A polygenic risk score-based analysis of gene-drug interactions in relation to type 2 diabetes risk

Jacob Russell Armitage

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Genetic Epidemiology

University of Washington
2023

Committee:
Luke Wander
Sara Lindström
Daniel Enquobahrie

Program Authorized to Offer Degree:
Institute for Public Health Genetics

University of Washington

Abstract

A polygenic risk score-based analysis of gene-drug interactions in relation to type 2 diabetes risk

Jacob Armitage

Committee Chair

Dr. Luke Wander

Statin usage is associated with a higher risk of developing type 2 diabetes (T2D); however, statins have been shown to be safe and effective at lowering the risk of cardiovascular disease and are a widely used first line therapy. Therefore, it is critical to understand the underlying mechanisms of the association between statin usage and T2D, including gene-environment interactions. This study's primary objective was to examine interactions between a T2D polygenic risk score (PRS) and statin usage on T2D risk. A secondary objective was to investigate interactions between statin usage and sex on T2D risk. This study utilized data from n=360,989 participants of the UK Biobank (UKB) without prevalent diabetes at baseline. Multiple logistic regression models were fit to test for multiplicative interaction between statin usage at baseline and measures of T2D PRS while controlling for various other covariates. Evidence for the interaction between statin usage at baseline and standard T2D PRS was not found ($p = 0.125$, and $p = 0.481$ for the PRS treated as a decile and standardized continuous value, respectively). When the data is subset by genetic sex the association of statin usage with T2D risk for females is (OR 1.86, 95% C.I 1.68, 2.06) and for males (OR 1.69, 95% C.I 1.58, 1.82) with the interaction estimate for statin usage and genetic sex (male) having $p = .050$. Future research on interactions between statin usage and genetic variants should examine interactions of statin usage with pathway- or cluster-specific T2D- related genetic variants such as variants contained in partitioned polygenic risk scores, while possible differential sex-based effects of statins on T2D risk may warrant fitting separate models dichotomized by sex.

Introduction

Diabetes is a major contributor to health care spending, costing over $327 billion in 2017 in United States, including direct costs and losses to work productivity[1]. The incidence and prevalence of diabetes continue to rise at startling rates across all age-groups and by 2050, 33% of all U.S. adults could have diabetes, predominantly type 2 diabetes (T2D). The projected percent of the those in the USA with diagnosed diabetes is expected to be 13.9% by 2030 with prevalence predicted to increase across all race-sex groups[2].

Statins, a class of lipid-lowering medications that is widely used among populations at risk of T2D, reduce illness and mortality in those who are at high risk of cardiovascular disease by lowering levels of LDL cholesterol. Statins exert this effect primarily through the inhibition of cholesterol synthesis in the liver but also by increasing the rate of LDL cholesterol uptake from the blood[3]. Statins also have pleiotropic effects that are still under investigation. These effects include potential differential gene expression related to various inflammatory processes[4]. Statins are not without adverse reactions, and it has been observed that statin use is associated with a higher risk of incident T2D [5–8]. A meta-analysis of 17 RCT occurring between 1994 and 2012 showed that atorvastatin at 80mg/day compared to placebo increased the risk of T2D by 15 percent [5], another meta-analysis of 170,255 patients across 76 RCTs found that 17 of these trials reported an increased risk for incident T2D with an average odds ratio of 1.09, with the other 59 trials showing no increased risk [6]. Observational studies have also found a higher risk of incident T2D in those taking statins [7,8]. These cohort studies also found evidence that statin use is inversely associated with insulin sensitivity and secretion.

Several mechanisms are hypothesized to underlie this association, including β-cell dysfunction via the disruption of β-cell cholesterol homeostasis[9]. Since statins have been shown to be safe and effective at lowering the risk of cardiovascular disease, they will remain a widely used first line therapy for many people, therefore it is clinically relevant to identify subgroups, including groups defined by genetic susceptibility, who are at higher risk of incident statin associated T2D so both preventive and therapeutic actions can be taken.

Over the past decade many GWAS studies have examined genetic risk variants for T2D. So far, the genetic variants that have been associated with the risk of T2D in these studies have almost all been common alleles (MAF > 5 percent) with relatively small individual effect sizes,

in total explaining a small amount of heritability with no strong evidence that rare variants play a major role in this risk [10].

Many of the existing gene-environment interaction studies for statins have been done on a few candidate genes and not at the GWAS level. This is likely because GWAS already suffers from reduced statistical power due to the amount of independent hypothesis tests, and adding gene-environmental interactions to these studies further lowers power to detect small effect sizes. While it may be practical to utilize a hypothesis-driven candidate gene approach to maintain the statistical power to identify small effect sizes, this approach has been criticized due to lack of reproducibility and other factors [11]. More recently gene-environment interaction analyses have used polygenic risk scores (PRS). These quantitative disease risk scores are generated by the summation of genetic variants weighted by their estimated effect size derived from GWAS data. In UK Biobank genotype data, individuals in the upper percentile for a PRS for T2D have a similar increased risk as those with rare monogenic mutations (96.5 percentile for T2D PRS had a greater than 3-fold increase and those within the 99th percentile with a 3.3-fold risk) [12]. Current PRS models for T2D perform well when they include age, sex, and other covariates in their models and can be useful tools to predict disease susceptibility [13].

Many previous studies using UKB data have investigated interactions of a PRS with environmental and drug exposures. Environmental exposure-PRS interaction studies in the UKB include a study on the interaction of alcohol consumption*Alzheimer disease PRS on brain cortical thickness [14] and cigarette smoking*forced expiratory volume PRS on reduced lung function [15]. Drug exposure-PRS interaction studies in the UKB include a study examining interactions of a PRS for coronary artery disease with alirocumab treatment on coronary artery disease [16] and a study examining interactions of a T2D PRS with sulfonylurea treatment on glycemic traits and various pharmacodynamic end points[17].

There is some evidence that women have a greater risk for statin-associated T2D due to differences in average body mass between males and females resulting in a greater effective dose among females [18]. Additionally, evidence exists in mice models that sex may impact the cellular response to statins[19].

For this analysis, we examined interaction of the PRS for T2D with statin usage on risk of incident T2D. We hypothesized that the association of statin use with incident T2D might depend on individual's PRS for T2D. If so, the diabetogenic effects of statins would be enhanced

synergistically by a high PRS, placing an upper percentile of individuals at a greater risk of T2D than what would be expected based on their PRS and statin usage alone. Secondary aims include testing for the presence of effect modification of the diabetogenic effect of statins by genetic sex on a multiplicative scale. Findings from this study can inform future mechanistic studies as well as clinical practice.

