Development and Performance of Polygenic Risk Scores for Genomic Reporting

Abstract

BACKGROUND: Polygenic risk scores (PRSs) take into account the effects of multiple common genetic variants enabling prediction of the future risk of the nonfamilial forms of diseases. PRSs have been proposed to be useful for lifestyle optimizations, therapeutic intervention, and disease screening despite ongoing discussions about their clinical utility and predictive performance.

METHODS: We have developed a set of PRSs using the variants identified in current genome-wide association studies (GWASs). The predictive performance of these models was evaluated using participants of the UK Biobank, a cohort study with deep genetic, physical and health data collected on ~500,000 individuals across the UK.

RESULTS: The performance, relative risk and disease prevalence have been provided for each model. The PRS for celiac disease, a condition with high heritability, performed the best (AUC 0.808). PRSs for Alzheimer's disease, multiple sclerosis, chronic lymphocytic leukemia, HDL cholesterol, osteoporosis, and Dupuytren's disease also had good performance (AUC 0.676 – 0.696). The AUC for remaining models was in the range: 0.552 - 0.658.

CONCLUSIONS: The PRSs for 32 common disorders have been implemented in the Simplify Genomics Genomic Clinical Report. The results are not recommended to be used to guide medical care but to rather guide lifestyle decisions.

Authors

Svetlana Bureeva, Lei Huang, Arya Iranmehr, Nafisa Bulsara, Rachana Sainger, Wayne Delport

Note

This work was originally published in March, 2020 by the authors above while they were employees of Human Longevity, Inc (HLI). The clinical genomics division of HLI was divested into Simplify Genomics in January, 2022. Wayne Delport is the senior author and is the Chief Technology Officer at Simplify Genomics. The UK Biobank datasets used in this validation study were enabled by an approved HLI UK Biobank Research Project.

Contents

PRS PERFORMANCE

1 Introduction

In recent years, results of comprehensive genomic studies[1] reinforced the conclusion that the genetic architecture of many common adult onset diseases is composed of:

- a familial form, responsible for 1–10% of disease incidence, linked to highly penetrant rare variants within a small set of genes known to drive familial disease.

- a nonfamilial form of disease that is mostly driven by a combination of common variants of small effect distributed throughout the genome.

Polygenic risk scores have been developed to take into account the effects of multiple common genetic variants into a single score, thus, allowing for the prediction of the future risk of the nonfamilial forms of diseases.

The scores have become increasingly popular in recent years, and studies reporting new PRSs for a variety of diseases appear regularly in the literature. There is a growing understanding that at least for some diseases PRSs could be used in conjunction with clinical, environmental, and monogenic risk factors for the following[1]:

- PRS-informed therapeutic intervention, e.g. stratifying individuals for starting statin therapy for heart disease[2].
- PRS-informed disease screening, e.g. stratifying women for the best age to start mammography for breast cancer. A recent study[3] modelled cost effectiveness and benefit to harm ratio of PRS-guided risk stratified screening for breast cancer, and established that not offering breast cancer screening to women at lower risk could improve the cost effectiveness of the screening program, reduce overdiagnosis, and maintain the benefits of screening. Another study estimated that the performance of the colorectal cancer PRS with AUC of at least 0.65 makes PRS risk stratified screening cost effective as the reduction in screening and treatment costs start covering the expected cost of polygenic testing[4].

- PRS informed life planning, e.g. lifestyle optimization. For coronary artery disease, individuals in the top quintile of genetic risk have the ability to offset much of this risk by maintaining optimal lifestyle habits, which reduces their overall risk of disease by nearly half[5]. For breast cancer, if healthy lifestyle choices were preferentially targeted to and employed by women in the top decile of the PRS-driven genetic risk, an estimated ~20% of all preventable breast cancer cases would be avoided[6].

Despite these applications, the clinical utility of PRSs remains under active discussion in the scientific and medical fields. In the case of coronary artery disease, many caution about PRS implementation in clinical practice. The limited discriminatory value of PRSs when used on their own as compared to other screening tests[7] as well as the modest or no improvement in the prediction accuracy provided by the PRSs' addition to the well established risk prediction models[8,9,10] are the main reasons for the skepticism. Limited translatability of PRSs across populations further decreases the utility of PRS models that are predominantly built based on genomes of people of European ancestry[11].

Despite these ongoing discussions, in recent years, several genetic companies have launched commercial PRS tests to improve disease risk assessment for cancer, heart attack, and a number of other disorders including Ambry Genetics, Myriad Genetics, Color Genomics, and MyHeritage[12,13]. Implementations of PRSs are also available from the direct to consumer providers like 23andMe[14,15] and Nebula Genomics[16].

Ambry Genetics implemented PRSs for prostate cancer[17] and breast cancer[18] that include 72 and 100 variants, and perform with AUC 0.65 and 0.61, respectively. Myriad Genetics developed and validated two breast cancer PRSs for women

of European[19] and Hispanic[20] ancestry including 86 and 87 variants, respectively. 23andMe developed and launched a type 2 diabetes model that includes over 1,000 variants. The model's performance was validated for several populations demonstrating AUC values between 0.65 for Europeans and 0.59 for African Americans[15]. Nebula Genomics has implemented dozens of PRSs for cancer, metabolic, mental, musculoskeletal, and other disorders.

This white paper describes our effort to develop and evaluate performance of PRSs for a number of common disorders. The variants comprising the models were identified from current published genome-wide association studies (GWASs). The predictive performance of the models, relative risk, and disease prevalence were evaluated using affected and nonaffected participants of the UK Biobank study[21] for each disease.

The PRSs are implemented in the Simplify Genomics Clinical Report. The reported results are not recommended to be used to guide medical care but rather to inform lifestyle decisions.

2 Results

The predictive performance of each model accompanied by observed and predicted relative risk and disease prevalence are provided in sections 2.1 – 2.32 and Table A. Cohort definitions are provided in Table B.

BOX 1. How to understand model performance and other metrics in this study.

AUC is the probability that the PRS of a case is greater than the PRS of a control. The values of 1.0 and 0.5 indicate the best and the worst performance, respectively.

PRS Percentile: PRS at 75th percentile means that 75% of the individuals in the population have a lower PRS and 25% have a higher a PRS.

Risk ratio indicates the relative risk of developing a disease for people with certain genetic profiles as compared to individuals in a reference group. In this study, the reference group was defined as individuals in the 40-60th PRS percentile bin as having "most common" genetic profiles.

- For example, a risk ratio of 2.80 means that the individual is at 2.8 fold increased risk for developing a disease as compared to the individuals with "most common" genetic profiles.
- A risk ratio of 0.40 means that the individual is at 60% decreased risk for developing a disease as compared to the individuals with "most common" genetic profiles.

Odds ratio is a metric closely related to risk ratio that indicates the association strength between PRS and disease. It gives a similar value to risk ratio when a disease is rare and relative risk is small or moderate.

Prevalence is the percent of people who developed a condition in a certain PRS percentile bin.

2.1 Basal Cell Carcinoma

Basal cell carcinoma is very common among people of some ethnicities; among Europeans, one in three men and one in four women will develop the condition during their lifetime[22]. The PRS model for basal cell carcinoma includes 29 variants and was developed based on the variants identified in a study that analyzed genomes of over 17,000 individuals of European ancestry affected by basal cell carcinoma[23]. Heritability of basal cell carcinoma is over 40%[24] and about 11%[23] of familial risk of basal cell carcinoma was explained by the variants identified in the study. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The predictive performance of the developed model (AUC 0.606) is slightly below compared to the published PRS models for basal cell carcinoma (AUC - 0.63[25] - 0.64[26]).

Number of individuals	408,885
Number of cases (%)	16,391 (4.0)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.606 (0.601, 0.610)
Covariates	0.666 (0.662, 0.669)
PRS + covariates	0.696 (0.692, 0.700)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

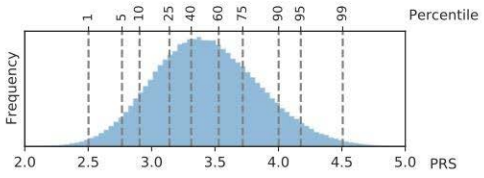

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.54 (0.50, 0.58)	0.54 (0.51, 0.59)
10-25	0.69 (0.65, 0.73)	0.70 (0.66, 0.74)
25-40	0.82 (0.77, 0.87)	0.83 (0.78, 0.87)
40-60	1.00 (0.95, 1.05)	1.00 (0.95, 1.05)
60-75	1.20 (1.14, 1.26)	1.19 (1.13, 1.25)
75-90	1.44 (1.37, 1.52)	1.42 (1.35, 1.49)
90-95	1.73 (1.62, 1.85)	1.68 (1.58, 1.79)
95-99	2.06 (1.92, 2.21)	1.98 (1.86, 2.11)
99-100	2.70 (2.42, 3.01)	2.54 (2.29, 2.80)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

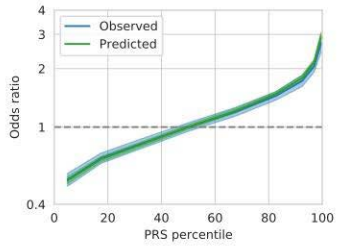

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

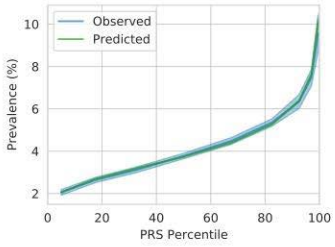

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.2 Bladder Cancer

Bladder cancer is a common disease; one in 27 men and one in 89 women will develop bladder cancer during their lifetime[27]. The PRS model for bladder cancer includes 12 variants and was developed based on the variants identified in a study that analyzed genomes of about 7,000 individuals of European ancestry affected by bladder cancer[28]. Heritability of bladder cancer is currently unknown; however, family history of bladder cancer is associated with about 2-fold increased risk, suggesting shared genetic and potential environmental contribution to its etiology[29].

The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The predictive performance of the developed model is modest (AUC 0.566) possibly attributing to the limited number of discovered variants and a strong contribution of lifestyle - cigarette smoking is thought to cause 50% of bladder cancer cases[30]. To the best of our knowledge, the performance of other bladder cancer PRS models has not been published.

Number of individuals	408,885
Number of cases (%)	2,758 (0.7)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.566 (0.557, 0.576)
Covariates	0.764 (0.758, 0.773)
PRS + covariates	0.770 (0.763, 0.779)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

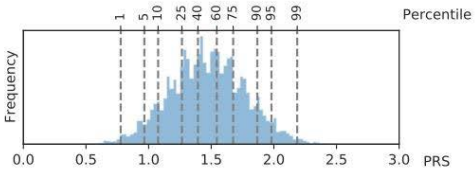

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.69 (0.58, 0.81)	0.69 (0.58, 0.82)
10-25	0.87 (0.76, 1.00)	0.87 (0.76, 1.00)
25-40	0.93 (0.81, 1.06)	0.93 (0.81, 1.06)
40-60	1.00 (0.89, 1.13)	1.00 (0.89, 1.13)
60-75	1.17 (1.03, 1.32)	1.17 (1.03, 1.32)
75-90	1.27 (1.13, 1.44)	1.27 (1.13, 1.44)
90-95	1.52 (1.29, 1.79)	1.51 (1.28, 1.78)
95-99	1.63 (1.37, 1.94)	1.63 (1.37, 1.93)
99-100	2.35 (1.79, 3.07)	2.33 (1.79, 3.04)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

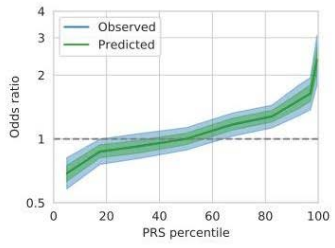

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

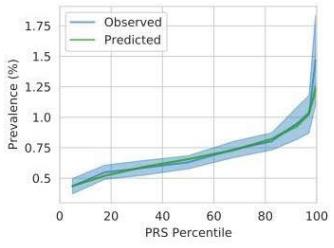

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.3 Breast Cancer

Breast cancer is a very common disease; one in eight women will develop the condition during their lifetime[31]. The PRS model for breast cancer includes 151 variants and was developed based on the variants identified in a study that analyzed genomes of over 120,000 women of European ancestry affected by breast cancer[32]. Heritability of breast cancer is about 30%[24,33] and about 18% of familial risk of breast cancer was explained by the variants discovered[32] . The performance of the developed PRS model was evaluated using affected and non-affected women from UK Biobank and the results are presented in Tables 1-3 and Figures 1-3. The predictive performance of the developed model (AUC 0.617) is comparable to the previously published PRS models for breast cancer (AUC 0.55 – 0.68)[18,25,34,35,36,37,38,39,40].

Number of individuals	221,005
Number of cases (%)	14,609 (6.6)
Median age (25 & 75 percentiles)	66.8 (59.4 & 71.9)

Table 1: Validation cohort characteristics.

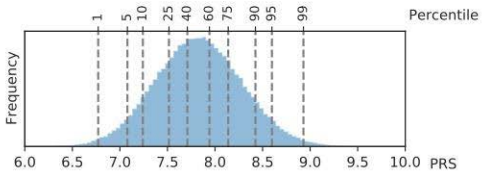

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

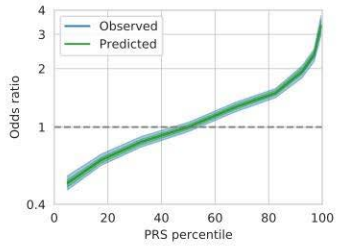

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

Features	AUC (95% CI)
PRS	0.617 (0.613, 0.621)
Covariates	0.600 (0.596, 0.604)
PRS + covariates	0.654 (0.650, 0.657)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.52 (0.47, 0.56)	0.53 (0.49, 0.58)
10-25	0.68 (0.63, 0.72)	0.69 (0.65, 0.74)
25-40	0.84 (0.79, 0.89)	0.85 (0.80, 0.90)
40-60	1.00 (0.95, 1.06)	1.00 (0.95, 1.05)
60-75	1.27 (1.20, 1.34)	1.25 (1.18, 1.31)
75-90	1.49 (1.41, 1.58)	1.45 (1.38, 1.53)
90-95	1.93 (1.79, 2.07)	1.82 (1.71, 1.95)
95-99	2.35 (2.18, 2.53)	2.17 (2.03, 2.32)
99-100	3.33 (2.96, 3.74)	2.92 (2.65, 3.22)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

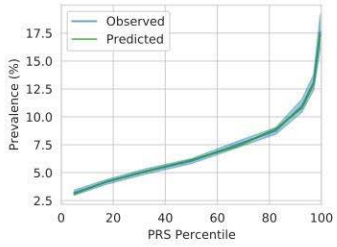

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.4 Colorectal Cancer

Colorectal cancer is a common disease; one in 23 people will develop the condition during their lifetime[41]. The PRS model for colorectal cancer includes 29 variants and was developed based on the variants identified in a study that analyzed genomes of about 37,000 individuals of European ancestry affected by colorectal cancer[42]. Heritability of colorectal cancer is about 35%[43] and about 12% of familial risk of colorectal cancer was explained by the variants discovered[42]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The predictive performance of the developed model (AUC 0.580) is below compared to the previously published PRS models for colorectal cancer 0.62 (women), 0.64 (men)[44].

Number of individuals	408,885
Number of cases (%)	6,016 (1.5)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.580 (0.572, 0.587)
Covariates	0.682 (0.676, 0.687)
PRS + covariates	0.698 (0.691, 0.703)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

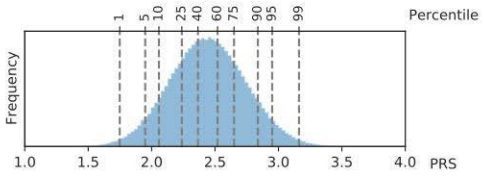

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.59 (0.52, 0.66)	0.59 (0.52, 0.66)
10-25	0.72 (0.65, 0.79)	0.72 (0.66, 0.79)
25-40	0.81 (0.74, 0.89)	0.82 (0.75, 0.89)
40-60	1.00 (0.92, 1.08)	1.00 (0.92, 1.08)
60-75	1.07 (0.98, 1.17)	1.07 (0.99, 1.16)
75-90	1.23 (1.13, 1.33)	1.22 (1.13, 1.33)
90-95	1.41 (1.26, 1.58)	1.40 (1.26, 1.56)
95-99	1.62 (1.44, 1.82)	1.60 (1.43, 1.80)
99-100	2.28 (1.91, 2.73)	2.24 (1.88, 2.67)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

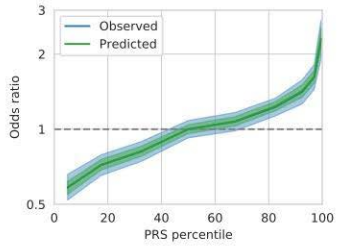

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

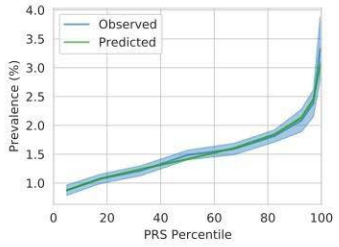

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.5 Chronic Lymphocytic Leukemia

Chronic lymphocytic leukemia is a rare disease; one in 175 people will develop the condition during their lifetime[45]. The PRS model for chronic lymphocytic leukemia includes 37 variants and was developed based on the variants identified in a study that analyzed genomes of about 4,500 individuals of European ancestry affected by the condition[46]. Heritability of chronic lymphocytic leukemia is about 20%[47] and about 25% of familial risk was explained by the variants discovered[46]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.676). To the best of our knowledge, the performance of other PRS models for chronic lymphocytic leukemia has not been published, however, a similar level of relative risk (OR 3.64 for the PRS percentile >80th as compared to 40-60th) was demonstrated by a comparable PRS model[48].

