

1                                    **Differential Results of Polygenic Risk Scoring for Multiple Sclerosis**  
2                                    **in European and African American Populations**

3  
4        Cyprien A. Rivier MD MSc<sup>1,2</sup>, Seyedmehdi Payabvash MD, Hongyu Zhao PhD, David A. Hafler MD<sup>1</sup>,  
5                                    Guido J. Falcone MD, ScD, MPH<sup>1,2\*</sup>, Erin E. Longbrake MD/PhD<sup>1\*</sup>

- 6  
7            1. Department of Neurology, Yale School of Medicine, New Haven, CT, USA.  
8            2. Yale Center for Brain and Mind Health, Yale School of Medicine, New Haven, CT, USA.

9  
10        \*Jointly supervised this work.

11  
12  
13  
14        **Corresponding author:**

15        Guido J. Falcone, MD, ScD, MPH	Erin Longbrake MD/PhD
16        100 York Street, Suite 1-N	6 Devine Street, Ste Suite 2B
17        PO Box 208018	
18        New Haven, CT 06511, USA	North Haven, CT 06473
19        Phone: +1 (203) 785-6288	
20        Fax: +1 (203) 737-4419	

21  
22        Title character count:    15                                    No. of references: 15  
23        Word count abstract:    292                                    No. of tables:    1  
24        Word count article:      1947                                    No. of figures:    2

25

26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

## Abstract

**Importance:** The risk of multiple sclerosis (MS) is significantly influenced by polygenic inheritance. Polygenic risk scores (PRS) for MS can help identify high-risk individuals and stratify populations for clinical trials. However, most genome-wide association studies (GWAS) have been conducted in populations of European ancestry, raising questions about the accuracy of these PRS in other ancestries.

**Objective:** To determine whether a PRS for MS can effectively stratify individuals of non-European ancestries.

**Design, setting and participants:** This cross-sectional study utilized prospectively collected data from the All of Us Research Program (2018-2023). It included participants who had both whole genome sequencing and electronic health record (EHR) data.

**Exposure(s):** A PRS comprising 282 independent single nucleotide variants for MS, divided into quintiles.

**Main Outcome(s) and Measure(s):** Prevalence of multiple sclerosis ascertained through ICD-10 or SNOMED codes.

**Results:** In this study population, MS cases comprised 1.0% (327 cases) of the European population, 0.56% (183 cases) of the African population, and 0.46% (150 cases) of the Latino/admixed American population. In analyses adjusting for age, sex, and genetic principal components, the PRS associates with MS risk in the European population, with a 141% increase in the risk of MS for individuals in the highest compared to the lowest PRS quintile (OR: 2.41 [1.69-3.50], test-for-trend  $p < 0.001$ ). Similarly, the PRS appropriately partitions the Latino/admixed population into increasing MS risk groups (OR: 2.56 [1.45-4.78], test-for-trend  $p$ -value  $< 0.001$ ). However, it did not significantly stratify the African population into distinct MS risk categories (OR: 1.45 [0.95-2.25], test-for-trend  $p = 0.10$ ).

56

57       **Conclusions and Relevance:** A PRS for MS effectively stratified individuals of European and  
58       Latino/admixed ancestries but not African ancestry. This highlights the need for ancestry-specific  
59       PRS development to ensure accurate risk prediction across diverse populations, emphasizing the  
60       importance of including non-European groups in MS genetic research.

61

62

63

## Introduction

64 Multiple sclerosis (MS) is a chronic autoimmune disorder characterized by inflammation and  
65 demyelination in the central nervous system, leading to a range of neurological symptoms and  
66 disabilities<sup>1</sup>. The etiology of MS is complex, involving both genetic and environmental factors<sup>1</sup>.  
67 Advances in genomic technologies have significantly contributed to our understanding of the genetic  
68 architecture of MS, revealing that polygenic inheritance plays a crucial role in the disease's  
69 development<sup>2</sup>. At present, 333 single nucleotide polymorphisms (SNPs) have been associated with  
70 risk of MS<sup>3</sup>, most are common variants that are associated with only modest increases in risk.  
71 However, the cumulative genetic burden of these loci may represent an important biomarker that is  
72 able to differentiate between those with high and low risk of future disease<sup>4</sup>.