Methods

Study Setting and Study Population

The UK Biobank is a prospective cohort study with genetic and phenotypic data collected on approximately 500,000 individuals from across the United Kingdom, aged between 40 and 69 at enrollment. Enrollment took place for 4 years starting in 2006 and participants are to be followed for 30 years [20]. Data collected during at an initial visit to an assessment center included surveys on health history and medication usage, socio-demographic information, anthropometric measurements, and biological samples.

Exclusion Criteria

Individuals with prevalent diabetes defined by the presence of one International Classification of Diseases, Tenth Revision (ICD-10) code E10, E11, E13, or E14 with a first occurrence date prior to or within one year after enrollment were excluded from the study. These codes do not include diabetes occurring during pregnancy, childbirth, and the puerperium as these diabetes types may indicate that the individual is at a higher risk for future T2D. Nor do they include malnutrition-related diabetes. To eliminate cryptic relatedness, kinship values between all pairs of individuals were calculated, and one participant from each pair was removed for those with at least a third-degree relatedness. In the event of a case–control pair the control was removed to preserve as many cases in the dataset as possible. In the case of a control-control pair the removal of one of the individuals was random. As a quality control measure those with missing genetic sex and genetic sex not matching with self-reported sex were removed. Those of self-reported non-British ethnicity were removed to optimize the predictive performance of the PRS in the model and to reduce genetic population structure. After these exclusions, there remained a final analytic sample of n=360,982 individuals.

Exposures:

Statin usage at baseline was obtained via a touchscreen questionnaire and a verbal interview with a trained nurse-examiner. Baseline statin usage is utilized as a binary indicator and broken down by type. For the T2D PRS, we utilized the UK Biobank's PRS release which includes optimized PRS scores for 53 diseases and various quantitate traits [21]. This PRS release has been shown to outperform various other published PRS scores and allows for increased power in analyzing the UKB cohort. This release contains both a "standard" and "enhanced" PRS with the former being available for almost all cohort members and the latter being available for a subset of 104,231 participants. The enhanced set has been trained on external data as well as additional data from a subset of UKB cohort members increasing the performance of the PRS set. In the evaluation of the performance of the PRS scores it was found that those with high-risk scores were conferred a similar increase in risk as carriers of single high-risk variants, however PRS performed worse in those of non-European ancestry due to the low diversity of the cohort. The performance of the Enhanced PRS for T2D is shown in Figure 1. in which one standard deviation in the log of the PRS is associated with an odds ratio of ~2 for those of European ancestry. For this analysis, we examined the performance of the PRS in two ways (standard PRS as continuous deciles and a standardized continuous variable).

Primary outcome:

The first occurrences data category contains generated fields to indicate first occurrences for a set of diagnostic codes covering a wide range of health outcomes sourced from self-report, primary care, hospital inpatient data and death data, mapped to ICD-10 codes; the primary outcome of interest is type 2 diabetes defined by ICD-10 code E11 contained in this data category.

Covariates:

Covariates to control confounding included a measure of alcohol consumption, smoking pack years, body fat percentage by impedance, age at recruitment and genetic sex. Those with "NA" for smoking pack years represents individuals who started and stopped smoking by age 16.

Statistical analysis

We examined distributions of baseline characteristics overall, stratified by statin usage at baseline, and stratified by whether T2D developed during the follow-up interval. We fit logistic regression models examining associations of the T2D PRS with incident T2D overall and

stratified by statin usage. We fit logistic regression models examining associations of statin usage with incident T2D overall and stratified by T2D PRS. In models including all participants, we tested for the presence of first-order multiplicative interactions between statin usage and PRS. We tested for the presence of a multiplicative interaction between statin usage and genetic sex. We also fit models stratified by genetic sex to estimate sex-specific associations of statin usage with T2D incidence.

Analyses were done using R version 4.2.2[22]. All hypothesis tests were two-sided and statistical significance level of .05 was utilized. Logistic regression models were fitted using R's GLM package. To reduce multicollinearity and make estimates more interpretable continuous control variables (bodyfat percentage, age at recruitment and smoking pack years) were centered and scaled using R's scale function. The first 10 genetic principal components were used to adjust for genetic population structure in the models. The effects of age and bodyfat percentage were allowed to vary with genetic sex via interaction terms giving the models more flexibility. Our null hypothesis was that the PRS-statin interaction terms would have an odds ratio of ~1, signifying that the diabetogenic effect of statins is not modified by T2D PRS. In the second aim our null hypothesis is that the genetic sex-statin interaction terms will have an odds ratio of ~1, signifying that the diabetogenic effect of statins is not modified by genetic sex. We estimated the probability of T2D for subgroups of sex, statin usage, and enhanced PRS decile using the underlying fitted logistic model using the *ggeffects* package[23].

Results

After filtering and exclusion, 360,989 participates remained; overall 54.7% of remaining participates were female with average participant age at recruitment 56.7 years (Table 1, Table 2) For those who at some point in the study became a T2D case, the mean time-duration between enrollment into the study and the first reported date of T2D was 394 weeks for those not using statins at baseline and 377 weeks for those using statins (Table 3). A standard T2D PRS score of 1 standard deviation was associated with a higher risk of T2D (OR 1.74, 95% C.I 1.68,1.79) and an increase by one decile of standard T2D PRS was associated with a higher risk of T2D (OR 1.21, 95% C.I 1.20, 1.22, Table 4). The interaction estimate between standard T2D PRS as a continuous decile and statin use at baseline had p value = .12 the interaction estimate for the standard T2D PRS as a standardized continuous value had p value = .48. Table 5 shows statin

usage at baseline was associated with an increased risk of T2D (OR 1.76, 95% C.I 1.66, 1.86) with statin subtype OR estimates differing slightly. When the data is subset by genetic sex (Table 6) the association of statin use with T2D risk for females is (OR 1.86, 95% C.I 1.68, 2.06) and for males (OR 1.69, 95% C.I 1.58, 1.82). The interaction estimate for statin use at baseline and genetic sex (male) had p = .0503. Both the enhanced and standard PRS distributions when dichotomized by T2D status showed expected rightward shifts for the positive T2D status grouping (Figure 2 and 4), this shift was not present when dichotomized by statin usage at baseline (Figure 3 and 5). Predicted probabilities for new-onset T2D grouped by genetic sex, statin usage and enhanced PRS decile show an expected increase in risk with statin use, male genetic sex, and increased enhanced PRS decile (Figure 6 and 7).

Discussion

Overall, we did not find evidence that statin associated T2D risk depended on an individual's genetic risk for T2D, as captured in the PRS. We also found that statin usage and T2D associations was slightly stronger in women compared to men. While the diabetogenic effect of statins may be worse in women, men were at a higher overall risk for T2D. To our knowledge this is the first time that a T2D PRS-statin interaction has been investigated in a large cohort.

