16%	2.83x
97th	17.40%	3.25x
99th	22.34%	4.17x

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

Conclusion

In this study, we evaluated our childhood asthma 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 childhood asthma risk consistent with the estimated US population prevalence, which is generally higher than observed prevalence in the UKB. The childhood asthma 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. Pividori, N. Schoettler, D. L. Nicolae, et al. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma. *Lancet Respir Med*, 7:509–522, 2019. doi:10.1016/S2213-2600(19)30055-4.
2. S. T. Holgate, S. Wenzel, D. S. Postma, et al. Asthma. *Nat Rev Dis Primers*, 1:15025, 2015. doi:10.1038/nrdp.2015.25.
3. Centers for Disease Control and Prevention. Most recent national asthma data. https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm, 2023. May 10, 2023. Accessed December 15, 2025.
4. 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.
5. 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.
6. 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.
7. FinnGen. FinnGen+MVP+UKBB Summary Statistics. https://mvp-ukbb.finngen.fi/about, 2025. Accessed December 5, 2025.
8. FinnGen. FinnGen+MVP+UKBB Phenotypes. https://mvp-ukbb.finngen.fi, 2025. Accessed December 5, 2025.
9. 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.

10. 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.

11. A. Fry, T. J. Littlejohns, C. Sudlow, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol*, 186:1026–1034, 2017. doi:10.1093/aje/kwx246.

12. 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
5.35%	2.01	4.05

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
3.3%	6.5%	8.74%

Table 4. Liability R^2. The estimated liability R^2 using a population prevalence of 6.5%.

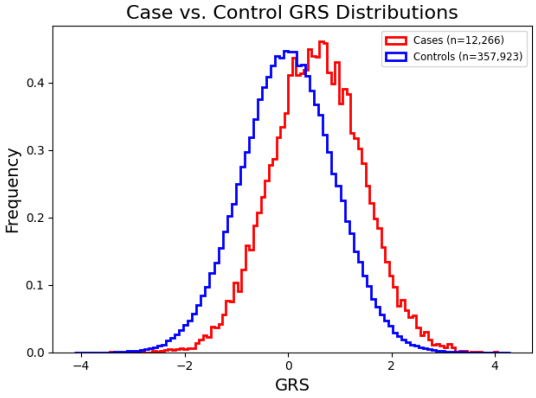

Figure 3. GRS histograms. GRS distributions for cases and controls. Both are approximately normal, with the case distribution shifted noticeably higher compared to the controls.

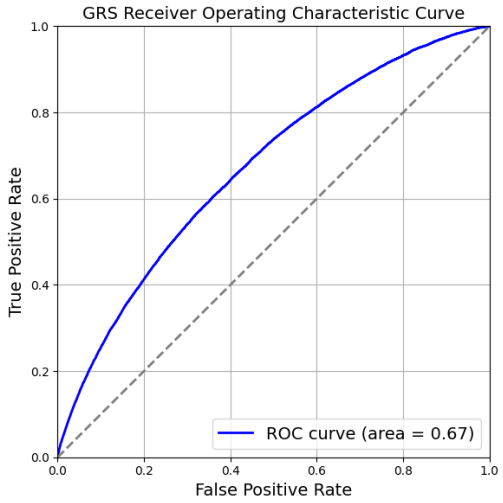

Figure 4. The receiver operating characteristic (ROC) used to compute the ROC area under the curve (AUC). The ROC curve is a graphical representation of a binary classifier's performance, plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) across different decision thresholds.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

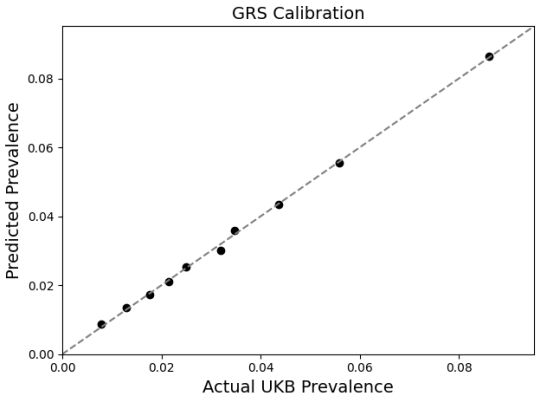

Figure 5. Calibration Curve. Calibration plot showing observed disease prevalence versus predicted risk across GRS deciles.