73

74 Polygenic risk scores (PRS) have emerged as a powerful tool to quantify an individual's cumulative  
75 genetic predisposition to complex diseases like MS<sup>5</sup>. PRS aggregate the effects of numerous genetic  
76 variants into a single score that can be used to predict disease risk. This method has shown promise  
77 in identifying high-risk individuals. Importantly, clinical trials evaluating lipid-lowering therapies have  
78 shown significant heterogeneity of treatment effects across categories of polygenic risk, with study  
79 participants at highest polygenic risk deriving larger benefits<sup>6</sup>. However, most genome-wide  
80 association studies (GWAS) that inform PRS development have predominantly involved populations  
81 of European ancestry<sup>7</sup>. This raises concerns about the applicability and accuracy of PRS in  
82 individuals of non-European ancestries, potentially leading to disparities in disease prediction and  
83 prevention efforts.

84

85 To address this gap, our study aims to evaluate the effectiveness of an MS-related PRS based on the  
86 known genetic architecture of the disease within different ancestry groups. The All of Us Research  
87 Program<sup>8</sup> aims to enroll an ethnically diverse group of 1 million Americans to accelerate biomedical  
88 research and reduce health disparities. All of Us is generating whole genome sequencing data and  
89 incorporates complete electronic health records (EHR) data. Using All of Us, we sought to determine  
90 whether a PRS developed primarily from European ancestry data could effectively stratify individuals  
91 of African and Latino/admixed American (L/A) ancestries into distinct MS risk categories. Given the  
92 underrepresentation of non-European populations in genetic research, this study provides an

93 essential assessment of PRS utility in diverse groups, thereby contributing to more equitable and  
94 accurate genetic risk assessment in MS.

95

96

## Methods

### 97 **Study design and inclusion criteria**

98 We conducted a cross-sectional genetic study within the All of Us Research Program<sup>8</sup>, an ongoing  
99 study in the US that aims to enroll 1 million Americans aged 18 and older. All of Us does not focus on  
100 any particular disease or health status and emphasizes recruitment of groups that have been  
101 historically underrepresented in biomedical research. All of Us collects baseline health surveys at  
102 enrollment and uses several means to collect longitudinal health data, including continuous  
103 abstraction of EHR data in the form of billing codes, laboratory and medication data, radiology reports,  
104 and linkage with other data sources, such as national death indexes, pharmacy data, health care  
105 claims data, and geospatially linked environmental data. The All of Us data used for this analysis were  
106 collected between May 2018 and July 2023. Data from All of Us (<https://www.researchallofus.org>) is  
107 available to researchers by application. The All of Us institutional review board approved the study  
108 protocol for this study. All participants or their legally designated surrogates provided written informed  
109 consent. In the present nested study, we included participants with both whole genome sequencing  
110 (WGS) data and EHR data.

111

### 112 **Genomic data**

113 DNA was isolated from blood and saliva samples obtained at dedicated research centers across the  
114 US, and WGS data were generated as described previously<sup>9</sup>. WGS data were quality controlled  
115 centrally using standardized pipelines that follow the most up-to-date standards in the field.

116

### 117 **Genetic ancestry ascertainment**

118 Ancestry was ascertained centrally by the All of Us team using principal components analysis on  
119 WGS data by comparing participants with reference datasets from diverse populations. Each  
120 participant was assigned to: African, East Asian, South Asian, West Asian, European, L/A or Other.  
121 For the rest of the analysis, we only included participants belonging to the three largest ancestry  
122 groups: European, African, and L/A.

123

#### 124 **Exposure ascertainment**

125 Our exposure of interest was polygenic susceptibility to MS modeled through a PRS, a well-  
126 established tool in statistical genetics that estimates an individual's genetic burden across numerous  
127 genetic risk variants.<sup>10</sup> For a given study participant, the polygenic risk score is the sum of the product  
128 of the risk allele counts for each variant multiplied by the allele's reported effect on MS. The score  
129 used genetic information on 282 independent genetic risk variants known to be associated with higher  
130 risk of MS. These genetic risk variants are single nucleotide polymorphisms with a minor allele  
131 frequency >1%, independent ( $r^2$ , a measure of the correlation between variants, <0.1), biallelic  
132 (involve two alleles only) and associated with the risk of MS at genome-wide levels ( $P < 5 \times 10^{-8}$ ),  
133 following best practices for PRS generation.<sup>3</sup>

134

#### 135 **Outcome assessment**

136 Corresponding ICD-10, and Systemized Nomenclature of Medicine (SNOMED) codes were used to  
137 ascertain MS. Relative to baseline assessment, both prevalent and incident MS events were counted.