There are several mechanisms that might explain why we did not find an interaction between the PRS and statin treatment at baseline in this analysis. First, this study may lack the statistical power to detect such an interaction due to various possible factors including but not limited to; the misclassification of T2D outcome, misclassification of statin usage and not considering time to event information.

Another possibility is that the lack of interaction between the PRS and statin usage indicate that genetic variants that interact with statins to cause T2D are either not contained in the T2D PRS or their effect is not detectable because they only make up small fraction of the PRS score. The diabetogenic effect of statins may be mediated in part by variants that create new mRNA target sites and/or those involved in statin metabolism. Such variants are not likely to be contained in a T2D PRS as they do not directly interact with T2D pathogenesis without the presence of statins, making them unlikely to be picked up in a marginal GWAS of T2D. The prevailing mechanistic hypotheses for the diabetogenic effect of statins involve pathways related

to beta cell dysfunction, with evidence for impaired glucose sensing and insulin secretion as well as disruptions to intracellular cholesterol homeostasis [9,9]. Therefore, it may still be advantageous to attempt to focus on genetic variants in areas of the genome related to the regulation and protein coding regions impacting these pathways to search for variants that interact with statin usage to raise NOD risk. Data from this study could be utilized for this task by imputing haplotypes based on SNP array results for UKB participants and fitting models with interaction terms between statins and individual variants or with a version of a genetic variant burden score.

In the UK Biobank population men across 4 ethnicities had a higher prevalence of T2D after standardizing for age, socioeconomic factors, BMI and lifestyle factors, with significant sex-based differences in white (men 6.0% vs. women 3.6%), South Asian (21.0% vs. 13.8%) and Black (13.3% vs. 9.7%) [25]. We see this higher prevalence among men partially reflected in the T2D probability graphs in figures 6 and 7, with men having a higher predicted probability of being an incident T2D case. However, even with an overall higher risk for T2D in males we still found evidence that the diabetogenic effect of statins is greater in females. Aside from the previously mentioned bodyweight differences between males and females possibly creating higher effective dosages for females, there may be differences in the cellular response to statin exposure. A study in mice demonstrated that both chromosomal and gonadal sex may impact the cellular response to statin treatment, particularly with changes to gene regulation in response to statin treatment in XY mice via the upregulation of cholesterol biosynthetic gene expression[19]. These changes to gene expression and subsequently cellular responses to statins require further investigation with human studies. Some research suggests that risk models for T2D be fit separately by sex, as the prediction power of common parameters (e.g., waist circumference) is higher in females [24]. This may support the usage of separate interaction models being fit for genetic sex data subsets for testing interactions between statins and measures of genetic risk factors.

This analysis has some strengths, including utilizing a very large sample size, polygenic risk scores with good predictive performance and detailed information on participants. There are also some limitations. Using the first occurrences data category to identify T2D cases will result in some amount of misclassification, likely in which not all T2D cases are captured. The usage of baseline covariates as control variables assumes consistent values relative to other participants over time. There is an assumption that participants using statins at baseline will not discontinue

their treatment in the future and that the number of participants who begin statin treatment after recruitment will be comparatively small. Removal of prevalent cases of T2D from the study sample may create sampling bias in which participants at a higher genetic risk for T2D are removed at a higher rate. A Cox proportional hazards model would likely perform better than a logistic regression model as time from exposure to outcome is accounted for and as such it would be advisable for subsequent investigations to explore the usage of this methodology. Death records exist in the UKB database for around 45,000 people, this study does not remove those who have died over the course of observation. Interaction is defined on the multiplicative scale but could also be investigated on the additive as this scale may have more relevance from a public health perspective but is more complex to model. The UK Biobank is not a representative sample of the general British population; being on average healthier, wealthier, and overrepresenting both females and white British people. Participants were also less likely to have high levels of smoking and drinking exposure as well as fewer self-reported health conditions[26]. While differential socioeconomic conditions in a sample population can have their own effects on the generalizability of disease-exposure associates, the lack of genetic diversity in the UKB cohort has its own consequences for the generalizability of genetic association studies. Those of European ancestry have been historically overrepresented in genetic studies, leading to the possible worsening of health disparities in nonwhite populations[27]. PRS have been found to have lower predictive power in diverse populations, especially in those of African ancestry[28]. The performance of PRS for usage across ancestries can be improved through the usage of diverse training data and should be pursued [29].

In summary, we did not find evidence that statin associated T2D risk depended on the PRS. Research is needed to define subpopulations who are at higher risk for developing statin associated T2D, as well as clarifying which individual genetic variants should be more closely examined to uncover the mechanisms involved. Future studies that investigate statin-gene interactions should try to utilize a more diverse and representative sample population to reduce the potential for future precision medicine-based interventions into this subject from possibly worsening health disparities.

Figures and Tables

Table 1. Baseline characteristics of UK Biobank participants of self-reported British ethnicity without diabetes at baseline, stratified by statin usage at baseline.

	No Statin Usage (N=314701)	Yes Statin Usage (N=46281)	Overall (N=360982)
Genetic Sex			
Female	179476 (57.0%)	17960 (38.8%)	197436 (54.7%)
Male	135225 (43.0%)	28321 (61.2%)	163546 (45.3%)
Age at Recruitment (years)			
Mean (SD)	55.9 (8.00)	61.8 (5.78)	56.7 (7.99)
Median [Min, Max]	57.0 [38.0, 73.0]	63.0 [40.0, 72.0]	58.0 [38.0, 73.0]
Smoking Pack Years			
Mean (SD)	21.6 (17.4)	28.5 (21.0)	22.8 (18.2)
Median [Min, Max]	17.5 [0, 301]	24.3 [0, 212]	18.5 [0, 301]
Missing	225812 (71.8%)	27937 (60.4%)	253749 (70.3%)
Reported Daily Drinking at Baseline			
Mean (SD)	0.212 (0.409)	0.237 (0.425)	0.215 (0.411)
Median [Min, Max]	0 [0, 1.00]	0 [0, 1.00]	0 [0, 1.00]
Bodyfat Percentage by Impedance			
Mean (SD)	31.2 (8.55)	31.9 (8.14)	31.3 (8.50)
Median [Min, Max]	30.9 [5.00, 69.8]	30.8 [5.40, 65.2]	30.9 [5.00, 69.8]
Missing	4851 (1.5%)	1313 (2.8%)	6164 (1.7%)
Developed T2D during follow-up			
Negative	303350 (96.4%)	41424 (89.5%)	344774 (95.5%)
Positive	11351 (3.6%)	4857 (10.5%)	16208 (4.5%)

Table 2. Baseline characteristics of UK Biobank participants of self-reported British ethnicity without diabetes at baseline, stratified by incident T2D over follow-up.