Number of individuals	408,885
Number of cases (%)	726 (0.2)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.676 (0.659, 0.695)
Covariates	0.699 (0.679, 0.715)
PRS + covariates	0.755 (0.740, 0.774)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

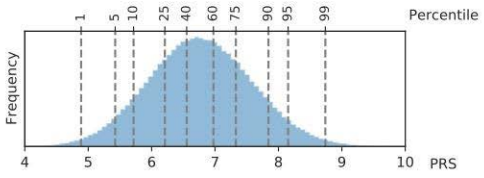

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.46 (0.29, 0.72)	0.46 (0.29, 0.72)
10-25	0.71 (0.51, 0.99)	0.71 (0.51, 0.99)
25-40	0.86 (0.63, 1.17)	0.86 (0.63, 1.17)
40-60	1.00 (0.76, 1.32)	1.00 (0.76, 1.32)
60-75	1.52 (1.16, 1.99)	1.52 (1.16, 1.98)
75-90	2.19 (1.71, 2.81)	2.19 (1.71, 2.81)
90-95	3.14 (2.34, 4.21)	3.13 (2.33, 4.20)
95-99	4.07 (3.04, 5.46)	4.06 (3.03, 5.43)
99-100	8.19 (5.69, 11.79)	8.12 (5.66, 11.66)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

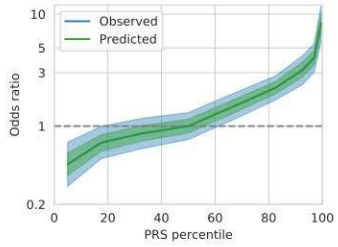

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

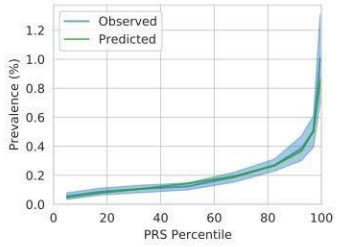

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.6 Melanoma

Melanoma is common among people of some ethnicities; one in 38 Europeans will develop melanoma during their lifetime[49]. The PRS model for melanoma includes 13 variants and was developed based on the variants identified in a study that analyzed genomes of over 6,600 individuals of European ancestry affected by melanoma[50]. Heritability of melanoma is about 60%[24] and about 9% of familial risk was explained by the variants discovered[50]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.578) comparable to the previously published PRS models for melanoma (AUC 0.58[25] - 0.61[26]).

Number of individuals	408,885
Number of cases (%)	5,003 (1.2)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.578 (0.572, 0.586)
Covariates	0.587 (0.578, 0.595)
PRS + covariates	0.617 (0.611, 0.625)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

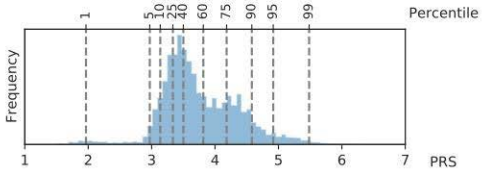

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.64 (0.57, 0.73)	0.65 (0.57, 0.73)
10-25	0.75 (0.68, 0.83)	0.76 (0.68, 0.84)
25-40	0.86 (0.78, 0.94)	0.86 (0.78, 0.94)
40-60	1.00 (0.92, 1.09)	1.00 (0.92, 1.09)
60-75	1.15 (1.05, 1.26)	1.15 (1.05, 1.26)
75-90	1.30 (1.19, 1.42)	1.29 (1.19, 1.41)
90-95	1.48 (1.31, 1.67)	1.47 (1.31, 1.66)
95-99	1.70 (1.50, 1.93)	1.69 (1.49, 1.91)
99-100	1.84 (1.47, 2.29)	1.82 (1.47, 2.26)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

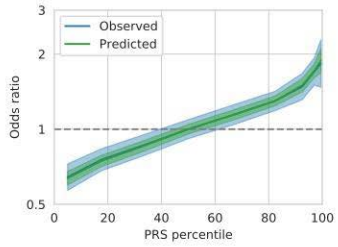

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

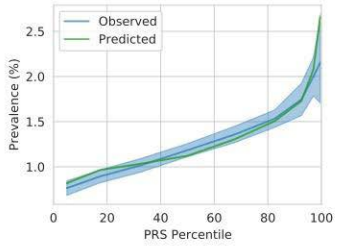

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.7 Pancreatic Cancer

Pancreatic cancer is a common disease; one in 63 people will develop the condition during their lifetime[51]. The PRS model for pancreatic cancer includes 22 variants and was developed based on the variants identified in a study that analyzed genomes of over 9,000 individuals of European ancestry affected by the condition[52]. Heritability of pancreatic cancer is unknown. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.595) comparable to the previously published PRS models for pancreatic cancer (AUC 0.61[53]).

Number of individuals	408,885
Number of cases (%)	939 (0.2)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.595 (0.578, 0.611)
Covariates	0.681 (0.664, 0.698)
PRS + covariates	0.704 (0.690, 0.722)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

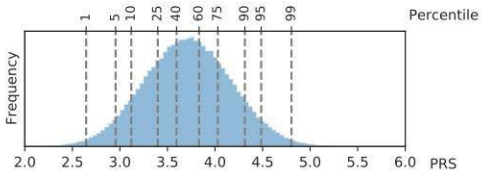

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.59 (0.43, 0.81)	0.59 (0.43, 0.81)
10-25	0.71 (0.55, 0.92)	0.71 (0.55, 0.92)
25-40	0.89 (0.70, 1.12)	0.89 (0.70, 1.12)
40-60	1.00 (0.81, 1.24)	1.00 (0.81, 1.24)
60-75	1.40 (1.13, 1.72)	1.40 (1.13, 1.72)
75-90	1.42 (1.15, 1.75)	1.42 (1.15, 1.75)
90-95	1.55 (1.17, 2.07)	1.55 (1.17, 2.06)
95-99	2.03 (1.54, 2.69)	2.03 (1.53, 2.68)
99-100	2.48 (1.57, 3.90)	2.47 (1.57, 3.88)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

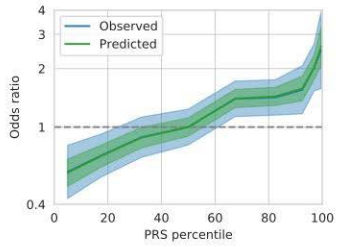

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

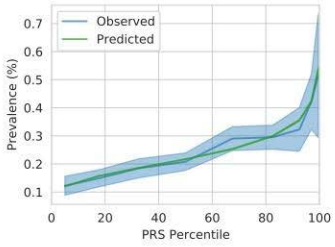

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.8 Prostate Cancer

Prostate cancer is a very common disease; one in nine men will be diagnosed with the condition during their lifetime[54]. The PRS model for prostate cancer includes 138 variants and was developed based on the variants identified in a study that analyzed genomes of ~80,000 men of European ancestry affected by the condition[55]. Heritability of prostate cancer is about 60%[56] and over 28% of familial risk was explained by the variants discovered[55]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.658) slightly exceeding the performance of the previously published PRS models for prostate cancer (AUC 0.59 – 0.65)[17,25,57,58,59].

Number of individuals	187,880
Number of cases (%)	8,155 (4.3)
Median age (25 & 75 percentiles)	67.5 (59.6 & 72.5)

Table 1: Validation cohort characteristics.

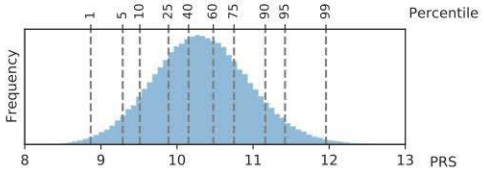

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

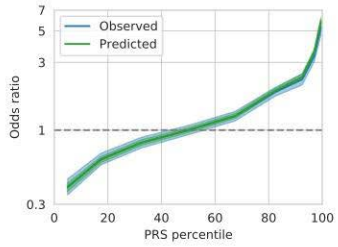

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

Features	AUC (95% CI)
PRS	0.658 (0.654, 0.665)
Covariates	0.721 (0.717, 0.726)
PRS + covariates	0.773 (0.768, 0.778)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.40 (0.35, 0.45)	0.41 (0.36, 0.46)
10-25	0.63 (0.57, 0.69)	0.63 (0.58, 0.70)
25-40	0.82 (0.75, 0.89)	0.82 (0.75, 0.89)
40-60	1.00 (0.93, 1.08)	1.00 (0.93, 1.08)
60-75	1.25 (1.16, 1.36)	1.24 (1.15, 1.34)
75-90	1.86 (1.74, 2.00)	1.81 (1.69, 1.94)
90-95	2.28 (2.08, 2.50)	2.18 (2.00, 2.38)
95-99	3.32 (3.04, 3.63)	3.06 (2.82, 3.33)
99-100	5.32 (4.66, 6.07)	4.60 (4.10, 5.15)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

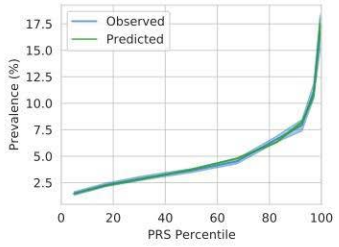

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.9 Thyroid Cancer

Thyroid cancer is a rare disease; one in 120 women and one in 300 men will be diagnosed with thyroid cancer during their lifetime[60]. The PRS model for thyroid cancer includes 10 variants and was developed based on the variants identified by a study that analyzed genomes of about 3,000 individuals of European ancestry affected by the condition[61]. Heritability of thyroid cancer is about 50%[62]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank and the results are presented in Tables 1-3 and Figures 1- 3. The developed model demonstrated modest predictive performance (AUC 0.616). To the best of our knowledge, the performance of other PRS models for thyroid cancer has not been published yet, however, a relative risk for the top quartile (OR 3.2, >75%/<25%) was reported for a comparable PRS model[63] .

Number of individuals	408,885
Number of cases (%)	619 (0.2)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.616 (0.596, 0.640)
Covariates	0.632 (0.611, 0.651)
PRS + covariates	0.683 (0.665, 0.700)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

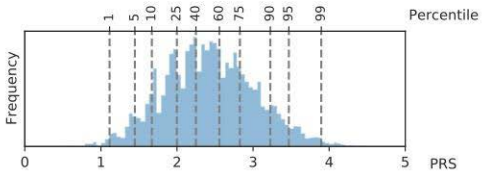

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.74 (0.50, 1.07)	0.74 (0.50, 1.07)
10-25	0.87 (0.63, 1.19)	0.87 (0.63, 1.19)
25-40	1.04 (0.77, 1.40)	1.04 (0.77, 1.40)
40-60	1.00 (0.75, 1.33)	1.00 (0.75, 1.33)
60-75	1.43 (1.08, 1.89)	1.43 (1.08, 1.89)
75-90	1.61 (1.23, 2.12)	1.61 (1.23, 2.11)
90-95	2.11 (1.50, 2.96)	2.11 (1.50, 2.95)
95-99	3.20 (2.33, 4.40)	3.19 (2.33, 4.39)
99-100	4.67 (2.96, 7.36)	4.65 (2.95, 7.32)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

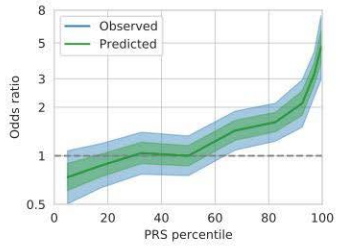

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

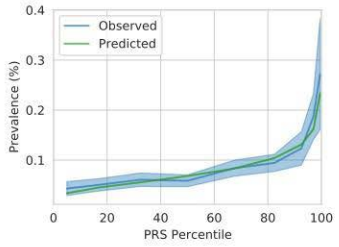

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.10 Celiac Disease

Celiac disease is a common condition; one in 100 people is affected by the disease[64]. The PRS model for celiac disease includes 56 variants: 55 non-HLA variants[65] and the HLA DQ2.5 haplotype represented by a proxy SNP[66] . This HLA allele is the most prevalent HLA allele associated with celiac disease among affected Europeans and found in over 90% of cases[67]. The model was developed based on the variants identified in a study of over 12,000 individuals of European ancestry affected by the condition[65,66]. Heritability of susceptibility to celiac disease is over 75%[68] and about 54% of heritability was explained by the variants identified[69]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.808), however, below the performance of a similar PRS model for celiac disease including more HLA alleles (AUC 0.85)[69].

Number of individuals	408,885
Number of cases (%)	2,783 (0.7)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.808 (0.799, 0.818)
Covariates	0.579 (0.570, 0.590)
PRS + covariates	0.814 (0.806, 0.823)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

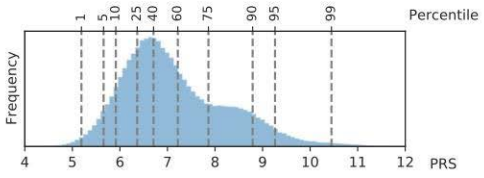

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.62 (0.46, 0.84)	0.62 (0.46, 0.84)
10-25	0.71 (0.55, 0.92)	0.71 (0.55, 0.92)
25-40	0.77 (0.61, 0.99)	0.77 (0.61, 0.99)
40-60	1.00 (0.81, 1.23)	1.00 (0.81, 1.23)
60-75	1.85 (1.52, 2.25)	1.85 (1.52, 2.24)
75-90	6.35 (5.39, 7.48)	6.28 (5.33, 7.39)
90-95	11.25 (9.45, 13.39)	11.01 (9.26, 13.09)
95-99	16.69 (14.07, 19.80)	16.15 (13.63, 19.12)
99-100	31.85 (26.21, 38.69)	29.88 (24.73, 36.11)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

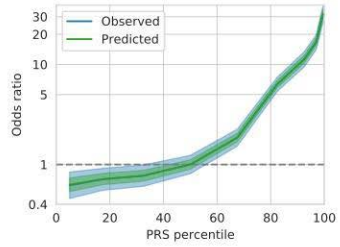

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

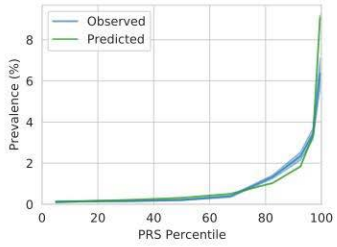

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.11 Inflammatory Bowel Disease

Inflammatory bowel disease is common; one in 80 adults in the US are currently affected with either ulcerative colitis or Crohn's disease[70]. The PRS model for inflammatory bowel disease includes 231 variants. The model was developed based on the variants identified by a study that analyzed genomes of over 37,000 individuals of European ancestry affected by either Crohn's disease or ulcerative colitis[71,72]. Heritability of inflammatory bowel disease is about 70%[73] and about 13% of heritability of Crohn's disease and 8.2% of heritability of ulcerative colitis was explained by the variants identified[71]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.641), however, below the performance of similar PRS models for Crohn's disease (AUC 0.75) and ulcerative colitis (AUC 0.70)[74].

Number of individuals	393,379
Number of cases (%)	5,938 (1.5)
Median age (25 & 75 percentiles)	67.0 (59.4 & 72.1)
Number of females (%)	181,896 (46.2)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.641 (0.635, 0.649)
Covariates	0.532 (0.525, 0.539)
PRS + covariates	0.645 (0.639, 0.652)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

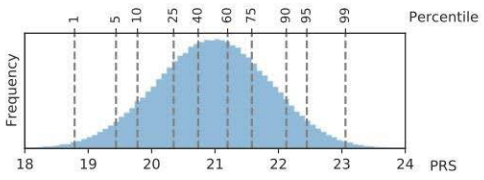

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.50 (0.43, 0.57)	0.50 (0.44, 0.58)
10-25	0.65 (0.58, 0.72)	0.65 (0.58, 0.72)
25-40	0.83 (0.75, 0.92)	0.83 (0.75, 0.92)
40-60	1.00 (0.92, 1.09)	1.00 (0.92, 1.09)
60-75	1.29 (1.18, 1.41)	1.28 (1.18, 1.40)
75-90	1.76 (1.62, 1.92)	1.75 (1.61, 1.90)
90-95	2.15 (1.94, 2.39)	2.12 (1.91, 2.35)
95-99	2.90 (2.61, 3.21)	2.83 (2.55, 3.13)
99-100	4.13 (3.54, 4.83)	3.98 (3.43, 4.61)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

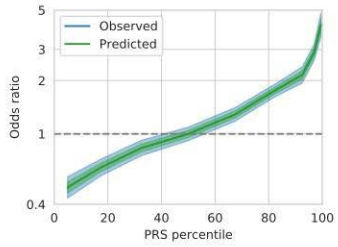

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

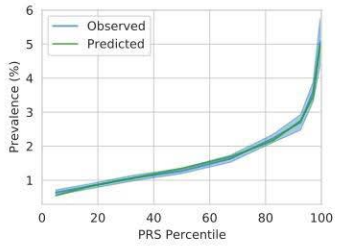

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.12 Atrial Fibrillation

Atrial fibrillation is a very common condition; one in four people 40 years or older will develop atrial fibrillation during their lifetime[75]. The PRS model for atrial fibrillation includes 82 variants and was developed based on the variants identified in a study that analyzed genomes of over 55,000 individuals of European ancestry affected by the condition[76]. Heritability of atrial fibrillation is ~60%[77] and over 42% of heritability was explained by the variants discovered[76]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.622), slightly below the performance of a previously published PRS model for atrial fibrillation (AUC 0.64)[25].