138

#### 139 **Statistical analysis**

140 We present discrete variables as counts (percentage [%]) and continuous variables as mean  
141 (standard deviation [SD]) or median (interquartile range [IQR]), as appropriate. Unadjusted  
142 comparisons were made using chi-square tests for discrete variables and t or ANOVA tests, as  
143 appropriate, for continuous variables. The three genetic ancestries considered (European, African,  
144 L/A) were randomly sampled to obtain three populations of equivalent sizes. In each population, we  
145 normalized the PRS (by subtracting the mean and dividing by the standard deviation) and divided it  
146 into quintiles to establish five risk categories (very low, low, intermediate, high and very high polygenic  
147 risk). To assess the relationship between the PRS and the risk of MS in each population, we used  
148 multivariable logistic regression models adjusting for age, sex, and the first four genetic principal  
149 components.

150

## 151 **Results**

### 152 **Cohort characteristics**

153 Of 413,457 participants included in the All of Us Research Program, 103,125 did not have EHR data,  
154 137,179 did not have WGS data, and 173,153 were included in our analysis (mean age: 51.6 (SD:  
155 17), 60% female). The distribution of genetic ancestry was as follows: 95,971 (55.4%) participants of  
156 European, 37,674 (21.8%) of African, 32,428 (18.7%) of L/A, 4,040 (2.3%) of East Asian, 2,365  
157 (1.4%) of South Asian, and 675 (0.4%) of Middle Eastern genetic ancestry (Table 1). To obtain three  
158 populations of equal sizes, we randomly sampled 32,428 participants from the African and European  
159 populations to match the number of L/A participants.

160

### 161 **Unadjusted analysis**

162 In each sample of 32,428 participants, there were 327 (1.0%) MS cases in the European group, 183  
163 (0.56%) in the African group, and 150 (0.46%) in the L/A group. In the European population, the  
164 proportion of MS cases ranged from 0.66% in the lowest PRS quintile to 1.59% in the highest PRS  
165 quintile (p-value < 0.001 - Figure 1). In the African population, the proportion of MS cases did not  
166 increase consistently across PRS categories with the same strength. The smallest proportion (0.45%)  
167 occurred in the intermediate (3rd) PRS quintile, while the highest (0.82%) was in the highest PRS  
168 quintile (p-value: 0.06). In the L/A population, the proportion of MS cases increased by PRS quintile,  
169 ranging from 0.23% in the lowest quintile to 0.63% in the highest (p-value < 0.001).

170

### 171 **Adjusted analysis**

172 In multivariable logistic regression adjusting for age, sex, and genetic principal components, we found  
173 that the PRS provided effective risk stratification in the European population. The odds of MS  
174 increased by 51% (Odds Ratio: 1.51, 95% CI: [1.03-2.25]) in the intermediate PRS quintile and by  
175 141% (OR: 2.41 [1.69-3.50]) in the highest PRS quintile compared to the lowest (test for trend p-  
176 value: <0.0001 - Figure 2). In the African population, the PRS did not significantly stratify participants  
177 into different MS risk categories (test-for-trend p-value: 0.10). Conversely, in the L/A population, the  
178 PRS effectively stratified risk across several categories. The odds of MS increased by 85% (Odds  
179 Ratio: 1.85, 95% CI: [1.00-3.56]) in the intermediate PRS quintile and by 156% (OR: 2.56 [1.45-4.78])  
180 in the highest quintile compared to the lowest (test-for-trend p-value: 0.0004).

181

182

## **Discussion**

183 In this study, we evaluated the effectiveness of a PRS for MS across different genetic ancestries using  
184 data from the All of Us Research Program. Our findings reveal significant disparities in the utility of the  
185 PRS among European, African, and L/A populations, highlighting the necessity for tailored genetic risk  
186 prediction tools in diverse populations.

187

188 The PRS effectively stratified MS risk in European and L/A individuals, with higher PRS quintiles  
189 corresponding to significantly increased odds of MS. This confirms the robustness of PRS derived  
190 from predominantly European GWAS data for these groups. Specifically, individuals in the highest  
191 PRS quintile had significantly elevated odds of developing MS compared to those in the lowest  
192 quintile, consistent with previous studies demonstrating the potential of PRS in risk stratification and  
193 early disease identification<sup>2,11</sup>.