	T2D Negative (N=344781)	T2D Positive (N=16208)
Genetic Sex		
Female	190482 (55.2%)	6958 (42.9%)
Male	154299 (44.8%)	9250 (57.1%)
Age at Recruitment (years)		
Mean (SD)	56.6 (8.00)	59.3 (7.23)
Median [Min, Max]	58.0 [38.0, 73.0]	61.0 [40.0, 70.0]
Smoking Pack Years		
Mean (SD)	22.2 (17.8)	31.1 (22.3)
Median [Min, Max]	18.0 [0, 301]	27.0 [0, 255]
Missing	244427 (70.9%)	9327 (57.5%)
Reported Daily Drinking		
No	269788 (78.2%)	13534 (83.5%)
Yes	74993 (21.8%)	2674 (16.5%)
Bodyfat Percentage by Impedance		
Mean (SD)	31.1 (8.47)	34.8 (8.37)
Median [Min, Max]	30.7 [5.00, 69.8]	33.9 [5.00, 68.0]
Missing	5708 (1.7%)	456 (2.8%)
Statin Usage at Baseline		
No	303356 (88.0%)	11351 (70.0%)
Yes	41425 (12.0%)	4857 (30.0%)
Statin Type		
atorvastatin	7431 (2.2%)	1061 (6.5%)
none	303356 (88.0%)	11351 (70.0%)
other	3011 (0.9%)	408 (2.5%)
simvastatin	30983 (9.0%)	3388 (20.9%)

Table 3. Time from assessment center visit to incident T2D stratified by statin usage at baseline among those of self-reported British ethnicity UK Biobank participants who were free from diabetes at baseline.

Number of Incident Cases	Statin Exposure at baseline	Mean Time Duration (weeks)
11351	No	394.46
4857	Yes	377.24

Table 4. Associations of standard T2D PRS with incident T2D, overall and stratified by statin usage at baseline among those of self-reported British ethnicity UK Biobank participants who were free from diabetes at baseline.

	Association of PRS with incident diabetes among participants without statin use at baseline**			Association of PRS with incident diabetes for all participants*			Association of PRS with incident diabetes among participants with statin use at baseline**			Interaction between statin usage at baseline and PRS for all participants*
	OR	95% C.I	p value	OR	95% C.I	p value	OR	95% C.I	p value	p value
PRS Decile	1.180	1.16, 1.20	1.23E-83	1.21	1.20, 1.22	2.72E-222	1.21	1.20, 1.23	5.72E-223	0.12
PRS Continuous	1.670	1.59, 1.75	1.09E-94	1.74	1.68, 1.79	4.70E-246	1.75	1.69, 1.81	6.57E-247	0.48

*Adjusting for age, genetic sex, daily alcohol consumption, smoking pack years, genetic pcs, and baseline statin usage

**Adjusting for age, genetic sex, daily alcohol consumption, smoking pack years and genetic pcs

Abbreviations: OR (odds ratio), C.I (Confidence Interval)

Table 5. Associations of statin usage at baseline with incident T2D, overall and stratified by PRS grouping among those of self-reported British ethnicity UK Biobank participants who were free from diabetes at baseline.

	Assocation between statin usage at baseline and incident diabetes for all participants*			Assocation between statin usage at baseline and incident diabetes for those of 1 SD or greater than 1 SD below the mean**			Assocation between statin usage at baseline and incident diabetes for those of 1 SD or greater than 1 SD above the mean**		
	OR	95% C.I	p value	OR	95% C.I	p value	OR	95% C.I	p value
Any statin	1.76	1.66, 1.86	4.03E-79	1.80	1.40, 2.29	2.10E-06	1.80	1.60, 2.00	5.95E-26
Atorvastatin	2.07	1.86, 2.29	5.01E-42	2.15	1.39, 3.21	3.26E-04	2.03	1.65, 2.48	9.33E-12
Simvastatin	1.63	1.52, 1.73	1.28E-46	1.63	1.22, 2.14	5.75E-04	1.70	1.50, 1.91	1.64E-17
Other	2.30	1.95, 2.70	9.71E-24	2.58	1.31, 4.61	2.83E-03	2.26	1.62, 3.09	6.35E-07

*Adjusting for age, genetic sex, daily alcohol consumption, smoking pack years, genetic pcs and standardized continuous PRS

**Adjusting for age, genetic sex, daily alcohol consumption, smoking pack years and genetic pcs

Abbreviations: OR (odds ratio), SD (standard deviation), C.I (Confidence Interval)

Table 6. Associations of statin usage with incident T2D stratified by genetic sex among those of self-reported British ethnicity UK Biobank participants who were free from diabetes at baseline.

	Assocation between statin usage at baseline and incident diabetes for those with male genetic sex**			Assocation between statin usage at baseline and incident diabetes for those with female genetic sex**			Interaction between statin usage at baseline and genetic sex*		
	OR	95% C.I	p value	OR	95% C.I	p value	OR	95% C.I	p value
Statin at usage baseline	1.69	1.58, 1.82	1.37E-47	1.86	1.68, 2.06	9.60E-33	0.88	0.78, 1.00	5.03E-02

*Adjusting for age, genetic sex, daily alcohol consumption, smoking pack years, genetic pcs and standardized continuous PRS

**Adjusting for age, daily alcohol consumption, smoking pack years, standardized continuous PRS and genetic pcs

Abbreviations: OR (odds ratio), C.I (Confidence Interval)