Number of individuals	408,885
Number of cases (%)	19,572 (4.8)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.622 (0.619, 0.627)
Covariates	0.740 (0.736, 0.742)
PRS + covariates	0.765 (0.762, 0.768)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

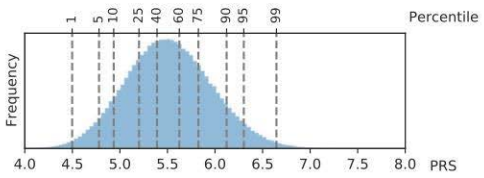

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.53 (0.50, 0.57)	0.54 (0.51, 0.58)
10-25	0.67 (0.63, 0.71)	0.68 (0.64, 0.72)
25-40	0.78 (0.74, 0.82)	0.79 (0.75, 0.83)
40-60	1.00 (0.95, 1.05)	1.00 (0.96, 1.05)
60-75	1.21 (1.15, 1.27)	1.20 (1.15, 1.26)
75-90	1.51 (1.44, 1.58)	1.48 (1.41, 1.54)
90-95	1.94 (1.83, 2.06)	1.87 (1.76, 1.97)
95-99	2.45 (2.31, 2.61)	2.31 (2.18, 2.44)
99-100	3.65 (3.32, 4.01)	3.27 (3.01, 3.55)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

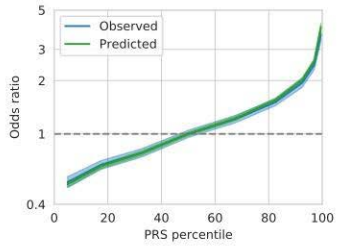

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

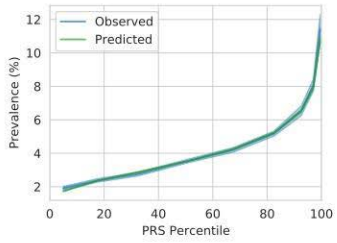

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.13 Coronary Artery Disease

Coronary artery disease is very common; one in two men and one in three women 40 years or older will develop the condition during their lifetime[78]. The PRS model for coronary artery disease includes 41 variants and was developed based on the variants identified in a study that analyzed genomes of over 60,000 individuals of European ancestry affected by the condition[79]. Heritability of coronary artery disease is about 50-60%[80].

The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.576), below the performance of previously published PRS models for coronary artery disease (AUC 0.65[81], cIndex 0.62[82]).

Number of individuals	408,885
Number of cases (%)	22,387 (5.5)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.576 (0.573, 0.580)
Covariates	0.757 (0.754, 0.760)
PRS + covariates	0.767 (0.764, 0.770)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

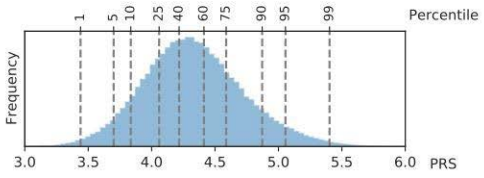

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.66 (0.63, 0.71)	0.68 (0.64, 0.72)
10-25	0.74 (0.70, 0.78)	0.75 (0.72, 0.79)
25-40	0.85 (0.81, 0.90)	0.86 (0.82, 0.90)
40-60	1.00 (0.96, 1.04)	1.00 (0.96, 1.04)
60-75	1.10 (1.05, 1.15)	1.10 (1.05, 1.14)
75-90	1.27 (1.22, 1.33)	1.25 (1.20, 1.30)
90-95	1.47 (1.38, 1.56)	1.43 (1.35, 1.51)
95-99	1.71 (1.60, 1.81)	1.64 (1.55, 1.74)
99-100	2.19 (1.98, 2.43)	2.06 (1.88, 2.26)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

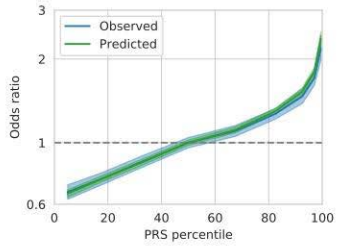

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

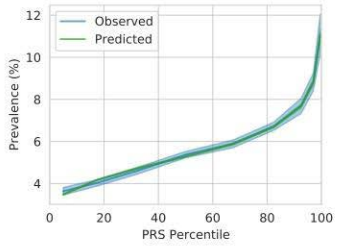

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.14 Venous Thromboembolism

Venous thromboembolism (VTE) is a very common condition; one in 12 people 45 years or older will develop the condition during their lifetime[83]. The PRS model for VTE includes 10 variants and was developed based on the variants identified in a study that analyzed genomes of ~3,300 individuals of European ancestry affected by the condition[84]. The model includes the factor V Leiden (F5) variant rs6025 (p.R506Q) and the F2 (prothrombin) variant G20210A. Heritability of VTE is about 40 - 50%[85] and over 5% was explained by the variants discovered[84]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.600). To the best of our knowledge, the performance of other PRS models for VTE has not been published yet, however, odds ratios for top PRS percentiles have been provided for a larger PRS model with either F5 or F2 variant included 4.62-4.90 (>95%/<95%)[86].

Number of individuals	407,351
Number of cases (%)	14,122 (3.5)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,166 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.600 (0.595, 0.605)
Covariates	0.616 (0.611, 0.619)
PRS + covariates	0.655 (0.651, 0.659)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

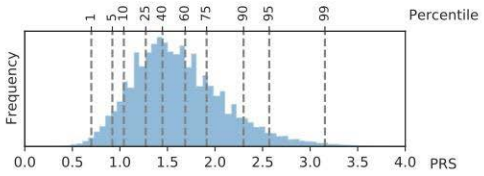

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.70 (0.64, 0.75)	0.70 (0.65, 0.76)
10-25	0.81 (0.76, 0.86)	0.81 (0.76, 0.86)
25-40	0.89 (0.84, 0.95)	0.90 (0.84, 0.95)
40-60	1.00 (0.95, 1.06)	1.00 (0.95, 1.06)
60-75	1.19 (1.12, 1.26)	1.18 (1.11, 1.25)
75-90	1.45 (1.37, 1.54)	1.43 (1.36, 1.51)
90-95	1.83 (1.70, 1.97)	1.78 (1.66, 1.91)
95-99	2.43 (2.26, 2.61)	2.33 (2.18, 2.49)
99-100	4.77 (4.32, 5.27)	4.28 (3.92, 4.68)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

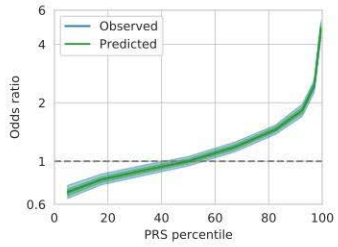

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

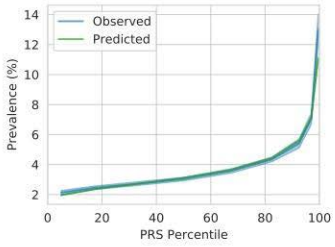

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.15 Schizophrenia

Schizophrenia is a common disease; one in 100 people will develop the condition[87]. The PRS model for schizophrenia includes 160 variants and was developed based on the variants identified in a study that analyzed genomes of over 40,000 individuals of European ancestry affected by the condition[88]. Heritability of schizophrenia is up to 80%[89]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.559), below the performance of previously published PRS models for schizophrenia (AUC 0.58[90,91] - 0.64[91]).

Number of individuals	408,885
Number of cases (%)	1,096 (0.3)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.559 (0.545, 0.573)
Covariates	0.584 (0.571, 0.601)
PRS + covariates	0.599 (0.585, 0.612)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

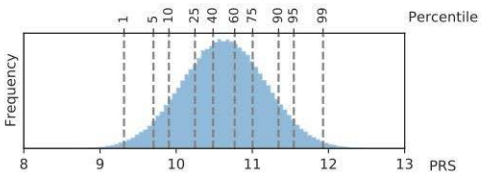

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.62 (0.47, 0.81)	0.62 (0.47, 0.81)
10-25	0.71 (0.57, 0.89)	0.71 (0.57, 0.89)
25-40	1.00 (0.82, 1.22)	1.00 (0.82, 1.22)
40-60	1.00 (0.83, 1.20)	1.00 (0.83, 1.20)
60-75	1.08 (0.88, 1.31)	1.08 (0.88, 1.31)
75-90	1.11 (0.91, 1.34)	1.11 (0.91, 1.34)
90-95	1.33 (1.03, 1.74)	1.33 (1.03, 1.73)
95-99	1.47 (1.11, 1.93)	1.46 (1.11, 1.93)
99-100	1.71 (1.07, 2.74)	1.71 (1.07, 2.73)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

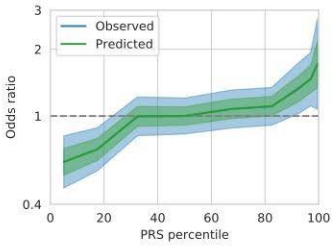

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

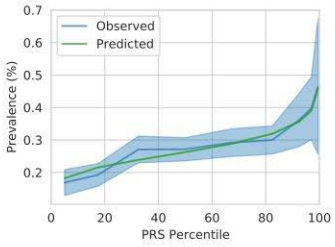

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.16 HDL Cholesterol

Having low levels of HDL cholesterol is very common; nearly one in three adults has low HDL cholesterol levels[92]. The PRS model for predicting low levels of HDL cholesterol includes 311 variants and was developed based on the variants identified in a study that analyzed genomes of ~300,000 multiethnic individuals (White - 72.8%, Black - 18.9%, Hispanic - 8.1%) with normal and abnormal HDL levels[93]. Heritability of HDL cholesterol levels is about 40-60%[94,95] and 12.3% was explained by the variants discovered[93]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.689).

Number of individuals	317,466
Number of cases (%)	30,610 (9.6)
Median age (25 & 75 percentiles)	66.3 (58.7 & 71.6)
Number of females (%)	125,061 (39.4)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.689 (0.686, 0.692)
Covariates	0.705 (0.703, 0.708)
PRS + covariates	0.786 (0.784, 0.788)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

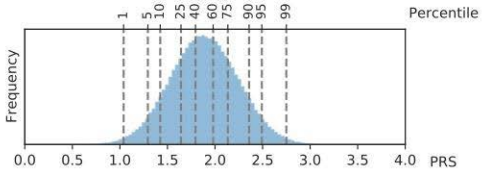

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.26 (0.24, 0.28)	0.28 (0.25, 0.30)
10-25	0.48 (0.45, 0.50)	0.50 (0.47, 0.52)
25-40	0.71 (0.68, 0.75)	0.73 (0.70, 0.76)
40-60	1.00 (0.96, 1.04)	1.00 (0.96, 1.04)
60-75	1.39 (1.33, 1.44)	1.34 (1.30, 1.39)
75-90	1.97 (1.89, 2.04)	1.82 (1.76, 1.88)
90-95	2.69 (2.56, 2.82)	2.36 (2.26, 2.46)
95-99	3.57 (3.40, 3.75)	2.94 (2.83, 3.06)
99-100	5.95 (5.51, 6.44)	4.22 (4.00, 4.46)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

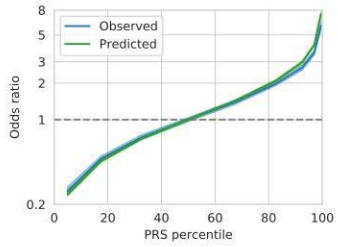

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

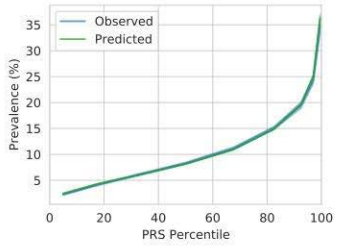

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.17 LDL Cholesterol

Having high cholesterol is very common; nearly one in three adults has a high cholesterol level[92]. The PRS model for predicting high levels of LDL cholesterol includes 218 variants and was developed based on the variants identified in a study that analyzed genomes of ~300,000 multiethnic individuals (White - 72.8%, Black - 18.9%, Hispanic - 8.1%) with normal and abnormal LDL levels[93]. Heritability of LDL cholesterol levels is ~30 - 50%[94,95] and 9.6% was explained by the variants discovered[9]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.643).

Number of individuals	345,736
Number of cases (%)	216,337 (62.6)
Median age (25 & 75 percentiles)	66.3 (58.7 & 71.6)
Number of females (%)	134,906 (39.0)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.643 (0.641, 0.644)
Covariates	0.567 (0.565, 0.569)
PRS + covariates	0.661 (0.660, 0.663)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

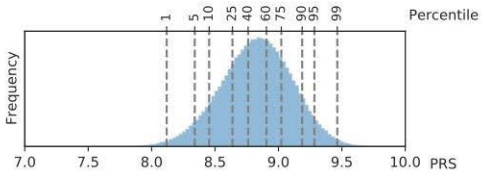

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.32 (0.31, 0.33)	0.57 (0.57, 0.58)
10-25	0.57 (0.55, 0.58)	0.79 (0.78, 0.80)
25-40	0.78 (0.76, 0.80)	0.91 (0.90, 0.92)
40-60	1.00 (0.98, 1.02)	1.00 (0.99, 1.01)
60-75	1.24 (1.21, 1.28)	1.07 (1.07, 1.08)
75-90	1.50 (1.47, 1.54)	1.13 (1.12, 1.14)
90-95	1.89 (1.82, 1.97)	1.20 (1.19, 1.21)
95-99	2.14 (2.05, 2.23)	1.23 (1.22, 1.24)
99-100	2.49 (2.28, 2.72)	1.27 (1.25, 1.29)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

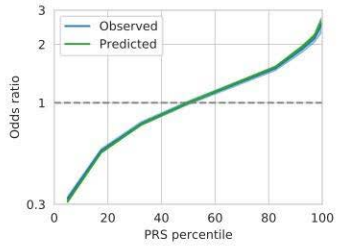

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

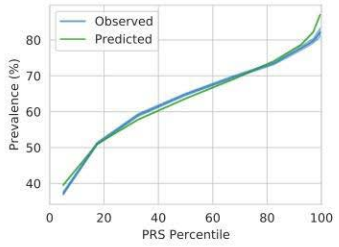

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.18 Triglycerides

Having a high triglyceride level is very common; one in four adults has a high triglyceride level[96]. The PRS model for predicting high levels of triglycerides includes 249 variants and was developed based on the variants identified in a study that analyzed genomes of ~300,000 multiethnic individuals (White - 72.8%, Black - 18.9%, Hispanic - 8.1%) with normal and abnormal triglyceridelevels[93] . Heritability of triglyceride levels is ~50%[97] and 11% was explained by the variants discovered[93] . The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.629).

Number of individuals	346,072
Number of cases (%)	135,959 (39.3)
Median age (25 & 75 percentiles)	66.3 (58.7 & 71.6)
Number of females (%)	135,043 (39.0)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.629 (0.627, 0.631)
Covariates	0.614 (0.613, 0.617)
PRS + covariates	0.672 (0.670, 0.674)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

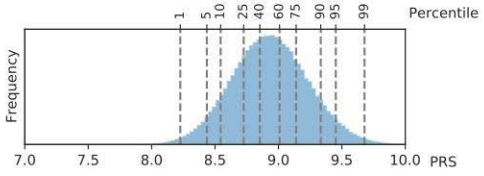

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.43 (0.42, 0.44)	0.55 (0.54, 0.56)
10-25	0.64 (0.62, 0.65)	0.74 (0.73, 0.75)
25-40	0.81 (0.79, 0.83)	0.87 (0.86, 0.88)
40-60	1.00 (0.98, 1.02)	1.00 (0.99, 1.01)
60-75	1.25 (1.22, 1.28)	1.14 (1.13, 1.16)
75-90	1.57 (1.54, 1.61)	1.29 (1.27, 1.30)
90-95	1.99 (1.92, 2.06)	1.44 (1.41, 1.46)
95-99	2.43 (2.34, 2.52)	1.56 (1.54, 1.59)
99-100	3.28 (3.05, 3.52)	1.74 (1.70, 1.78)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

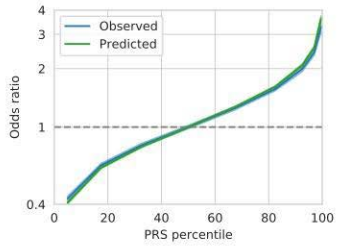

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

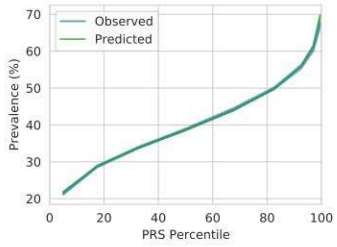

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.19 Total Cholesterol

Having high cholesterol is very common; nearly one in three adults has a high cholesterol level[92]. The PRS model for predicting high levels of total cholesterol includes 225 variants and was developed based on the variants identified in a study that analyzed genomes of ~300,000 multiethnic individuals (White - 72.8%, Black - 18.9%, Hispanic - 8.1%) with normal and abnormal total cholesterol levels[93]. Heritability of total cholesterol levels is ~30 - 50%[94,95] and 8.9% was explained by the variants discovered[93]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.628) similar to the previously published PRS model for total cholesterol (AUC 0.63[25]).