194

195 Conversely, the PRS did not significantly stratify MS risk among individuals of African ancestry. This  
196 discrepancy may stem from several factors. Primarily, the limited representation of African populations  
197 in existing GWAS likely results in PRS that are less applicable to this group<sup>12</sup>. The genetic variants  
198 contributing to MS risk in Europeans may not have the same effect in African populations, indicating  
199 differences in the genetic architecture of MS across ancestries<sup>13</sup>. Additionally, gene-environment  
200 interactions could play a role; genetic variants might interact with environmental factors differently  
201 across populations, affecting disease risk in ways that a PRS developed in one population cannot  
202 capture<sup>14</sup>. For instance, lifestyle, diet, and exposure to certain environmental factors can vary widely  
203 between populations, influencing how genetic predispositions manifest as disease. Moreover, African  
204 populations possess greater genetic diversity compared to European populations<sup>15,16</sup>. This diversity  
205 means the distribution of genetic variants is broader, including many that may be rare or absent in  
206 European populations. As a result, a PRS developed using European data may not account for the full  
207 range of genetic factors influencing MS risk in African populations. This genetic diversity complicates  
208 the direct application of PRS across different ancestries, necessitating the development of population-  
209 specific PRS to ensure accurate and equitable risk prediction.

210

211 Our study also highlights the importance of utilizing large, diverse datasets like the All of Us Research  
212 Program to evaluate genetic tools' performance across populations. Such datasets are invaluable for



213 understanding complex diseases' genetic basis in diverse groups and developing inclusive genetic  
214 risk prediction models.

215

216 This study has several limitations that warrant consideration. Firstly, the absolute risk of MS in the  
217 population is low, which may limit the practical utility of the PRS as a stand-alone screening tool.

218 Given the multifactorial nature of MS, a comprehensive screening approach that incorporates genetic,  
219 environmental, and lifestyle factors may be necessary to more accurately identify individuals at risk.

220 Additionally, the performance of the PRS in the All of Us dataset was lower compared to its

221 performance in other large datasets, such as the UK Biobank, even when considering individuals of

222 European ancestry only<sup>2</sup>. This discrepancy suggests that specific factors or characteristics unique to

223 the All of Us dataset, beyond genetic ancestry, may influence PRS performance. Factors such as

224 differences in data collection methods, population heterogeneity, or unmeasured environmental

225 influences could contribute to this variation.

226

227 In conclusion, while the PRS for MS is effective in European and L/A populations, it is less effective

228 for African ancestry individuals. This emphasizes the urgent need for inclusive genetic research and

229 the development of ancestry-specific PRS to ensure equitable benefits from genetic risk prediction

230 and personalized healthcare. Future studies should focus on expanding genetic diversity in research

231 cohorts, improving underrepresented populations' representation in GWAS, and refining PRS

232 methodologies to enhance their applicability across all populations, ultimately advancing the goal of

233 personalized medicine for all.

234

#### 235 **Data availability statement**

236 This study used data from the All of Us Research Program's Controlled Tier Dataset C2022Q4R9 and

237 Registered Tier Dataset R2022Q4R9, available to authorized users on the Researcher Workbench via

238 [www.allofus.nih.gov](http://www.allofus.nih.gov). All data access and analyses were conducted within a secure informatic

239 workspace provided by the National Institutes of Health.