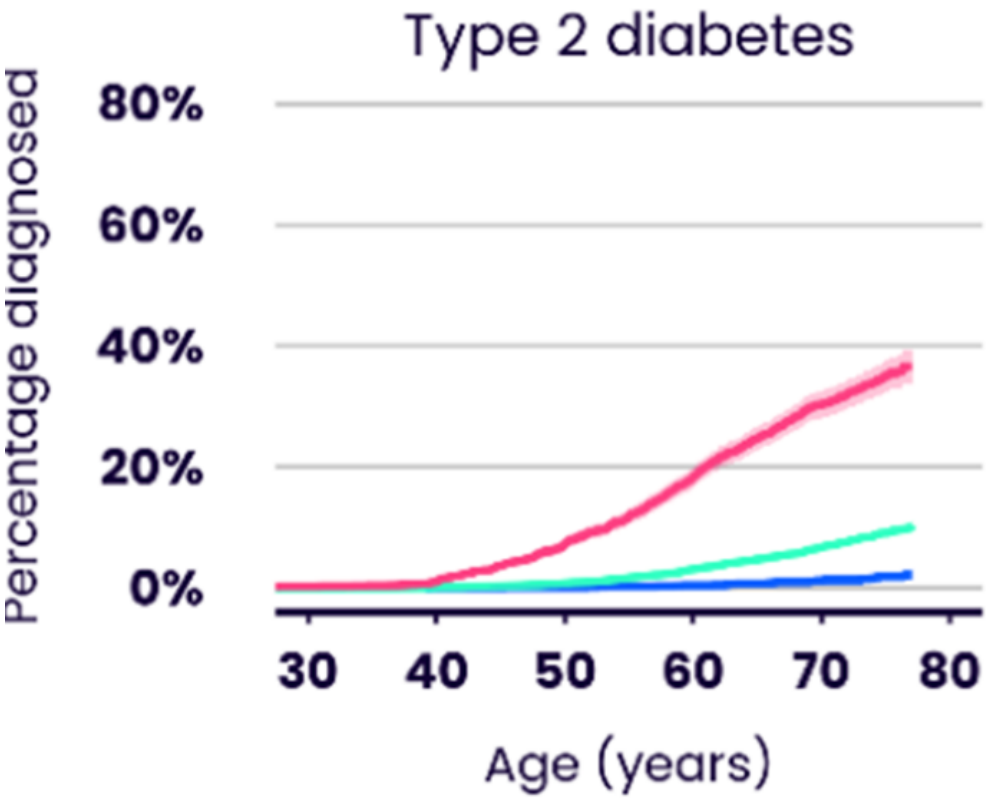

Figure 1. Cumulative incidence of T2D by Enhanced PRS group in UKB; Red: highest 3%, Green: median 40%-60%, Blue: lowest 3%. Odds ratio for 1 Sd of log(PRS) is ~2 for those of European ancestry. Adopted from "Thompson, D. J. et al. UK Biobank release and systematic evaluation of optimized polygenic risk scores for 53 diseases and quantitative traits. 2022.06.16.22276246 Preprint at https://doi.org/10.1101/2022.06.16.22276246 (2022)."

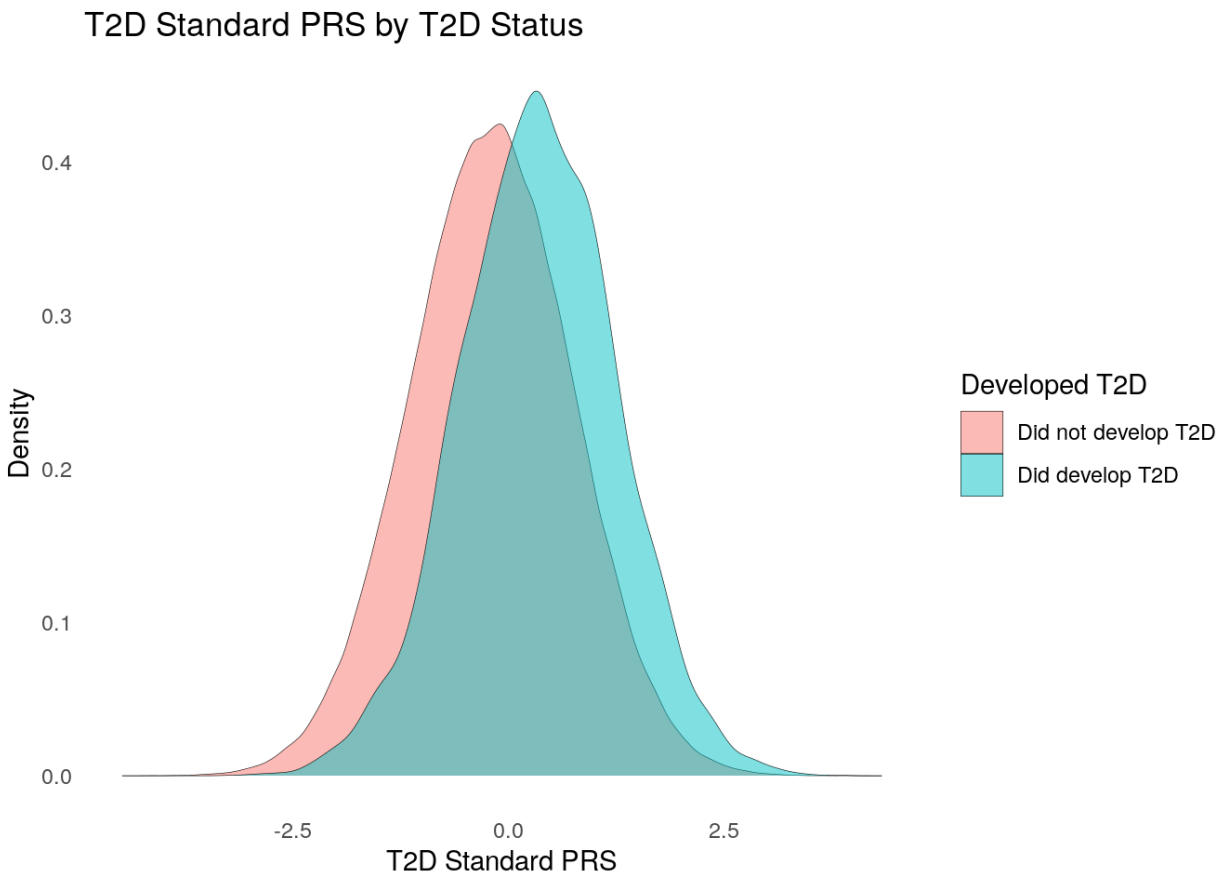

Figure 2. T2D Standard Polygenic Risk Score distribution among UK Biobank participants of self-reported British ethnicity who were free from diabetes at baseline dichotomized by development of T2D.