Number of individuals	346,335
Number of cases (%)	251,566 (72.6)
Median age (25 & 75 percentiles)	66.3 (58.7 & 71.6)
Number of females (%)	135,188 (39.0)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.628 (0.626, 0.630)
Covariates	0.567 (0.566, 0.569)
PRS + covariates	0.647 (0.645, 0.648)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

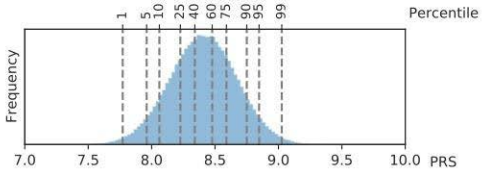

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.41 (0.40, 0.42)	0.73 (0.72, 0.73)
10-25	0.62 (0.61, 0.64)	0.86 (0.86, 0.87)
25-40	0.81 (0.79, 0.83)	0.94 (0.93, 0.95)
40-60	1.00 (0.98, 1.02)	1.00 (0.99, 1.01)
60-75	1.23 (1.20, 1.27)	1.05 (1.05, 1.06)
75-90	1.52 (1.48, 1.56)	1.10 (1.09, 1.10)
90-95	1.84 (1.76, 1.93)	1.13 (1.13, 1.14)
95-99	2.20 (2.09, 2.32)	1.17 (1.16, 1.17)
99-100	2.67 (2.40, 2.96)	1.19 (1.18, 1.21)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

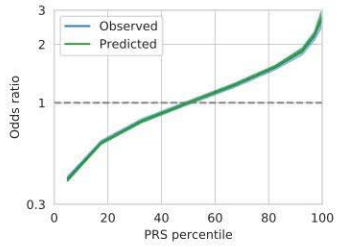

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

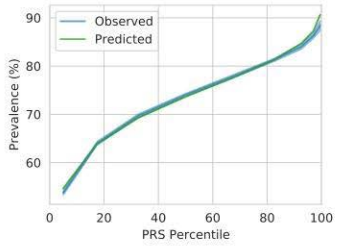

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.20 Type 2 Diabetes

Type 2 diabetes is a very common disease; two out of five people will develop the condition during their lifetime[98]. The PRS model for type 2 diabetes includes 47 variants and was developed based on the variants identified in a study that analyzed genomes of over 41,000 individuals of European ancestry affected by the condition[99]. Heritability of type 2 diabetes is about 70%[100]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.593), below the performance of previously published PRS models for type 2 diabetes (AUC - 0.64[25] and 0.58 – 0.65[15]).

Number of individuals	408,885
Number of cases (%)	28,649 (7.0)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.593 (0.590, 0.597)
Covariates	0.649 (0.647, 0.652)
PRS + covariates	0.674 (0.670, 0.677)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

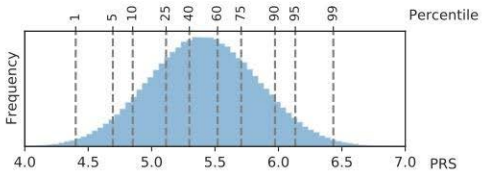

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.60 (0.56, 0.63)	0.61 (0.58, 0.65)
10-25	0.75 (0.71, 0.78)	0.76 (0.73, 0.79)
25-40	0.87 (0.83, 0.91)	0.87 (0.84, 0.91)
40-60	1.00 (0.96, 1.04)	1.00 (0.96, 1.04)
60-75	1.22 (1.17, 1.27)	1.20 (1.16, 1.25)
75-90	1.41 (1.36, 1.47)	1.37 (1.32, 1.42)
90-95	1.64 (1.56, 1.73)	1.57 (1.50, 1.65)
95-99	1.94 (1.83, 2.05)	1.82 (1.74, 1.92)
99-100	2.58 (2.36, 2.82)	2.34 (2.16, 2.52)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

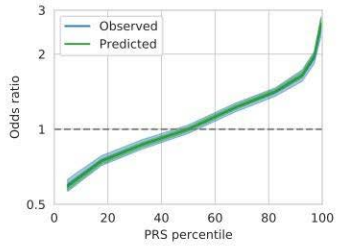

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

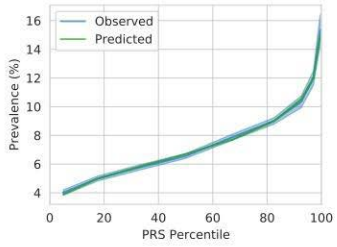

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.21 Dupuytren's Disease

Dupuytren's disease is very common among people of some ethnicities; up to one in four people of European descent are affected by the condition[101]. The PRS model for Dupuytren's disease includes 25 variants and was developed based on the variants identified in a study that analyzed genomes of about 8,000 individuals of European ancestry affected by the condition[101]. Heritability of Dupuytren's disease is about 80%[102], of which 11% was explained by the variants identified[101]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.681). To the best of our knowledge, the performance of other PRS models for Dupuytren's disease has not been published, however, a relative risk for developing a bilateral diseases for the top quartile (OR 1.78, >75%/<25%) was reported for a similar PRS model[103].

Number of individuals	408,885
Number of cases (%)	4,318 (1.1)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.681 (0.675, 0.689)
Covariates	0.737 (0.729, 0.744)
PRS + covariates	0.790 (0.784, 0.795)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

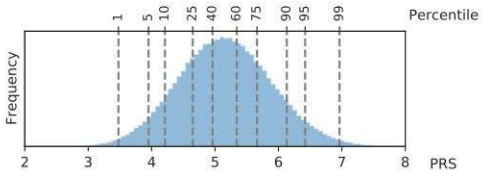

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.41 (0.34, 0.49)	0.41 (0.34, 0.50)
10-25	0.60 (0.52, 0.68)	0.60 (0.52, 0.69)
25-40	0.79 (0.70, 0.90)	0.79 (0.70, 0.90)
40-60	1.00 (0.90, 1.12)	1.00 (0.90, 1.11)
60-75	1.48 (1.33, 1.65)	1.47 (1.33, 1.64)
75-90	2.10 (1.90, 2.32)	2.08 (1.89, 2.30)
90-95	2.81 (2.49, 3.18)	2.77 (2.46, 3.12)
95-99	4.09 (3.64, 4.60)	3.99 (3.56, 4.48)
99-100	5.99 (5.08, 7.08)	5.77 (4.91, 6.77)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

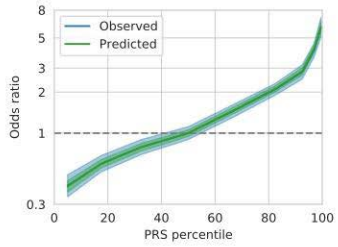

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

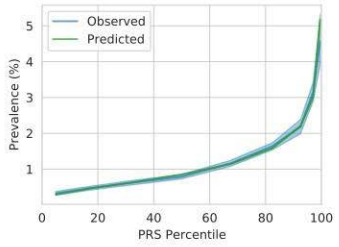

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.22 Gout

Gout is a common disease; one in 50 people will develop the condition during their lifetime[104]. Gout happens when uric acid builds up in the body. The PRS model for gout includes 29 variants and was developed based on the variants identified in a study that analyzed genomes of ~140,000 individuals of European ancestry with normal and abnormal serum urate levels as well as people affected by gout[105]. Heritability of urate levels is up to 60%[106] and about 7% of it was explained by the variants identified[105]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.641), but below the performance of a previously published PRS model for gout (AUC 0.68[25]).

Number of individuals	408,885
Number of cases (%)	8,574 (2.1)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.641 (0.635, 0.647)
Covariates	0.775 (0.771, 0.780)
PRS + covariates	0.807 (0.803, 0.812)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

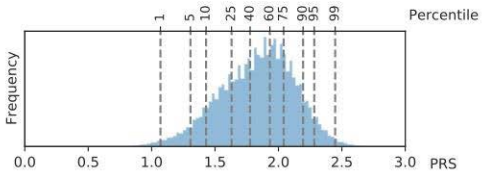

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.39 (0.34, 0.44)	0.39 (0.35, 0.45)
10-25	0.61 (0.56, 0.67)	0.61 (0.56, 0.67)
25-40	0.81 (0.74, 0.88)	0.81 (0.75, 0.88)
40-60	1.00 (0.93, 1.07)	1.00 (0.93, 1.07)
60-75	1.21 (1.12, 1.30)	1.20 (1.12, 1.29)
75-90	1.52 (1.42, 1.63)	1.51 (1.41, 1.61)
90-95	2.01 (1.84, 2.20)	1.98 (1.82, 2.15)
95-99	2.85 (2.62, 3.10)	2.75 (2.54, 2.99)
99-100	4.21 (3.71, 4.78)	3.97 (3.53, 4.47)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

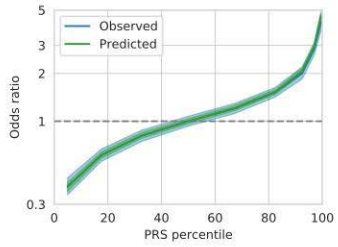

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

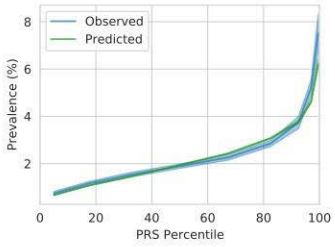

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.23 Osteoporosis

Osteoporosis is a very common disease; one in three women and one in five men 50 years or older will have fractures due to osteoporosis during their lifetime[107]. The PRS model for osteoporosis includes 1,146 variants and was developed based on the variants identified in a study that analyzed genomes of ~400,000 individuals of European ancestry with normal and abnormal bone mineral density[108] . Heritability of osteoporosis is about 60 - 90%[109,110]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.678), exceeding the performance of the previously published PRS models for osteoporosis (AUC 0.59[111], 0.604[108]).

Number of individuals	272,623
Number of cases (%)	3,799 (1.4)
Median age (25 & 75 percentiles)	67.2 (59.7 & 72.5)
Number of females (%)	124,702 (45.7)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.678 (0.671, 0.687)
Covariates	0.661 (0.653, 0.670)
PRS + covariates	0.725 (0.718, 0.735)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

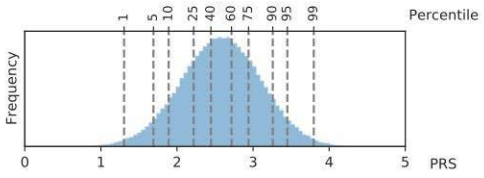

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.46 (0.38, 0.56)	0.46 (0.39, 0.56)
10-25	0.53 (0.46, 0.62)	0.54 (0.46, 0.62)
25-40	0.78 (0.68, 0.89)	0.78 (0.68, 0.89)
40-60	1.00 (0.89, 1.12)	1.00 (0.89, 1.12)
60-75	1.36 (1.21, 1.52)	1.35 (1.21, 1.51)
75-90	1.96 (1.76, 2.17)	1.94 (1.75, 2.15)
90-95	2.87 (2.53, 3.26)	2.81 (2.49, 3.19)
95-99	3.82 (3.37, 4.33)	3.70 (3.28, 4.18)
99-100	6.58 (5.54, 7.80)	6.20 (5.28, 7.29)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

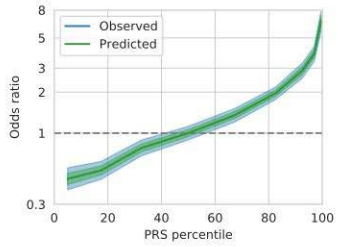

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

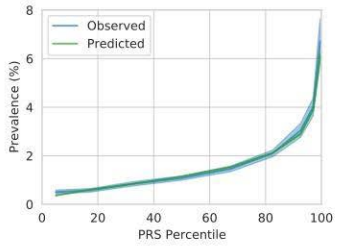

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.24 Rheumatoid Arthritis

Rheumatoid arthritis (RA) is a common disease; one in 28 women and one in 60 men will develop the condition during their lifetime[112]. The PRS model for RA includes 94 variants and was developed based on the variants identified in a study that analyzed genomes of over 18,000 individuals of European ancestry and over 11,700 individuals of Asian ancestry affected by RA[113]. The developed model includes a tag SNP for the HLADRB1*0401 allele, a strong genetic risk factor for RA conferring 3.35-fold increased risk for the disease[114]. Heritability of RA is about 60%[115], of which 4.7% and 3.8% was explained by variants discovered in Europeans and Asians, respectively[113]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.579). To the best of our knowledge, the performance of other PRS models for RA has not been published yet.

Number of individuals	408,885
Number of cases (%)	7,844 (1.9)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.579 (0.572, 0.584)
Covariates	0.631 (0.624, 0.636)
PRS + covariates	0.652 (0.646, 0.656)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

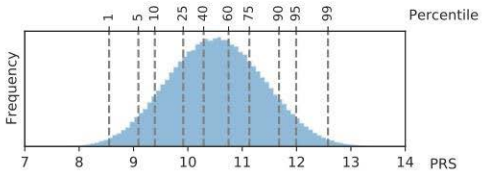

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.79 (0.72, 0.87)	0.79 (0.72, 0.88)
10-25	0.90 (0.83, 0.98)	0.90 (0.83, 0.98)
25-40	0.88 (0.81, 0.95)	0.88 (0.81, 0.96)
40-60	1.00 (0.93, 1.08)	1.00 (0.93, 1.08)
60-75	1.18 (1.09, 1.27)	1.17 (1.09, 1.26)
75-90	1.41 (1.31, 1.52)	1.40 (1.31, 1.51)
90-95	1.79 (1.62, 1.97)	1.76 (1.60, 1.94)
95-99	2.09 (1.89, 2.31)	2.05 (1.86, 2.26)
99-100	2.70 (2.30, 3.16)	2.62 (2.25, 3.05)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

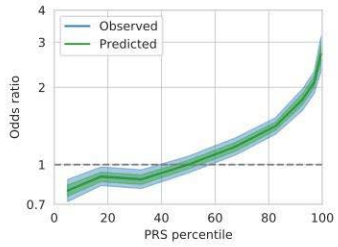

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

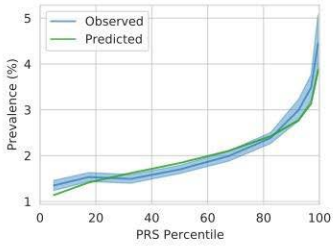

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.25 Alzheimer's Disease

Alzheimer's disease is a very common condition; one in 10 people 65 years or older will develop the disease during their lifetime[116]. The PRS model for Alzheimer's disease includes 20 variants and was developed based on the variants identified in a study that analyzed genomes of over 25,500 individuals of European ancestry affected by Alzheimer's disease[117]. The developed model includes the APOE e4 variant, a strong genetic risk factor conferring ~3.60 fold increased risk for Alzheimer's disease by itself[118]. Heritability of Alzheimer's disease is ~60-80%[119]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.696), slightly below the performance of comparable published PRS models (AUC 0.715 – 0.745[120]).

Number of individuals	408,885
Number of cases (%)	873 (0.2)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.696 (0.678, 0.714)
Covariates	0.787 (0.778, 0.802)
PRS + covariates	0.830 (0.818, 0.844)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

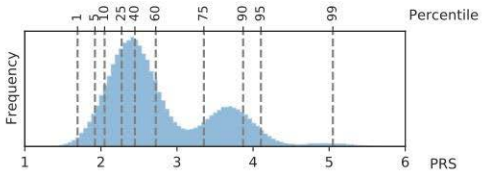

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.71 (0.48, 1.03)	0.71 (0.48, 1.03)
10-25	0.89 (0.65, 1.21)	0.89 (0.65, 1.21)
25-40	0.90 (0.66, 1.22)	0.90 (0.66, 1.22)
40-60	1.00 (0.76, 1.32)	1.00 (0.76, 1.32)
60-75	1.31 (0.99, 1.72)	1.31 (0.99, 1.72)
75-90	2.87 (2.27, 3.63)	2.86 (2.26, 3.62)
90-95	3.94 (2.99, 5.19)	3.92 (2.98, 5.16)
95-99	5.03 (3.82, 6.62)	5.00 (3.80, 6.58)
99-100	15.37 (11.41, 20.70)	15.10 (11.25, 20.26)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

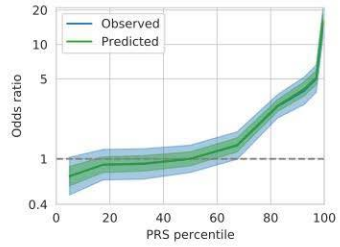

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

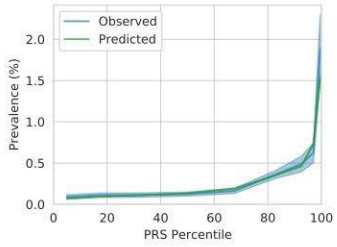

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.26 Multiple Sclerosis

Multiple sclerosis is a rare disease; one in 750 people will develop multiple sclerosis during their lifetime[121]. The PRS model for multiple sclerosis includes 102 variants and was developed based on the variants identified in studies that analyzed genomes of over 46,000 individuals of European ancestry affected by multiple sclerosis[122,123]. The developed model includes both non-HLA and HLA alleles as the both types contribute to the disease pathogenesis[123]. Heritability of multiple sclerosis is ~25-76%[124] and ~28% was explained by the variants discovered[122]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated good predictive performance (AUC 0.692), comparable to the performance of a previously published model (AUC 0.70[125]).

Number of individuals	408,885
Number of cases (%)	1,871 (0.5)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.692 (0.680, 0.706)
Covariates	0.623 (0.613, 0.634)
PRS + covariates	0.724 (0.714, 0.737)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

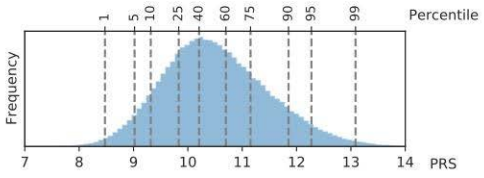

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.39 (0.29, 0.53)	0.39 (0.29, 0.53)
10-25	0.62 (0.50, 0.76)	0.62 (0.50, 0.76)
25-40	0.82 (0.68, 1.00)	0.82 (0.68, 1.00)
40-60	1.00 (0.84, 1.19)	1.00 (0.84, 1.19)
60-75	1.43 (1.21, 1.69)	1.43 (1.21, 1.69)
75-90	2.24 (1.92, 2.61)	2.23 (1.91, 2.59)
90-95	3.39 (2.83, 4.05)	3.36 (2.82, 4.02)
95-99	4.62 (3.87, 5.50)	4.56 (3.84, 5.43)
99-100	6.40 (4.99, 8.20)	6.29 (4.92, 8.03)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

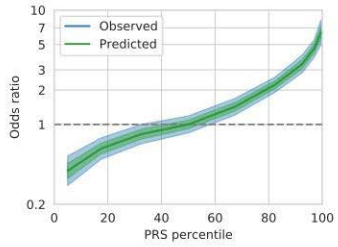

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

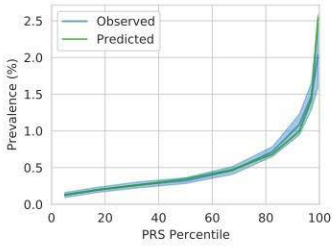

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.27 Parkinson's Disease

Parkinson's disease is a common condition; one in 100 people 65 years or older will develop the disease during their lifetime[126]. The PRS model for Parkinson's disease includes 46 variants and was developed based on the variants identified in a study that analyzed genomes of over 26,000 individuals of European ancestry affected by Parkinson's disease[127]. Heritability of Parkinson's disease is ~35-60%[128]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.593), slightly below the performance of a published PRS model for Parkinson's disease (AUC 0.61[129]).