240

241

## References

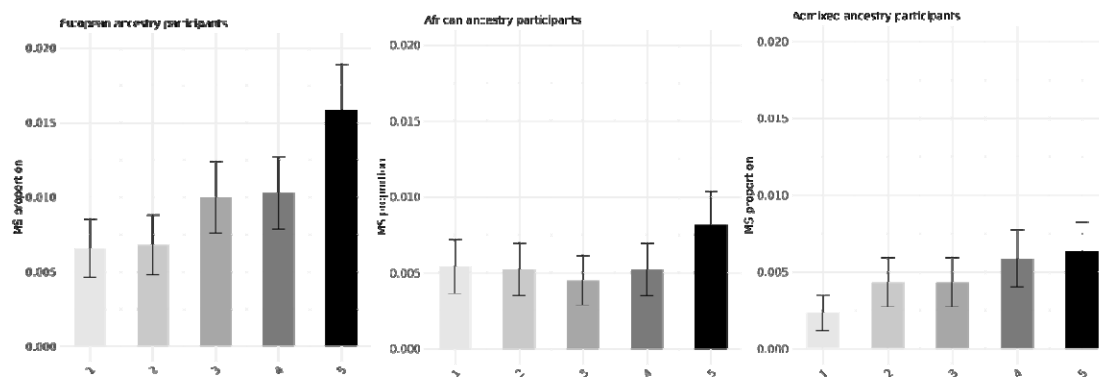
1. Dobson R, Giovannoni G. Multiple sclerosis - a review. *Eur J Neurol.* 2019;26(1):27-40. doi:10.1111/ene.13819
2. Shams H, Shao X, Santaniello A, et al. Polygenic risk score association with multiple sclerosis susceptibility and phenotype in Europeans. *Brain.* 2023;146(2):645-656. doi:10.1093/brain/awac092
3. INTERNATIONAL MULTIPLE SCLEROSIS GENETICS CONSORTIUM. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365(6460):eaav7188. doi:10.1126/science.aav7188
4. Loginovic P, Wang F, Li J, et al. Applying a genetic risk score model to enhance prediction of future multiple sclerosis diagnosis at first presentation with optic neuritis. *Nat Commun.* 2024;15(1):1415. doi:10.1038/s41467-024-44917-9
5. Choi SW, Mak TSH, O'Reilly PF. A guide to performing Polygenic Risk Score analyses. *Nat Protoc.* 2020;15(9):2759-2772. doi:10.1038/s41596-020-0353-1
6. Natarajan P, Young R, Stitzel NO, et al. Polygenic Risk Score Identifies Subgroup With Higher Burden of Atherosclerosis and Greater Relative Benefit From Statin Therapy in the Primary Prevention Setting. *Circulation.* 2017;135(22):2091-2101. doi:10.1161/CIRCULATIONAHA.116.024436
7. Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell.* 2019;179(3):589-603. doi:10.1016/j.cell.2019.08.051
8. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. *N Engl J Med.* 2019;381(7):668-676. doi:10.1056/NEJMs1809937
9. Devaney S. All of Us. *Nature.* 2019;576(7785):S14-S17. doi:10.1038/d41586-019-03717-8
10. Vaura F, Kauko A, Suvila K. Polygenic Risk Scores Predict Hypertension Onset and Cardiovascular Risk. *Hypertens Dallas Tex.* 1979;2021;77(4):1119-1127. doi:10.1161/HYPERTENSIONAHA.120.16471
11. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. *Nat Rev Cardiol.* Published online November 22, 2021:1-11. doi:10.1038/s41569-021-00638-w
12. Ruan Y, Lin YF, Feng YCA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet.* 2022;54(5):573-580. doi:10.1038/s41588-022-01054-7
13. Beecham AH, Amezcua L, Chinea A, et al. The genetic diversity of multiple sclerosis risk among Hispanic and African American populations living in the United States. *Mult Scler.* 2020;26(11):1329-1339. doi:10.1177/1352458519863764
14. Virolainen SJ, VonHandorf A, Viel KCMF, Weirauch MT, Kottyan LC. Gene-environment interactions and their impact on human health. *Genes Immun.* 2023;24(1):1-11. doi:10.1038/s41435-022-00192-6
15. Pereira L, Mutesa L, Tindana P, Ramsay M. African genetic diversity and adaptation inform a precision medicine agenda. *Nat Rev Genet.* 2021;22(5):284-306. doi:10.1038/s41576-020-00306-8
16. Yu N, Chen FC, Ota S, et al. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics.* 2002;161(1):269-274.

### Tables and figures (3 max)

**Table 1. Cohort characteristics**

	All of Us		
	MS (N = 1,364)	No MS (N= 171,789)	P value
Age at enrollment, mean (SD)	52.27 (13.7)	51.58 (17.0)	0.138
Female	1,055 (77.3)	103,141 (60.0)	<b>&lt;0.001</b>
Genetic ancestry			
EUR	971 (71.2)	95,000 (55.3)	<b>&lt;0.001</b>
AFR	214 (15.7)	37,460 (21.8)	
AMR	150 (11.0)	32,278 (18.8)	
EAS	<20	4,030 (2.3)	
SAS	<20	2,352 (1.4)	
MID	<20	669 (0.4)	
Smoking			
Current	176 (13.2)	27,640 (16.6)	<b>&lt;0.001</b>
Previous	403 (30.3)	38,903 (23.3)	
Body Mass Index			
<25	403 (31.0)	47,034 (28.3)	<b>0.047</b>
25-30	397 (30.5)	50,400 (30.3)	
>=30	501 (38.5)	69,008 (41.5)	

**Figure 1. Proportions of MS cases by PRS quintile in European, African, and L/A ancestry populations**



**Figure 2. Adjusted odds ratios of MS diagnosis by PRS quintile in European, African, and L/A ancestry populations**