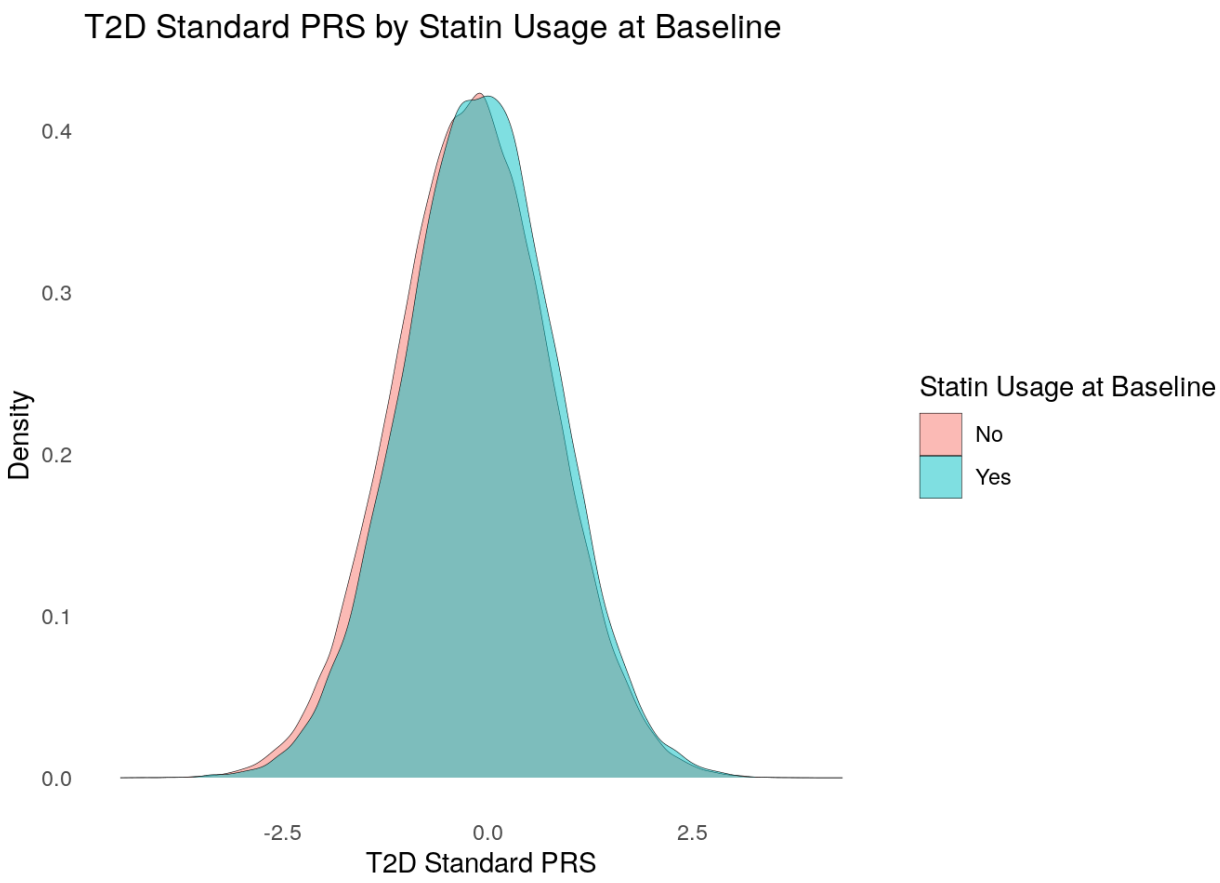

Figure 3. T2D Standard Polygenic Risk Score distribution among UK Biobank participants of self-reported British ethnicity who were free from diabetes at baseline dichotomized by statin usage at baseline.

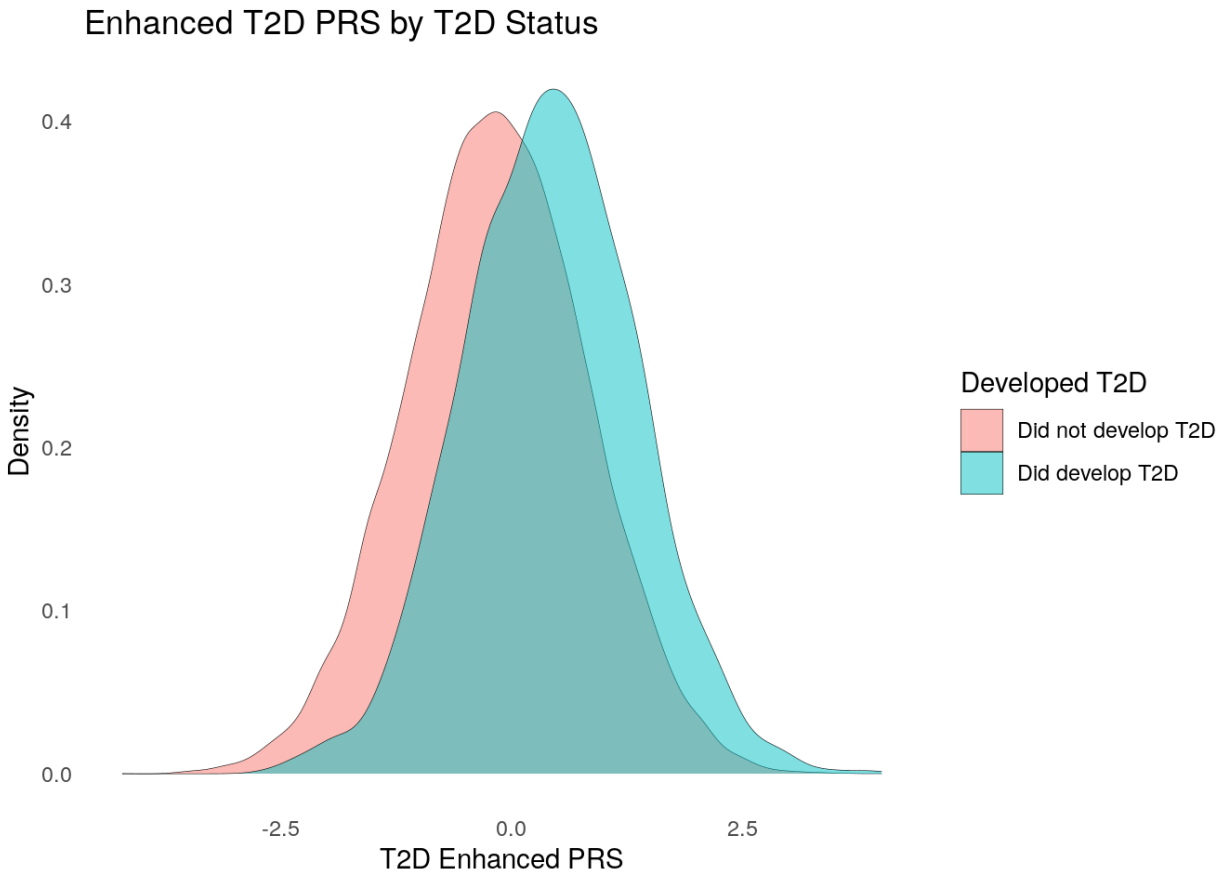

Figure 4. T2D Enhanced Polygenic Risk Score distribution among UK Biobank participants of self-reported British ethnicity who were free from diabetes at baseline dichotomized by development of T2D.