Number of individuals	408,885
Number of cases (%)	1,762 (0.4)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.593 (0.581, 0.602)
Covariates	0.739 (0.730, 0.749)
PRS + covariates	0.756 (0.747, 0.765)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

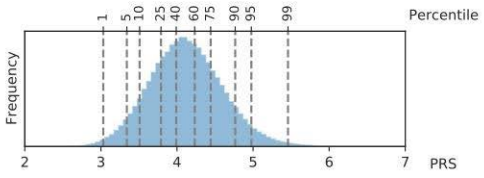

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.64 (0.51, 0.79)	0.64 (0.51, 0.79)
10-25	0.65 (0.54, 0.78)	0.65 (0.54, 0.78)
25-40	0.91 (0.77, 1.07)	0.91 (0.77, 1.07)
40-60	1.00 (0.86, 1.16)	1.00 (0.86, 1.16)
60-75	1.19 (1.02, 1.39)	1.19 (1.02, 1.39)
75-90	1.34 (1.15, 1.56)	1.33 (1.15, 1.55)
90-95	1.63 (1.33, 1.99)	1.63 (1.33, 1.99)
95-99	1.95 (1.59, 2.39)	1.94 (1.59, 2.38)
99-100	2.67 (1.95, 3.66)	2.65 (1.94, 3.62)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

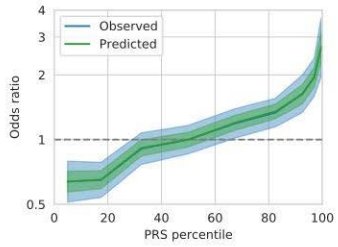

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

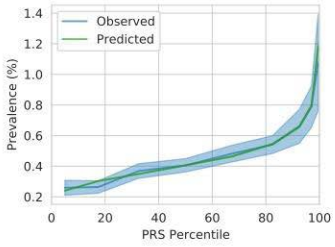

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.28 Asthma

Asthma is a very common disease; one in 13 people currently have asthma[130]. The PRS model for asthma includes 17 variants and was developed based on the variants identified in a study that analyzed genomes of ~20,000 individuals of European ancestry affected by asthma[131]. Heritability of asthma is ~35-95%[132] and ~3.5% was explained by the variants identified[131]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.559), below the performance of the previously published PRS model for asthma (AUC 0.63[25]).

Number of individuals	321,105
Number of cases (%)	57,385 (17.9)
Median age (25 & 75 percentiles)	67.4 (59.8 & 72.4)
Number of females (%)	148,704 (46.3)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.559 (0.556, 0.561)
Covariates	0.540 (0.537, 0.542)
PRS + covariates	0.570 (0.568, 0.573)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

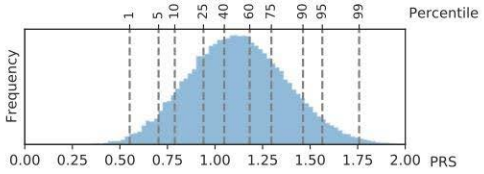

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.71 (0.69, 0.74)	0.75 (0.73, 0.78)
10-25	0.84 (0.81, 0.87)	0.86 (0.84, 0.89)
25-40	0.90 (0.87, 0.92)	0.91 (0.89, 0.94)
40-60	1.00 (0.97, 1.03)	1.00 (0.98, 1.02)
60-75	1.09 (1.06, 1.12)	1.07 (1.05, 1.10)
75-90	1.22 (1.18, 1.26)	1.18 (1.15, 1.20)
90-95	1.41 (1.35, 1.47)	1.32 (1.27, 1.36)
95-99	1.55 (1.48, 1.62)	1.41 (1.37, 1.46)
99-100	1.76 (1.62, 1.91)	1.55 (1.46, 1.65)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

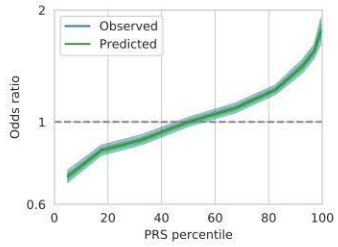

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

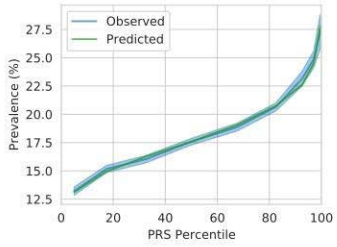

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.29 COPD

COPD is a very common disease; one in five people 55 years or older will develop the condition during their lifetime[133]. The PRS model for COPD includes 22 variants and was developed based on the variants identified in a study that analyzed genomes of ~25,000 multiethnic individuals (Europeans, African Americans, Koreans, and Hispanics) affected by COPD[134]. Heritability of COPD is ~60%[135]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.566), slightly below published PRS models for COPD (AUC 0.58 – 0.60[136]).

Number of individuals	112,543
Number of cases (%)	10,555 (9.4)
Median age (25 & 75 percentiles)	65.8 (58.5 & 71.3)
Number of females (%)	46,025 (40.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.566 (0.559, 0.571)
Covariates	0.665 (0.661, 0.670)
PRS + covariates	0.676 (0.672, 0.681)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

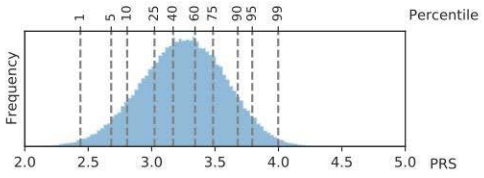

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.67 (0.62, 0.74)	0.70 (0.64, 0.75)
10-25	0.76 (0.70, 0.81)	0.77 (0.72, 0.83)
25-40	0.88 (0.82, 0.94)	0.89 (0.83, 0.95)
40-60	1.00 (0.94, 1.07)	1.00 (0.94, 1.06)
60-75	1.08 (1.01, 1.15)	1.07 (1.01, 1.14)
75-90	1.25 (1.17, 1.33)	1.22 (1.15, 1.29)
90-95	1.37 (1.25, 1.50)	1.32 (1.22, 1.43)
95-99	1.52 (1.38, 1.67)	1.45 (1.33, 1.57)
99-100	1.58 (1.33, 1.88)	1.50 (1.29, 1.74)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

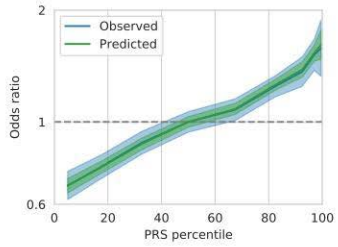

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

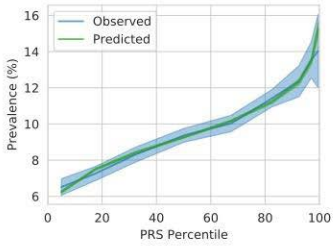

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.30 Age-Related Macular Degeneration

Age-related macular degeneration is a very common disease; one in 15 adults age 40 and older has some degree of macular degeneration[137]. The PRS model for age-related macular degeneration includes 46 variants and was developed based on the variants identified in a study that analyzed genomes of over 16,000 individuals of European ancestry affected by the condition[138]. Heritability of age-related macular degeneration is 45-70%[139] and 27.2% was explained by the variants discovered[138]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.564), significantly below the performance of a previously published PRS model for age-related macular degeneration (AUC 0.82[140]).

Number of individuals	408,885
Number of cases (%)	6,648 (1.6)
Median age (25 & 75 percentiles)	67.1 (59.5 & 72.2)
Number of females (%)	187,880 (45.9)

Table 1: Validation cohort characteristics.

Features	AUC (95% CI)
PRS	0.564 (0.557, 0.571)
Covariates	0.706 (0.700, 0.710)
PRS + covariates	0.714 (0.709, 0.720)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

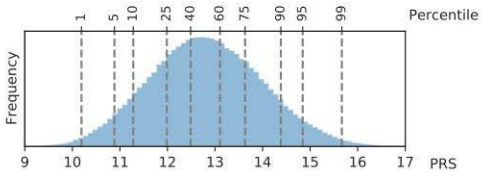

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.90 (0.81, 1.00)	0.90 (0.81, 1.00)
10-25	0.96 (0.87, 1.04)	0.96 (0.88, 1.04)
25-40	0.94 (0.86, 1.03)	0.94 (0.86, 1.03)
40-60	1.00 (0.92, 1.08)	1.00 (0.92, 1.08)
60-75	1.10 (1.01, 1.20)	1.10 (1.01, 1.20)
75-90	1.21 (1.12, 1.32)	1.21 (1.11, 1.31)
90-95	1.65 (1.49, 1.84)	1.64 (1.47, 1.82)
95-99	2.22 (2.00, 2.47)	2.18 (1.97, 2.42)
99-100	3.79 (3.27, 4.40)	3.64 (3.16, 4.20)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

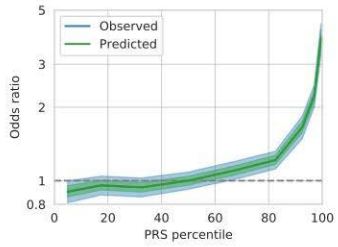

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

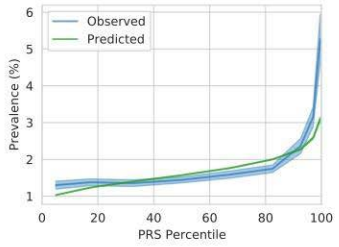

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.31 Endometriosis

Endometriosis is a very common condition; one in ten women will develop endometriosis during their reproductive years[141]. The PRS model for endometriosis includes 19 variants and was developed based on the variants identified in a study that analyzed genomes of ~15,000 women of European ancestry affected by the condition[142]. Heritability of endometriosis is ~50%[143] and 1.75% was explained by the variants discovered[142]. The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.552).

Number of individuals	221,005
Number of cases (%)	6,734 (3.0)
Median age (25 & 75 percentiles)	66.8 (59.4 & 71.9)

Table 1: Validation cohort characteristics.

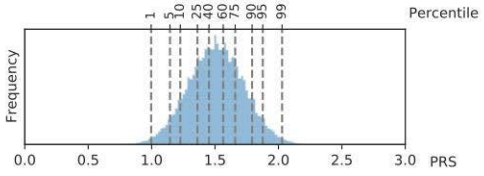

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

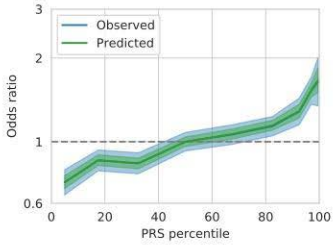

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

Features	AUC (95% CI)
PRS	0.552 (0.545, 0.558)
Covariates	0.624 (0.616, 0.630)
PRS + covariates	0.633 (0.626, 0.640)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.71 (0.64, 0.79)	0.72 (0.65, 0.80)
10-25	0.86 (0.79, 0.93)	0.86 (0.79, 0.94)
25-40	0.84 (0.77, 0.91)	0.84 (0.77, 0.91)
40-60	1.00 (0.93, 1.08)	1.00 (0.93, 1.08)
60-75	1.06 (0.98, 1.15)	1.06 (0.98, 1.15)
75-90	1.14 (1.05, 1.23)	1.13 (1.05, 1.22)
90-95	1.28 (1.15, 1.43)	1.27 (1.14, 1.42)
95-99	1.53 (1.36, 1.71)	1.50 (1.35, 1.67)
99-100	1.64 (1.34, 2.01)	1.61 (1.33, 1.95)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

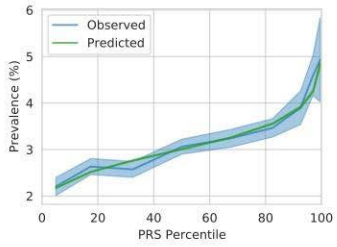

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

2.32 Uterine Fibroids

Uterine fibroids are a very common condition; one in two women will have fibroids by age 50[144]. The PRS model for uterine fibroids includes 21 variants and was developed based on the variants identified in a study that analyzed genomes of over 16,500 women of European ancestry affected by the condition[145]. Heritability of uterine 80%[146] fibroids is 80%[146] and 10.8% was explained by the variants discovered[145].

The performance of the developed PRS model was evaluated using affected and non-affected individuals from UK Biobank (see Table B for cohort definitions) and the results are presented in Tables 1-3 and Figures 1-3. The developed model demonstrated modest predictive performance (AUC 0.571).

Number of individuals	204,801
Number of cases (%)	16,092 (7.9)
Median age (25 & 75 percentiles)	66.9 (59.4 & 71.9)

Table 1: Validation cohort characteristics.

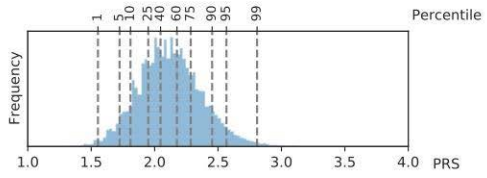

Figure 1: Distribution of the PRS in the validation cohort with select percentiles.

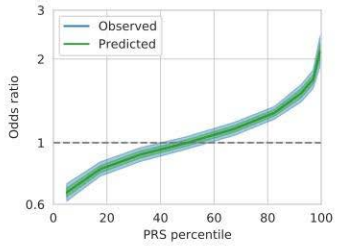

Figure 2: Odds ratios with 95% CI for each PRS percentile bin. The predicted odds ratios were calculated using logistic regression to adjust for the covariates.

Features	AUC (95% CI)
PRS	0.571 (0.567, 0.576)
Covariates	0.506 (0.501, 0.509)
PRS + covariates	0.571 (0.567, 0.576)

Table 2: The predictive performance of the genetics-only PRS model as compared to the covariates-only and the combined (PRS+covariates) models. Covariates refer to age, sex and ancestry. For "Covariates" and "PRS+covariates" logistic regression was used to calculate the AUC.

PRS bin	Odds ratio (95% CI)	Risk ratio (95% CI)
0-10	0.66 (0.62, 0.71)	0.68 (0.63, 0.73)
10-25	0.80 (0.76, 0.85)	0.82 (0.77, 0.86)
25-40	0.91 (0.86, 0.96)	0.91 (0.87, 0.96)
40-60	1.00 (0.95, 1.05)	1.00 (0.95, 1.05)
60-75	1.12 (1.06, 1.18)	1.11 (1.06, 1.17)
75-90	1.28 (1.22, 1.35)	1.25 (1.20, 1.32)
90-95	1.51 (1.40, 1.62)	1.45 (1.36, 1.55)
95-99	1.69 (1.57, 1.83)	1.61 (1.50, 1.72)
99-100	2.13 (1.88, 2.42)	1.96 (1.76, 2.19)

Table 3: Observed odds ratios and risk ratios with 95% CI for each PRS percentile bin. The 40-60 bin was defined as the reference group. Odds ratio and risk ratio (also called "relative risk") for each bin were calculated by comparing to this group.

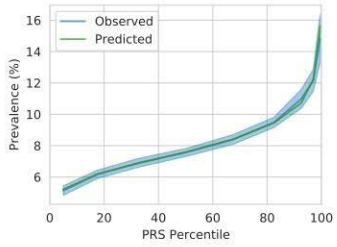

Figure 3: Prevalences with 95% CI for each PRS percentile bin. The observed prevalences were calculated by dividing the number of cases in each PRS percentile bin by the total number of individuals in the bin. The predicted prevalences were calculated using logistic regression on PRS and the covariates.

3.1 Development of PRS models

3.1.1 Variant Extraction for PRS Models

The variants included in the PRS models were extracted from genome-wide association studies (GWASs, Table A). For each disease of interest, a literature search was conducted and the GWAS articles were prioritized using the following criteria:

- Recent studies published in a respected journal by a well-established research group with a good track record (e.g., a consortium of researchers that periodically publish GWASs in a given disease area)
- Studies with well-defined cohorts, e.g., criteria to diagnose a disorder, sex, ethnicity, age, and other details of the studied individuals were provided in the article
- Studies with large sample sizes, preferably exceeding 5,000 cases
- Studies with reproducible results, e.g. discovered variants or loci were confirmed as significant by other studies
- Studies that included a step of confirming loci independence
- Studies that are accompanied by building a PRS model and evaluating its performance. Additional criteria were utilized to select articles based on the PRSs performance:
    - Studies that describe PRSs performing with higher AUC
    - Studies that describe the percentage of disease heritability that the model explains
    - Studies that calculate relative and/or absolute risk for developed models

After selecting the best GWAS article for a given disease using the criteria above, an additional set of parameters were used to select variants to be included in a PRS for a given disease. The following logic was used:

- If a PRS model was built in the same or a subsequent study, the variants comprising this PRS model were extracted.
- If no PRS was built, but explained heritability was evaluated by a specific set of variants, this set of variants was extracted.
- In cases when PRS model or explained heritability estimates were not available or variants comprising them were not clearly stated, the variants were extracted based on the p-value threshold. Specifically:
    - For studies including discovery and replication phases, discovery phase p-value should be $\leq 5x10^{8}$ and replication phase ≤ 0.05.
    - For studies including discovery and meta-analysis, p-value of the discovery phase should be $\leq 5x10^{8}$ and p-value of meta-analysis phase should be lower than that in the discovery phase.
    - For studies including meta-analysis only, p-value should be $\leq 5x10^{8}$.

After the variants and their associated weights for each model were extracted, a set of QC steps were performed to avoid human errors, inaccuracies from hg19 to hg38 liftover and other artefacts. The variant coordinates, alleles, and identifiers were verified with records of dbSNP[147], variants' expected allele frequency was verified using allele frequencies in gnomAD[148] and UK Biobank. Concordance with variant records extracted by the GWAS catalog[149] curators was also evaluated and discrepancies investigated.

As the last step of defining the variant sets for each PRS model, each variant was evaluated on whether it can be successfully detected by Simplify Genomics' CLIA-validated whole-genome sequencing assay. The variants meeting the

following criteria were further filtered out: 1) variants that are not single nucleotide variants (SNV), small insertions (≤ 7bp), or small deletions (≤ 7bp); 2) variants outside of the high-confidence region (~93.25% of whole genome) defined by Simplify Genomics based on the Illumina platinum genome for the GRCh38/hg38 reference assembly; 3) variants that are not successfully called in at least 95% of samples. The resultant sets of variants included in each PRS model is provided in the supplement to the Genomic Clinical Report[150].

3.1.2 PRS Calculation

A polygenic risk score for a given disease was calculated as the sum of the number of risk alleles multiplied by a variant-specific weight, i.e.,

$$PRS = \sum_{i \in S}^{N} w_i x_i \qquad (1)$$

where S is the set of N variants comprising the PRS model, w_i is the weight associated with variant i, and x_i is the number of risk alleles for variant i observed in the input genome.

3.1.3 PRS Percentile Calculation

A percentile is a measure indicating the value below which a given percentage of observations in a group of observations falls. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found.

For the reporting in the Genomic Clinical Report, a percentile rank was calculated for each PRS result as compared to PRS of research participants who had European ancestry. The European ancestry of research participants was determined based on the ancestry computed from the individuals' genomes and defined by the combined proportion of categories "Western European", "Eastern European" and "South European" being equal to or exceeding 60% for each included individual as suggested previously[151].

3.2 Validation of PRS Models Using UK Biobank

3.2.1 Cohorts in the Validation Dataset

The UK Biobank is a prospective cohort study with deep genetic, physical and health data collected on ~500,000 individuals across the United Kingdom from 2006-2010[21]. We use data of the UK Biobank cohort to evaluate performance of the PRS models included in the Genomic Clinical Report. Only individuals of European ancestry (as defined by code 1002 in table 22006 "Genetic ethnic grouping") were included in the validation cohort. For each disease, case ascertainment was done based on a combination of self-report in an interview with a trained nurse, hospital admission diagnosis records (ICD10 and ICD9 codes), operation procedure records (OPCS4 codes), and records of causes of death. For some diseases and phenotypes, e.g., myopia and lipid levels, specific measurements were used instead of the diagnosis codes. Unaffected individuals served as controls. Full details on the generation of the case and control cohorts for each disease are provided in Table B.

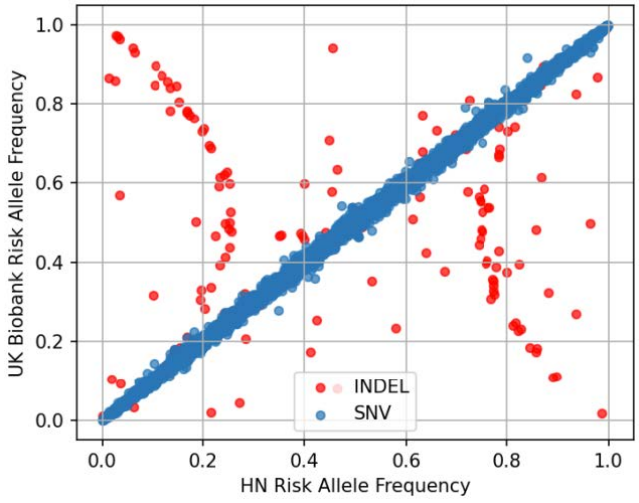

Figure 4: Risk allele frequency in the validation cohort of UK Biobank and the cohort of research participants.

3.2.2 PRS Calculation in the Validation Dataset

We evaluated 32 PRS models listed in Table A. As a validation dataset, we considered UK Biobank genetic dataset version 3, which contains imputed genomic data for 97,059,328 variants for 487,395 participants.

In order to the prepare the UK Biobank data for computing PRSs we proceeded as follows:

- Liftover. The UK Biobank dataset uses the hg19 reference genome, whereas risk alleles for PRS models were defined using the hg38 reference genome. We converted the risk alleles' distinct genomic coordinates from hg38 to hg19 using "pyliftover 0.4". All genomic coordinates were successfully mapped to hg19 with the exception of rs386000, for which we used the hg19 coordinates as reported in the original publication[93].
- Extraction and Filtering. We used HAIL 0.2 to extract UK Biobank genotypes corresponding to the PRS models' loci. Only variants for which the imputation score was greater than 0.8 were used, the remainder were imputed to allele frequencies observed in the research cohort.

After the transformation described above, the formula

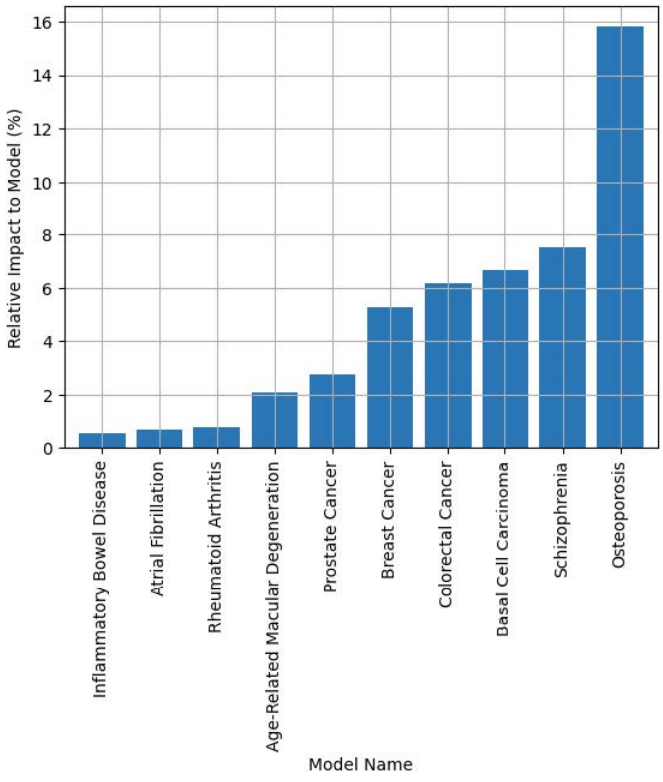

Figure 5: Contribution of indels to PRS models, as fraction of effects from indels to the effect of all model variants.

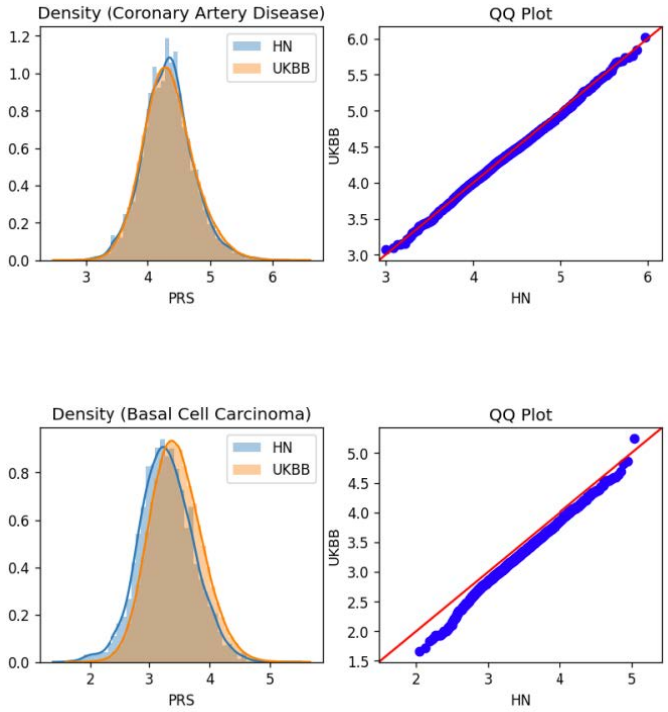

Figure 6: Comparison of PRS distribution and quantiles for basal cell carcinoma (the most discordant model) and coronary artery disease (a concordant model) in UK Biobank and research participants.

(1) was used to calculate PRSs for each disease and each participant of the UK Biobank cohort.

As UK Biobank and Simplify Genomics use different assay types, genotyping and whole-genome sequencing, respectively, we evaluated concordance of variant detection between the validation cohort (N=409,677) and the cohort of research participants (N=3,091) defined as described above. Specifically, we calculated and compared risk alleles frequencies in both cohorts. Frequencies of single nucleotide variants were concordant but frequencies of many indels (insertions or deletions) had significant discrepancies with some indels being more frequent in one cohort or another (Figure 4).

To evaluate the impact of indel detection discrepancies on the validation results, we assessed the contribution of indels to each model by calculating the ratio of sum of indel weights to the weights of all model variants (Figure 5). Osteoporosis model had the highest contribution of indels (15.8% of total weight) and its predictive performance could have been impacted the most. We compared the performance of the PRS models with and without indels in the UK Biobank cohort by calculating the AUCs (see section 3.2.3 for detail). We did not observe any significant differences in the AUCs for any model. Osteoporosis model benefited the most from including indels, but it was reflected as only 0.5% AUC increase in its discriminative power.

To further confirm concordance of the PRS distributions in the validation cohort of UK Biobank and the cohort of research participants, we performed a calibration analysis. We calculated and compared the PRS quantiles using QQplots. The UK Biobank and research PRS distributions differed at the tails (very high (>90th) and very low (<10th) percentiles) - the result of differences in indel allele frequencies between these cohorts as described above. However, the distributions did not display any major discrepancies for the rest of the PRSs (10 – 90th percentiles). The distribution for the most discordant model (with the most extreme Mann– Whitney U test's P-value), a model for basal cell carcinoma, is shown in Figure 6 for illustration.

As indels allele frequency was not concordant between the research and the validation datasets, and their exclusion did not significantly impact the discriminative power of the models, indels were removed from all models for the purposes of the validation. All metrics including AUCs and percentiles provided in this paper are presented for the PRS models with indels excluded. The Genomic Clinical Report provides results with indels included; the variants included in in each model are listed in the report supplement[150] .

3.2.3 AUC Calculation

To evaluate the discrimination performance of the PRS models, i.e., the ability of PRS to distinguish between low-risk and high-risk individuals, we calculated the area under the receiver operating characteristic curve (AUC), which is a metric measuring how far apart the PRS distributions for cases and controls are. We calculated the AUC of each RPS model using the Python package scikitlearn (function metrics.roc_auc_score). We used bootstrap resampling with replacement to generate 95% confidence intervals for AUC.

To assess the utility of PRS in predicting disease risk, we compared AUCs of PRS models to (i) models that comprise common clinically relevant covariates (age, sex and the first four principal components of ancestry), and (ii) models that combine PRS and these covariates.

Such an approach is often used to evaluate the utility of genetics-only models and the contribution of genetic factors to disease risk prediction[81,152] . As the latter two scenarios contain multiple features, we first used a logistic regression model to fit the outcome, and then used the predicted probabilities returned by the models to calculate the AUC.

3.2.4 Odds Ratio and Risk Ratio Calculation

Odds ratio and risk ratio (also known as "relative risk") are two important metrics measuring the strength of association between a risk factor and risk. Odds ratio is well-suited for case-control studies and offers a bridge to logistic regression models where covariates can be adjusted (more below), while risk ratio is easily interpretable; hence the two metrics complement each other. In case of PRS, a risk ratio shows how much increased (or decreased) risk is associated with having a certain level of PRS compared to the population average. For conditions with a small proportion of cases (e.g., < 5%) or a moderate effect of PRS (e.g., odds ratio < 3), which include most of the diseases included in this report, odds ratios give very similar values to risk ratios.

There are two ways to calculate the odds ratio: with or without adjustment of covariates. To calculate odds ratios without adjustment of covariates, we first defined a reference subset of the validation cohort to be the individuals within the 40 to 60 PRS percentile bin ("control group") as this group includes people having the average scores. Then, we compared subsets of the cohort in other percentile ranges ("carrier groups": 0 – 10th, 10 – 25th, 25 – 40th, 60 – 75th, 75 – 90th, 90 – 95th, 95 – 99th , 99 - 100th) to the reference subset. Specifically, for each comparison we tabulated the number of cases and controls in both the control and carrier groups in a 2x2 contingency table and the observed odds ratios were calculated following the standard odds ratio formula for the contingency table. The observed risk ratios were calculated in a similar fashion.

To calculate odds ratios with adjustment of covariates, we defined the control and carrier groups similarly as before. Then, we used a logistic regression model to predict the case/control outcome of individuals in the two groups, where the predictors were the binary group membership variable, and the covariates. The adjusted odds ratio is equal to the exponential of the coefficient of the binary group membership variable. Because of such different procedures, we referred to the non-adjusted odds ratios as "observed" and the adjusted odds ratios as "predicted" (Table 3 and Fig. 2 for each disease). Aside from adjusting for the potential confounding bias introduced by the covariates, the adjusted odds ratios also typically enjoy a tighter confidence interval, since the covariates explain part of the residual variance.

3.2.5 Prevalence Calculation

We also examined the risk of actually developing a disease for the individuals with a given genetic burden. For each of the PRS percentile bins as defined above (Figure 3 for each disease), we calculated the observed prevalence by dividing the number of cases by the total number of individuals in each bin. Like in case of the AUC and odds ratio, we additionally calculated the predicted prevalence as the average of the predicted probabilities of the individuals in each percentile range returned by logistic regression models with or without the covariates.

3.3 Limitations

Even though UK Biobank is the most comprehensive dataset for the evaluation of PRS models, and we have utilized high-quality current GWAS articles for PRS models, there are a number of limitations that need to be considered.

3.3.1 Limitations Affecting Model Development

- PRS models only describe the contribution of genetic background and do not account for other important contributors, such as lifestyle and environmental factors. This limitation is the main reason for the lower AUCs displayed by genetic-only models when attempting to predict disease risk.
- PRS models only describe the contribution of genetic background and do not account for other important contributors, such as lifestyle and environmental factors. This limitation is the main reason for the lower AUCs displayed by genetic-only models when attempting to predict disease risk.
- PRS models provide genetic risk information based on an assessment of specific genetic variants described in the report supplement[150]. The models do not report on the entire genetic profile and most importantly, by design, exclude rare variants that contribute to the development of monogenic forms of diseases.
- Some variants that are significant in the GWAS studies were removed from the models since they fall outside of the CLIA-validated scope of Simplify Genomics' Whole Genome Sequencing Assay. See section 3.1.1 for more detail.

3.3.2 Limitations Affecting Model Validation

- UK Biobank is affected by the "healthy volunteer effect"[153] as people who volunteer for research studies tend to be, on average, more healthconscious. Compared with the general population, UK Biobank participants were less likely to be obese, to smoke, or to drink alcohol on a daily basis and had fewer self-reported health conditions[154] . At age 70–74 years, rates of all-cause mortality and total cancer incidence were 46.2% and 11.8% lower, respectively, in men and 55.5% and 18.1% lower, respectively, in women than in the general population of the same age. As contribution of lifestyle factors to the development of the studied common diseases is significant, it is possible that due to healthier lifestyle, less participants of the dataset developed the corresponding conditions than expected in general population, thus affecting the models' performance.
- For some models (venous thromboembolism and osteoporosis) the validation data set is not independent since the original GWASs were performed using UK Biobank.
- No validation was done in other ethnicities. The PRSs developed were based on GWASs conducted based on, predominantly, individualsof European descent. The resultant models are expected to provide less accurate results for individuals of other ethnicities. However, no validation has been performed to quantify these differences.
- The variant sets for each model implemented in the Genomic Clinical Report are not exactly identical to the ones used for the validation in UK Biobank. The differences arise from exclusion of indels from PRS calculation in the UK Biobank cohort due to differences in their detection between assays utilized by UK Biobank and Simplify Genomics (see section 3.2.2 for more detail).

simplify genomics

PRS PERFORMANCE

Table A: The PRS models' details (number of variants, initial GWAS article, explained disease heritability) and performance metrics (AUC, odds ratio) accompanied by the performance metrics found in the literature for comparable models. Covariates refer to age, sex and ancestry.