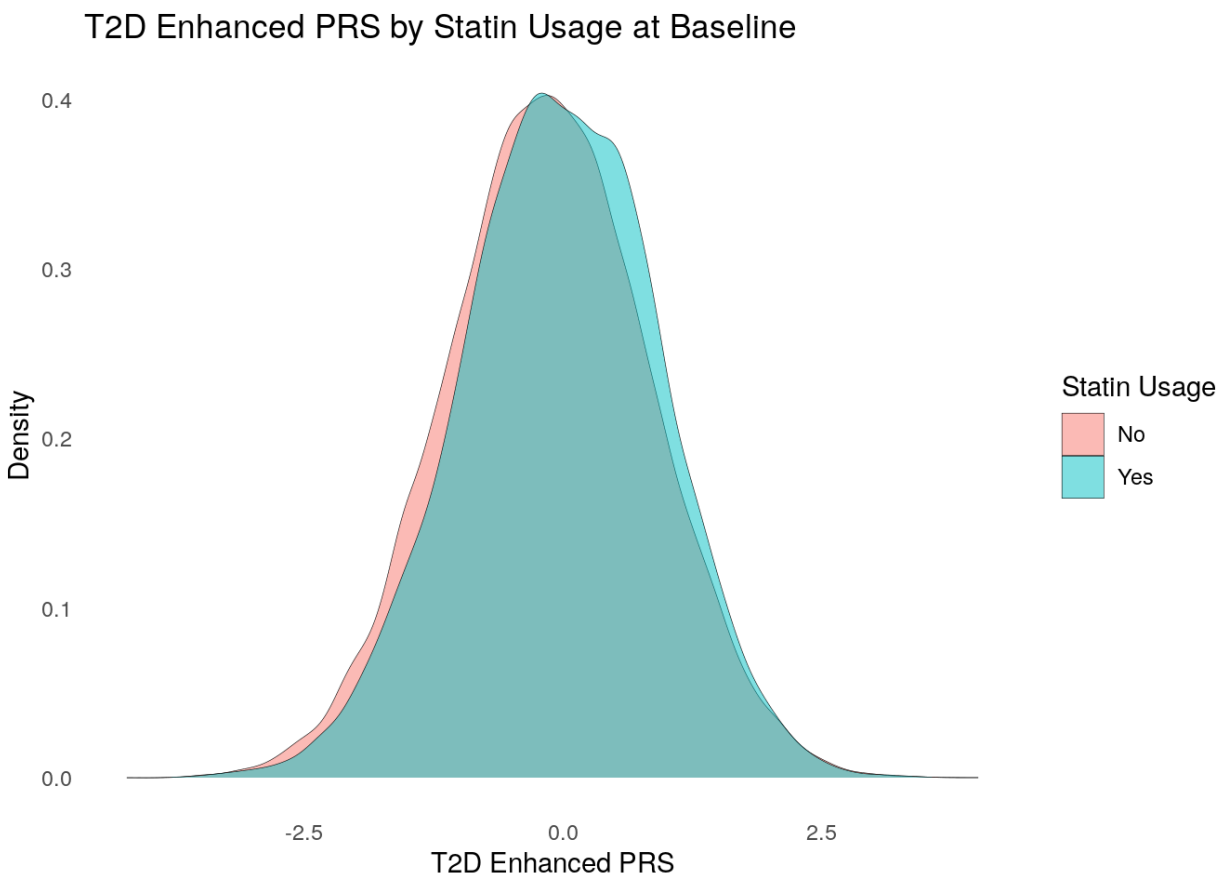

Figure 5. T2D Enhanced Polygenic Risk Score distribution among UK Biobank participants of self-reported British ethnicity who were free from diabetes at baseline dichotomized by statin usage at baseline.

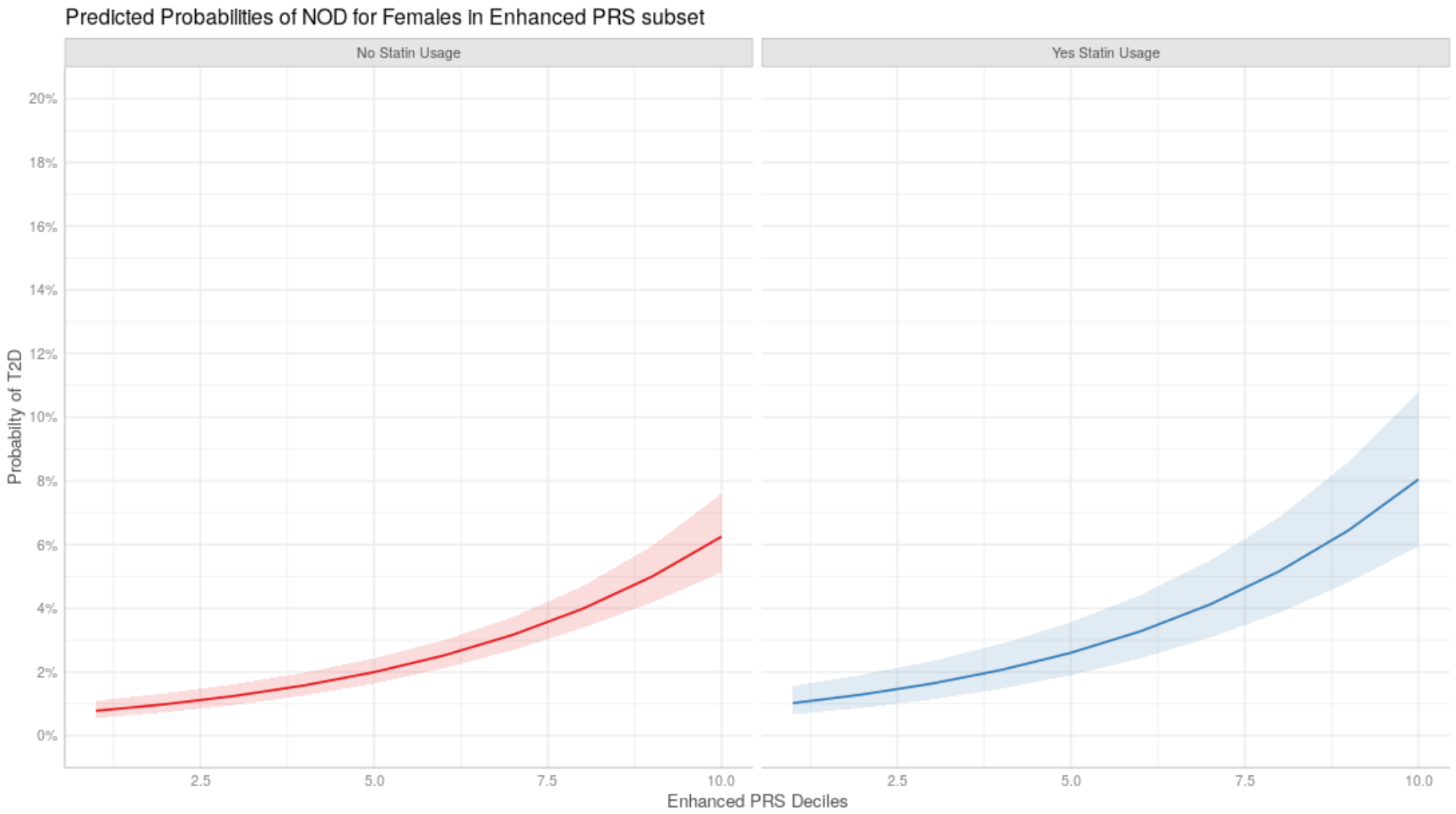

Figure 6. Predicted probabilities for incident T2D for women in UK Biobank free from diabetes at baseline contained in enhanced PRS subset, by Enhanced PRS deciles with 95% C.Is. dichotomized by statin exposure reported at baseline, other predictors ~ bodyfat percentage and smoking pack years are set at their mean value for the enhanced female data subset.