Category	Disease	Number of variants in PRS model	Original GWAS	Validation Cohort Size	Number of Cases (%) in Validation Cohort	AUC (PRS)	AUC (Covariates)	AUC (PRS+ Covariates)	AUC (PRS) (Literature)	Disease Heritability (Literature)	Heritability Explained (Literature)	Odds ratio (95-99%/40-60%)	Odds ratio (Literature)
Cancer Predisposition	Basal Cell Carcinoma	29	[23]	408,885	16,391 (4.0)	0.606	0.666	0.696	- 0.63[25] - 0.64[26]	40%[24]	10.89%[23]	2.06	- 2.64 (>96%/<96%)[25]
	Bladder Cancer	12	[28]	408,885	2,758 (0.7)	0.566	0.764	0.770	-	Unknown	-	1.63	- 2.94 (>75%/<25%)[155]
	Breast Cancer	151	[32]	221,005	14,609 (6.6)	0.617	0.600	0.654	- 0.55 – 0.68[18,25,34-40]	30%[24,33]	18%[32]	2.35	- 2.32 (>90%/0-89%)[156] - 1.80 (>96%/<96%)[25]
	Colorectal Cancer	29	[42]	408,885	6,016 (1.5)	0.580	0.682	0.698	- 0.62 (women), 0.64 (men)[44]	35%[43]	11.90%[42]	1.62	- 1.69 (90-99%/25-75%)[157]
	Chronic Lymphocytic Leukemia	37	[46]	408,885	726 (0.2)	0.676	0.699	0.755	-	20%[47]	25%[46]	4.07	- 3.64 (>80%/40-60%)
	Melanoma	13	[50]	408,885	5,003 (1.2)	0.578	0.587	0.617	- 0.58[25] - 0.61[26]	60%[24]	9%[50]	1.70	-
	Pancreatic Cancer	22	[52]	408,885	939 (0.2)	0.595	0.681	0.704	- 0.61[53]	-	-	2.03	- 1.98 (>80%/40-60%)[53]
	Prostate Cancer	138	[55]	187,880	8,155 (4.3)	0.658	0.721	0.773	- 0.59 – 0.65[17,25,57-59]	60%[56]	28.4%[55]	3.32	- 3.3 (>75%/<25%)[63] - 1.58 (>75%/<25%)[25]
	Thyroid Cancer	10	[61]	408,885	619 (0.2)	0.616	0.632	0.683	-	-	-	3.20	- 3.2 (>75%/<25%)[63]
Digestive Health	Celiac Disease	56	[65, 66]	408,885	2,783 (0.7)	0.808	0.579	0.814	- 0.85[69]	75%[68]	54%[69]	16.69	- 2.5 (>80%/40-60%)[69] - 7.2 (>80%/0-20%)[69]
	Inflammatory Bowel Disease	231	[71, 72]	393,379	5,938 (1.5)	0.641	0.532	0.645	- 0.75 (CD) and 0.70 (UC)[74]	70%[73]	13% (CD) and 8.2% (UC)[71]	2.90	- 40.64 (CD) and 16.45 (UC) (>90%/<10%)[72] - 2.43 (>90%/0-90%)[156]
Heart and Vascular Health	Atrial Fibrillation	82	[76]	408,885	19,572 (4.8)	0.622	0.740	0.765	-0.64[25]	60%[77]	42%[76]	2.45	- 2.74 (>90%/0-90%)[156] - 2.81 (>90%/0-89%)[25]
	Coronary Artery Disease	41	[79]	408,885	22,387 (5.5)	0.576	0.757	0.767	- 0.65[81] - 0.62 (c-Index)[82]	50 – 60%[80]	-	1.71	- 4.17 (>80%/<20%)[82] - 2.89 (>90%/0-90%)[156]
	Venous Thromboembolism	10	[84]	407,351	14,122 (3.5)	0.600	0.616	0.655	-	40 - 50%[85]	5.24%[84]	2.43	- 4.62-4.90 (>95%/<95%)[86]
Mental Health	Schizophrenia	160	[88]	408885	1096 (0.3)	0.559	0.584	0.599	- 0.58[90,91] - 0.64[91]	80%[89]	-	1.47	- 2.17 (hazard ratio, >90%/<10%)[158]
Metabolic Health	HDL Cholesterol	311	[93]	317,466	30,610 (9.6)	0.689	0.705	0.786	-	40-60%[94,95]	12.3%[93]	3.57	-
	LDL Cholesterol	218	[93]	345,736	216,337 (62.6)	0.643	0.567	0.661	-	30-50%[94,95]	9.6%[93]	2.14	-
	Triglycerides	249	[93]	346,072	135,959 (39.3)	0.629	0.614	0.672	-	50%[97]	11%[93]	2.43	-
	Total Cholesterol	225	[93]	346,335	251,566 (72.6)	0.628	0.567	0.647	- 0.63[25]	30-50%[94,95]	8.9%[93]	2.20	- 2.54 (>96%/<96%)[25]
	Type 2 Diabetes	47	[99]	408,885	28,649 (7.0)	0.593	0.649	0.674	- 0.64[25] - 0.58 – 0.65[15]	70%[100]		1.94	- 2.49 (>90%)[156] - 2.04 (>90%)[25]
Musculoskeletal Health	Dupuytren's disease	25	[101]	408,885	4,318 (1.1)	0.681	0.737	0.790	-	80%[102]	11%[101]	4.09	1.78 (>75%/<25%)[103]
	Gout	29	[105]	408,885	8,574 (2.1)	0.641	0.775	0.807	- 0.68[25]	60%[106]	7%[105]	2.85	- 1.16 (>90%/<10%)[159] - 8.20 (>90%/<10%)[25]
	Osteoporosis	1,146	[108]	272,623	3,799 (1.4)	0.678	0.661	0.725	- 0.59[111] - 0.604[108]	60-90%[109,110]	-	3.82	-
	Rheumatoid Arthritis	94	[113]	408,885	7,844 (1.9)	0.579	0.631	0.652	-	60%[115]	3.8-4.7%[113]	2.09	-
Neurological Health	Alzheimer's Disease	20	[117]	408,885	873 (0.2)	0.696	0.787	0.830	- 0.715 – 0.745[120]	60-80%[119]	-	5.03	-
	Multiple Sclerosis	102	[122, 123]	408,885	1,871 (0.5)	0.692	0.623	0.724	- 0.70[125]	25-76%[124]	28%[122]	4.62	-
	Parkinson's Disease	46	[127]	408,885	1,762 (0.4)	0.593	0.739	0.756	- 0.61[129]	35-60%[128]	-	1.95	- 3.16 (>80%/<20%)[129]
Respiratory Health	Asthma	17	[131]	321,105	57,385 (17.9)	0.559	0.540	0.570	- 0.63[25]	35-95%[132]	3.5%[131]	1.55	- 2.71 (>96%/<96%)[25]
	COPD	22	[134]	112,543	10,555 (9.4)	0.566	0.665	0.676	- 0.58 – 0.60[136]	60%[135]	-	1.52	-
Sensory Health	Age-Related Macular Degeneration	46	[138]	408,885	6,648 (1.6)	0.564	0.706	0.714	- 0.82[140]	45-70%[139]	27.2%[138]	2.22	- 44.0 (>90%/<10%)[138]
Women's Health	Endometriosis	19	[142]	221,005	6,734 (3.0)	0.552	0.624	0.633	-	50%[143]	1.75%[142]	1.53	-
	Uterine Fibroids	21	[145]	204,801	16,092 (7.9)	0.571	0.506	0.571	-	80%	10.8%	1.69	- 1.58 – 2.94 (>75%/<25%)[160]

Table B: Case and control cohort definitions in UK Biobank. Cases: The ICD-9, ICD-10, and OPCS-4 codes were used to identify individuals in the tables: 41202 (Diagnoses - main ICD10), 41204 (Diagnoses - secondary ICD10), 40001 (Underlying (primary) cause of death: ICD10), 41203 (Diagnoses - main ICD9), 41205 (Diagnoses - secondary ICD9), 41200 (Operative procedures - main OPCS4), 41210 (Operative procedures - secondary OPCS4). When provided ICD-9, ICD-10, or OPCS-4 code have subcodes, all subcodes were included (e.g., code C50 has subcodes C50.0-C50.9 and all subcodes were included). Additionally, tables and codes specified in columns "Other" were used, primarily, 20001 (Cancer code, self-reported) and 20002 (Non-cancer illness code, self-reported). Controls: participants who did not have the codes specified for cases were included in the controls cohorts. Where ICD-10, ICD-9 and OPCS-4 codes are provided for controls, those individuals were excluded from controls.

Category	Disease	Case Cohort Definitions				Control Cohort Definitions		
		ICD-10	ICD-9	OPCS-4	Other	ICD-10 (excl.)	ICD-9 (excl.)	OPCS-4 (excl.)
Cancer Predisposition	Basal Cell Carcinoma	C44	173		20001: 1061			
	Bladder Cancer	C67	188		20001: 1035			
	Breast Cancer	C50	174		20001: 1002			
	Chronic Lymphocytic Leukemia	C911	2041		20001: 1055			
	Colorectal Cancer	C18, C19, C20	153, 154		20001: 1020, 1022, 1023			
	Melanoma	C43	172		20001: 1059			
	Pancreatic Cancer	C25	157		20001: 1026			
	Prostate Cancer	C61	185	M61	20001: 1044			
	Thyroid Cancer	C73	193		20001: 1065			
Digestive Health	Celiac Disease	K900	5790		20002: 1456, 21068: 1			
	Inflammatory Bowel Disease	K50, K51	555		20002: 1461, 1462, 1463	K52		
Heart and Vascular Health	Atrial Fibrillation	I48	4273	K571, K621, K622, K623, K624	20002: 1471			
	Coronary Artery Disease	I21, I22, I23, I241, I252	410, 4110, 412, 42979	K410, K402, K403, K404, K411, K412, K413, K414, K451, K452, K453, K454, K455, K491, K492, K498, K499, K502, K751, K752, K753, K754, K758, K759	20002: 1075, 6150: 1			
	Venous Thromboembolism	I801, I802, I822, I260, I269	4511, 4532, 4151	L791, L902	20002: 1068, 1094	I81, I820, I800, I803, I808, I809, D68	4510, 4512, 4518, 4519, 4538, 286	
Mental Health	Schizophrenia	F20, F21, F22, F213, F24, F25, F28, F29	295		20002: 1289, 20544: 2			
Metabolic Health	HDL Cholesterol	Only participants who had HDL cholesterol measurements in table 30760 were considered. People who were on a lipid-lowering medication or did not respond whether they are on a lipid-lowering medication were excluded from both cases and controls. Participants with HDL cholesterol level <= 40 mg/dL were considered cases and participants with HDL cholesterol level > 40 mg/dL were considered controls.						
	LDL Cholesterol	Only participants who had LDL cholesterol measurements in table 30780 were considered. People who were on a lipid-lowering medication or did not respond whether they are on a lipid-lowering medication were excluded from both cases and controls. Participants with LDL cholesterol level >= 130 mg/dL were considered cases and participants with LDL cholesterol level < 130 mg/dL were considered controls.						
	Total Cholesterol	Only participants who total cholesterol measurements in table 30780 were considered. People who were on a lipid-lowering medication or did not respond whether they are on a lipid-lowering medication were excluded from both cases and controls. Participants with total cholesterol level >= 200 mg/dL were considered cases and participants with total cholesterol level < 200 mg/dL were considered controls.						
	Triglycerides	Only participants who had triglycerides measurements in table 30780 were considered. People who were on a lipid-lowering medication or did not respond whether they are on a lipid-lowering medication were excluded from both cases and controls. Participants with triglycerides level >= 150 mg/dL were considered cases and participants with triglycerides level < 150 mg/dL were considered controls.						
	Type 2 Diabetes	E11			20002: 1223, 2443: 1, 6177: 3			
Musculoskeletal Health	Dupuytren's disease	M720	7286		20002: 1544			
	Gout	M10		274	20002: 1466			
	Osteoporosis	Only participants who had heel bone mineral density (BMD) T-score (fields 77, 78) measurements or responded on whether they had any fractured/broken bones in last 5 years (field 2463) prior to the baseline assessment were considered. Participants with either BMD T-score of less than -2.5 or having a fracture were considered cases and the rest were considered controls.						
	Rheumatoid Arthritis	M05, M06	7140		20002: 1464			
Neurological Health	Alzheimer's Disease	G30, F00			20002: 1263			
	Multiple Sclerosis	G35	340		20002: 1261			
	Parkinson's Disease	G20	332		20002: 1262			
Respiratory Health	Asthma	J45	493		6152: 8, 20002: 1111, 22127: 1	J30, J31, L20, L21, L22, L23, L24, L25, L26, L27, L28, L29, L30, T784, J40, J41, J42, J43, J44	477, 690, 691, 692, 698, 9953, 490, 491, 492, 496	
	COPD	Only participants who had spirometric measurements FEV1, predicted FEV1 and FVC in tables 20150, 20153 and 20151, respectively, were considered. Participants with FEV1 / predicted FEV1 ratio < 80% and FEV1 / FVC ratio < 70% were considered cases, and participants with the two ratios > 80% and > 70%, respectively, were considered controls as described in[134].						
Sensory Health	Age-Related Macular Degeneration	H353	3625		20002: 1528, 6148: 5			
Women's Health	Endometriosis	N80	6197		20002: 1402			
	Uterine Fibroids	D25	218		20002: 1351			Q092, Q07, Q08, Q09, Q16, Q17

4 References

1. A. Torkamani et al., The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 19, 581590 (2018).
2. J. L. Mega et al., Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. Lancet. 385, 22642271 (2015).
3. N. Pashayan et al. 2018. Costeffectiveness and BenefittoHarm Ratio of RiskStratified Screening for Breast Cancer: A LifeTable Model. JAMA Oncol. 4, 15041510 (2018).
4. S. K. Naber et al., CostEffectiveness of RiskStratified Colorectal Cancer Screening Based on Polygenic Risk: Current Status and Future Potential. JNCI Cancer Spectr. 4, pkz086 (2019).
5. A. Khera et al., Genetic risk, adherence to a healthy lifestyle, and coronary disease. N. Engl. J. Med. 375, 2349–2358 (2016).
6. P. Maas et al., Breast cancer risk from modifiable and nonmodifiable risk factors among white women in the United States. JAMA Oncol. 2, 1295–1302 (2016).
7. N. J. Wald et al., The illusion of polygenic disease risk prediction. Genet Med. 21, 17051707 (2019).
8. J. Elliott et al., Predictive Accuracy of a Polygenic Risk ScoreEnhanced Prediction Model vs a Clinical Risk Score for Coronary Artery Disease. JAMA. 323, 636645 (2020).
9. J.D. Mosley et al., Predictive Accuracy of a Polygenic Risk Score Compared With a Clinical Risk Score for Incident Coronary Heart Disease. JAMA. 323, 627635 (2020).
10. A. T. Kunzmann et al., Information on Genetic Variants Does Not Increase Identification of Individuals at Risk of Esophageal Adenocarcinoma Compared to Clinical Risk Factors. Gastroenterology. 156, 4345 (2019).
11. A. R. Martin et al., Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet. 51, 584591 (2019).
12. Online [https://www.genomeweb.com/moleculardiagnostics/clinicalutilitypolygenicriskscoresquestionedcommentary#.XLX8EthKhmA]. Accessed February 2020.
13. Online [https://www.businesswire.com/news/home/20190520005426/en/MyHeritageExpandsHealthLaunchesNewDNATest]. Accessed March 2020.
14. Online [https://blog.23andme.com/healthtraits/type2diabetes/]. Accessed February 2020.
15. Online [https://permalinks.23andme.com/pdf/23_19Type2Diabetes_March2019.pdf]. Accessed February 2020.
16. Online [https://blog.nebula.org/polygenicscores/]. Accessed February 2020.
17. Online [https://www.ambrygen.com/file/material/view/985/AmbryScore_Prostate_WhitePaper_0918_final.pdf]. Accessed February 2020.
18. Online [https://www.ambrygen.com/file/material/view/704/AmbryScore_0318_final.pdf]. Accessed February 2020.
19. E. Hughes et al., Development and Validation of a Combined Residual Risk Score to Predict Breast Cancer Risk in Unaffected Women Negative for Mutations on a MultiGene Hereditary Cancer Panel. Presented at SABCS (2017). Online [https://myriadmyrisk.com/scientificpresentations/]. Accessed February 2020.
20. E. Hughes et al. Development and Validation of a Polygenic Score to Predict Breast Cancer Risk in Unaffected Hispanic Women Negative for Mutations on a MultiGene Hereditary Cancer Panel. Presented at SABCS (2018). Online [https://myriadmyrisk.com/scientificpresentations/]. Accessed February 2020.
21. Online [https://www.ukbiobank.ac.uk/]. Accessed March 2020.
22. Online [http://www.americanskin.org/resource/bcc.php]. Accessed March 2020.
23. H. S. Chahal et al., Genomewide association study identifies 14 novel risk alleles associated with basal cell carcinoma. Nat Commun. 7, 12510 (2016).
24. L. A. Mucci et al., Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA. 315, 6876 (2016).
25. L. Lello et al., Genomic Prediction of 16 Complex Disease Risks Including Heart Attack, Diabetes, Breast and Prostate Cancer. Sci Rep. 9, 15286 (2019).
26. L. G. Fritsche et al., Exploring various polygenic risk scores for skin cancer in the phenomes of the Michigan genomics initiative and the UK Biobank with a visual catalog: PRSWeb. PLoS Genet. 15, e1008202 (2019).
27. Online [https://www.cancer.org/cancer/bladdercancer/about/keystatistics.html]. Accessed March 2020.
28. J. D. Figueroa et al., Genomewide association study identifies multiple loci associated with bladder cancer risk. Hum Mol Genet. 23, 13871398 (2014).

29. A. F. Kantor et al., Familial and environmental interactions in bladder cancer risk. Int. J. Cancer, 35, 703 – 706 (1985)
30. M. Burger et al., Epidemiology and risk factors of urothelial bladder cancer. Eur Urol. 63, 234241 (2013).
31. Online [https://www.breastcancer.org/symptoms/understand_bc/risk/understanding]. Accessed March 2020.
32. K. Michailidou et al., Association analysis identifies 65 new breast cancer risk loci. Nature. 551, 9294 (2017).
33. P. Lichtenstein et al., Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 343, 7885 (2000).
34. N. Mavaddat et al., Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 107, djv036 (2015).
35. G. S. Dite et al., Breast Cancer Risk Prediction Using Clinical Models and 77 Independent RiskAssociated SNPs for Women Aged Under 50 Years: Australian Breast Cancer Family Registry. Cancer Epidemiol Biomarkers Prev. 25, 359365 (2016).
36. M. E. Mealiffe et al., Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. J Natl Cancer Inst. 102, 16181627 (2010).
37. Y. Shieh et al., Breast cancer risk prediction using a clinical risk model and polygenic risk score. Breast Cancer Res Treat. 159, 513525 (2016).
38. H. Li et al., Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. Genet Med. 19, 3035 (2017).
39. S. Sawyer et al., A role for common genomic variants in the assessment of familial breast cancer. J Clin Oncol. 30, 43304336 (2012).
40. R. Allman et al., SNPs and breast cancer risk prediction for African American and Hispanic women. Breast Cancer Res Treat. 154, 583589 (2015).
41. Online [https://www.cancer.org/cancer/colonrectalcancer/about/keystatistics.html]. Accessed March 2020.
42. S. L. Schmit et al., Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst. 111, 146157 (2019).
43. N. A. Al-Tassan et al., A new GWAS and metaanalysis with 1000 Genomes imputation identifies novel risk variants for colorectal cancer. Sci Rep. 5, 10442 (2015).
44. Saunders CL et al., External validation of risk prediction models incorporating common genetic variants for incident colorectal cancer using UK Biobank. Cancer Prev Res (Phila), (2019).
45. Online [https://www.cancer.org/cancer/chronic-lymphocytic-leukemia/about/key-statistics.html]. Accessed March 2020.
46. P. J. Law et al., Genomewide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. Nat Commun. 8, 14175 (2017).
47. J. R. Cerhan et al., Familial predisposition and genetic risk factors for lymphoma. Blood. 126, 22652273 (2015).
48. G. Kleinstern et al., Association of polygenic risk score with the risk of chronic lymphocytic leukemia and monoclonal Bcell lymphocytosis. Blood. 131, 25412551 (2018).
49. Online [https://www.cancer.org/cancer/melanomaskincancer/about/keystatistics.html]. Accessed March 2020.
50. K. J. Ransohoff et al., Twostage genomewide association study identifies a novel susceptibility locus associated with melanoma. Oncotarget. 8, 1758617592 (2017).
51. Online [https://www.cancer.org/cancer/pancreaticcancer/about/keystatistics.html]. Accessed March 2020.
52. A. P. Klein et al., Genomewide metaanalysis identifies five new susceptibility loci for pancreatic cancer. Nat Commun. 9, 556 (2018).
53. M. Nakatochi et al., Prediction model for pancreatic cancer risk in the general Japanese population. PLoS One. 13, e0203386 (2018).
54. Online [https://www.cancer.org/cancer/prostatecancer/about/keystatistics.html]. Accessed March 2020.
55. F. R. Schumacher et al., Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nat Genet. 50, 928936 (2018).
56. J. B. Hjelmborg et al., The heritability of prostate cancer in the Nordic Twin Study of Cancer. Cancer Epidemiol Biomarkers Prev. 23, 23032310 (2014).
57. J. Sun, et al., Genetic score is an objective and better measurement of inherited risk of prostate cancer than family history. Eur Urol. 63, 585587 (2013).
58. C. A. Conran et al., Populationstandardized genetic risk score: the SNPbased method of choice for inherited risk assessment of prostate cancer. Asian J Androl. 18, 520524 (2016).
59. M. Aly et al., Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm1 cohort study. Eur Urol. , 2128 (2011).

60. R. M. Tuttle et al., Thyroid Carcinoma. Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 8, 1228 – 1274 (2010).
61. J. Gudmundsson et al., A genomewide association study yields five novel thyroid cancer risk loci. Nat Commun. 8, 14517 (2017).
62. K. Czene et al., Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish FamilyCancer Database. Int J Cancer. 99, 260266 (2002).
63. L. G. Fritsche et al., Association of Polygenic Risk Scores for Multiple Cancers in a Phenomewide Study: Results from The Michigan Genomics Initiative. Am J Hum Genet. 102, 10481061 (2018).
64. B. Lebwohl et al., Celiac disease and nonceliac gluten sensitivity. BMJ. 351, h4347 (2015).
65. G. Trynka et al., Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 43, 11931201 (2011).
66. P. C. Dubois et al., Multiple common variants for celiac disease influencing immune gene expression. Nat Genet. 42, 295302 (2010).
67. J. A. Tye-Din et al., Appropriate clinical use of human leukocyte antigen typing for coeliac disease: an Australasian perspective. Intern Med J. 45, 441-450 (2015).
68. R. Kuja-Halkola et al., Heritability of non-HLA genetics in coeliac disease: a population-based study in 107 000 twins. Gut. 65, 1793-1798 (2016).
69. J. Romanos et al., Improving coeliac disease risk prediction by testing non-HLA variants additional to HLA variants. Gut. 63, 415-422 (2014).
70. 70. Online [https://www.cdc.gov/ibd/data-statistics.htm]. Accessed March 2020.
71. J. Z. Liu et al., Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 47, 979986 (2015).
72. K. M. de Lange et al., Genomewide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet. 49, 256261 (2017).
73. H. Gordon et al., Heritability in inflammatory bowel disease: from the first twin study to genomewide association studies. Inflamm Bowel Dis. 21, 14281434 (2015).
74. G. B. Chen et al., Performance of risk prediction for inflammatory bowel disease based on genotyping platform and genomic risk score method. BMC Med Genet. 18, 94 (2017).
75. D. M. Lloyd-Jones et al., Lifetime risk for development of atrial fibrillation: the Framingham Heart Study. Circulation. 110, 10421046 (2004).
76. C. Roselli et al., Multiethnic genomewide association study for atrial fibrillation. Nat Genet. 50, 12251233 (2018).
77. L. C. Weng et al., Heritability of Atrial Fibrillation. Circ Cardiovasc Genet. 10, e001838 (2017).
78. D. M. LloydJones et al., Lifetime risk of developing coronary heart disease. Lancet. 353, 8992 (1999).
79. M. Nikpay et al., A comprehensive 1,000 Genomesbased genomewide association metaanalysis of coronary artery disease. Nat Genet. 47, 11211130 (2015).
80. X. Dai et al., Genetics of coronary artery disease and myocardial infarction. World J Cardiol. 8, 123
81. D. Curtis, Clinical relevance of genomewide polygenic score may be less than claimed. Ann Hum Genet. 83, 274277 (2019).
82. M. Inouye et al., Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol. 72, 18831893 (2018).
83. E. J. Bell et al., Lifetime Risk of Venous Thromboembolism in Two Cohort Studies. Am J Med. 129, 339, e1926 (2016).
84. D. Klarin et al., Genetic Analysis of Venous Thromboembolism in UK Biobank Identifies the ZFPM2 Locus and Implicates Obesity as a Causal Risk Factor. Circ Cardiovasc Genet. 10, e001643 (2017).
85. B. Zöller et al., A sibling based design to quantify genetic and shared environmental effects of venous thromboembolism in Sweden. Thromb Res. 149, 8287 (2017).
86. D. Klarin et al., Genomewide association analysis of venous thromboembolism identifies new risk loci and genetic overlap with arterial vascular disease. Nat Genet. 51, 15741579 (2019).
87. International Schizophrenia Consortium et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 460, 748752 (2009).
88. A. F. Pardiñas et al., Common schizophrenia alleles are enriched in mutationintolerant genes and in regions under strong background selection. Nat Genet. 50, 381389 (2018).
89. R. Hilker et al., Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. Biol Psychiatry. 83, 492498 (2018).

90. S. Ripke et al., Biological insights from 108 schizophreniaassociated genetic loci. Nature. 511, 421427 (2014).
91. T. Vivian-Griffiths et al., Predictive modeling of schizophrenia from genomic data: Comparison of polygenic risk score with kernel support vector machines approach. Am J Med Genet B Neuropsychiatr Genet. 180, 8085 (2019).
92. M. R. Cobain et al., Lifetime risk for developing dyslipidemia: the Framingham Offspring Study. Am J Med. 120, 623630 (2007).
93. D. Klarin et al., Genetics of blood lipids among ~300,000 multiethnic participants of the Million Veteran Program. Nat Genet. 50, 15141523 (2018).
94. S. V. de Miranda Chagas et al., Environmental factors, familial aggregation and heritability of total cholesterol, low density lipoproteincholesterol and high density lipoproteincholesterol in a Brazilian population assisted by the Family Doctor Program. Public Health. 125, 329337 (2011).
95. D. WeissglasVolkov et al., Genetic causes of high and low serum HDLcholesterol. J Lipid Res. 51, 20322057
96. Online [https://www.cdc.gov/nchs/products/databriefs/db198.htm]. Accessed March 2020.
97. H. Snieder et al., Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. Arterioscler Thromb Vasc Biol. 19, 28262834 (1999).
98. E. W. Gregg et al., Trends in lifetime risk and years of life lost due to diabetes in the USA, 19852011: a modelling study. Lancet Diabetes Endocrinol. 2, 867874 (2014).
99. R. A. Scott et al., An Expanded GenomeWide Association Study of Type 2 Diabetes in Europeans. Diabetes. 66, 28882902 (2017).
100. G. Willemsen et al., The Concordance and Heritability of Type 2 Diabetes in 34,166 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN) Consortium. Twin Res Hum Genet. 18, 762-771 (2015).
101. M. Ng et al., A Genomewide Association Study of Dupuytren Disease Reveals 17 Additional Variants Implicated in Fibrosis. Am J Hum Genet. 101, 417427 (2017).
102. S. Larsen et al., Genetic and environmental influences in Dupuytren's disease: a study of 30,330 Danish twin pairs. J Hand Surg Eur Vol. 40, 171176 (2015).
103. S. A. Riesmeijer et al., A Weighted Genetic Risk Score Predicts Surgical Recurrence Independent of HighRisk Clinical Features in Dupuytren's Disease. Plast Reconstr Surg. 143, 512518 (2019).
104. P. Richette et al., Gout. Lancet. 375, 318328 (2010).
105. A. Köttgen et al., Genomewide association analyses identify 18 new loci associated with serum urate concentrations. Nat Genet. 45, 145154 (2013).
106. T. R. Merriman et al., An update on the genetic architecture of hyperuricemia and gout. Arthritis Res Ther. 17, 98 (2015).
107. Online [https://www.iofbonehealth.org/factsstatistics]. Accessed March 2020.
108. S. K. Kim, Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. PLoS One. 13, e0200785 (2018).
109. S. H. Ralston et al., Genetics of osteoporosis. Endocr Rev. 31, 629662 (2010).
110. . E. L. Duncan et al., Clinical review 2: Genetic determinants of bone density and fracture risk--state of the art and future directions. J Clin Endocrinol Metab. 95, 25762587 (2010).
111. K. Estrada et al., Genomewide metaanalysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. Nat Genet. 44, 491501 (2012).
112. C. S. Crowson et al., The lifetime risk of adultonset rheumatoid arthritis and other inflammatory autoimmune rheumatic diseases. Arthritis Rheum. 63, 633639 (2011).
113. Y. Okada et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 506, 376381 (2014).
114. W. X. Liu et al., HLADRB1 shared epitope allele polymorphisms and rheumatoid arthritis: a systemic review and metaanalysis. Clin Invest Med. 39, E182E203 (2016).
115. . A. J. MacGregor et al., Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum. 43, 3037 (2000).
116. Online [https://www.alz.org/alzheimersdementia/factsfigures]. Accessed March 2020.
117. J. C. Lambert et al., Metaanalysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 45, 14521458 (2013).
118. E. Genin et al. APOE and Alzheimer Disease: A Major Gene with SemiDominant Inheritance. Mol Psychiatry. 16, 903–907 (2011).
119. C. Van Cauwenberghe et al., The genetic landscape of Alzheimer disease: clinical implications and perspectives. Genet Med. 18, 421430 (2016).

120. V. Escott-Price et al., Common polygenic variation enhances risk prediction for Alzheimer's disease. Brain. 138, 36733684 (2015).
121. Online [https://www.nationalmssociety.org/WhatisMS/WhoGetsMS]. Accessed March 2020.
122. International Multiple Sclerosis Genetics Consortium (IMSGC) et al., Analysis of immunerelated loci identifies 48 new susceptibility variants for multiple sclerosis. Nat Genet. 45, 13531360 (2013).
123. L. Moutsianas et al., Class II HLA interactions modulate genetic risk for multiple sclerosis. Nat Genet. 47, 11071113 (2015).
124. C. H. Hawkes et al., Twin studies and the heritability of MS: a conclusion. Mult Scler. 15, 661667 (2009).
125. P. L. De Jager et al., Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. Lancet Neurol. 8, 11111119 (2009).
126. Online [https://www.sciencedirect.com/topics/medicineanddentistry/parkinsonsdisease]. Accessed March 2020.
127. D. Chang et al., A metaanalysis of genomewide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet. 49, 15111516 (2017).
128. S. Pierce et al., Parkinson's diseaseassociated genetic variation is linked to quantitative expression of inflammatory genes. PLoS One. 12, e0175882 (2017).
129. M. A. Nalls et al., Largescale metaanalysis of genomewide association data identifies six new risk loci for Parkinson's disease. Nat Genet. 46, 989993 (2014).
130. Online [http://www.aafa.org/asthmafacts/]. Accessed March 2020.
131. F. Demenais et al., Multiancestry association study identifies new asthma risk loci that colocalize with immunecell enhancer marks. Nat Genet. 50, 4253 (2018).
132. C. Ober et al., The genetics of asthma and allergic disease: a 21st century perspective. Immunol Rev. 2011 Jul;242(1):1030.
133. Y. M. van Durme et al., Prevalence, incidence, and lifetime risk for the development of COPD in the elderly: the Rotterdam study. Chest. 135, 368377 (2009).
134. B. D. Hobbs et al., Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. Nat Genet. 49, 426432 (2017).
135. T. Ingebrigtsen et al., Genetic influences on Chronic Obstructive Pulmonary Disease - a twin study. Respir Med. 104, 18901895 (2010).
136. R. Busch et al., Genetic Association and Risk Scores in a Chronic Obstructive Pulmonary Disease Metaanalysis of 16,707 Subjects. Am J Respir Cell Mol Biol. 57, 3546 (2017).
137. Online [https://www.allaboutvision.com/conditions/amd.htm]. Accessed March 2020.
138. L. G. Fritsche et al., A large genomewide association study of agerelated macular degeneration highlights contributions of rare and common variants. Nat Genet. 48, 134143 (2016).
139. J. M. Seddon et al., The US twin study of agerelated macular degeneration: relative roles of genetic and environmental influences. Arch Ophthalmol. 123, 321327 (2005).
140. F. Grassmann et al., Modelling the genetic risk in agerelated macular degeneration. PLoS One. 7, e37979 (2012).
141. Online [http://endometriosis.org/resources/articles/factsaboutendometriosis/]. Accessed March 2020.
142. Y. Sapkota et al., Metaanalysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. Nat Commun. 8, 15539 (2017).
143. S. A. Treloar et al., Genetic influences on endometriosis in an Australian twin sample. Fertil Steril. 71, 701710 (1999).
144. Online [https://medlineplus.gov/ency/article/000914.htm]. Accessed March 2020.
145. T. Rafnar et al., Variants associating with uterine leiomyoma highlight genetic background shared by various cancers and hormonerelated traits. Nat Commun. 9, 3636 (2018).
146. N. Christacos et al, Glob. libr. women's med., (ISSN: 17562228); DOI 10.3843/GLOWM.10364. (2009).
147. S. T. Sherry et al., dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29, 308311 (2001).
148. K. J. Karczewski et al., Variation across 141,456 human exomes and genomes reveals the spectrum of lossoffunction intolerance across human proteincoding genes. In bioRxiv (p. 531210). https://doi.org/10.1101/531210. (2019).
149. A. Buniello et al., The NHGRIEBI GWAS Catalog of published genomewide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005D1012 (2019).
150. Online [https://www.humanlongevity.com/HNGenomicClinicalGeneList50.pdf]. Accessed March 2020.
151. L. Dumitrescu et al., Assessing the accuracy of observerreported ancestry in a biorepository linked to electronic medical records.

Genet Med. 12, 648–650 (2010).
152. Online [https://www.phgfoundation.org/documents/prsreportfinalweb.pdf]. Accessed March 2020.
153. M. DelgadoRodríguez et al., Bias. J Epidemiol Community Health. 58, 635641 (2004).
154. A. Fry et al., Comparison of Sociodemographic and HealthRelated Characteristics of UK Biobank Participants With Those of the General Population. Am J Epidemiol. 186, 10261034 (2017).
155. M. GarciaClosas et al., Common genetic polymorphisms modify the effect of smoking on absolute risk of bladder cancer. Cancer Res. 73, 22112220 (2013).
156. A. V. Khera et al., Genomewide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 50, 12191224 (2018).
157. S. L. Schmit et al., Novel Common Genetic Susceptibility Loci for Colorectal Cancer. J Natl Cancer Inst. 111, 146157 (2019).
158. H. T. Horsdal et al., Association of Childhood Exposure to Nitrogen Dioxide and Polygenic Risk Score for Schizophrenia With the Risk of Developing Schizophrenia. JAMA Netw Open. 2, e1914401 (2019).
159. T. G. Richardson et al., An atlas of polygenic risk score associations to highlight putative causal relationships across the human phenome. Elife. 8, e43657 (2019).
160. N. Välimäki et al., Genetic predisposition to uterine leiomyoma is determined by loci for genitourinary development and genome stability. Elife. 7, e37110 (2018).