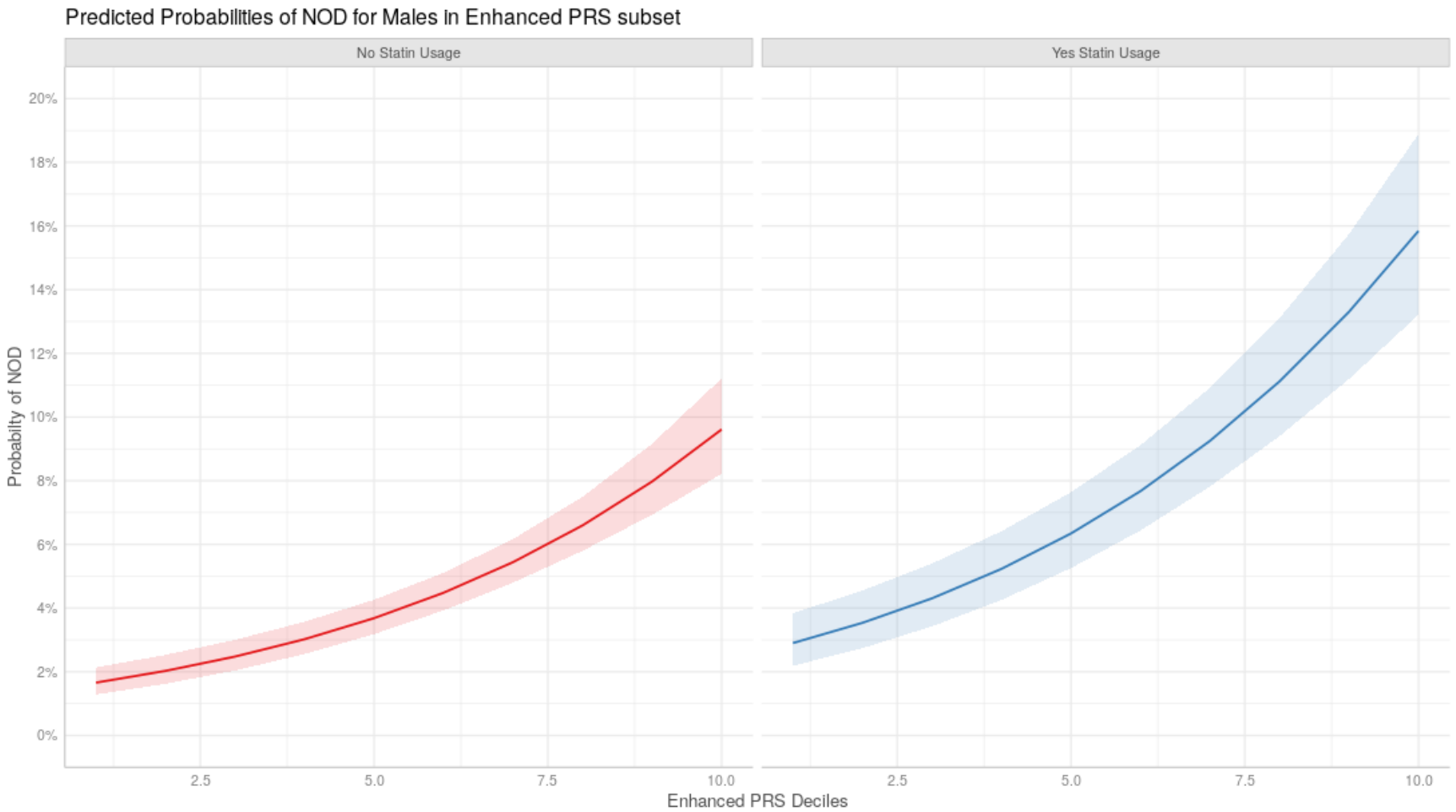

Figure 7. Predicted probabilities for NOD for males UK Biobank free from diabetes at baseline contained in enhanced PRS subset, plotted against Enhanced PRS deciles with 95% C.Is. dichotomized by statin exposure at baseline, other predictors ~ bodyfat percentage and smoking pack years are set at their mean value for the enhanced male data subset.

Supplemental Information 1. UKB Data Fields Used

Variable Name	**Type**	**UKB Code**
Age at Recruitment	Recorded at Baseline	21022
Body Fat Percentage	Recorded at Baseline	23099
Alcohol Intake Frequency	Recorded at Baseline	1558
Pack Years Smoking	Recorded at Baseline	20161
Genetic sex	Genetically Derived	22001
Genetic principal components	Genetically Derived	22009
Calculated Standard Polygenic Risk Score	Genetically Derived	26286
Medication at Baseline	Recorded at Baseline	20003

Drug	Code
Atorvastatin	1141146234
Rosuvastatin	1141192410
Simvastatin	1140861958
Fluvastatin	1140888594
Pravastatin	1140888648

References

1. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care*. 2018;41(5):917-928. doi:10.2337/dci18-0007

2. Lin J, Thompson TJ, Cheng YJ, et al. Projection of the future diabetes burden in the United States through 2060. *Population Health Metrics*. 2018;16(1):9. doi:10.1186/s12963-018-0166-4

3. Stancu C, Sima A. Statins: mechanism of action and effects. *J Cell Mol Med*. 2001;5(4):378-387. doi:10.1111/j.1582-4934.2001.tb00172.x

4. Liao JK, Laufs U. PLEIOTROPIC EFFECTS OF STATINS. *Annu Rev Pharmacol Toxicol*. 2005;45:89-118. doi:10.1146/annurev.pharmtox.45.120403.095748

5. Navarese EP, Buffon A, Andreotti F, et al. Meta-Analysis of Impact of Different Types and Doses of Statins on New-Onset Diabetes Mellitus. *American Journal of Cardiology*. 2013;111(8):1123-1130. doi:10.1016/j.amjcard.2012.12.037

6. Mills EJ, Wu P, Chong G, et al. Efficacy and safety of statin treatment for cardiovascular disease: a network meta-analysis of 170,255 patients from 76 randomized trials. *QJM*. 2011;104(2):109-124. doi:10.1093/qjmed/hcq165

7. Cederberg H, Stančáková A, Yaluri N, Modi S, Kuusisto J, Laakso M. Increased risk of diabetes with statin treatment is associated with impaired insulin sensitivity and insulin secretion: a 6 year follow-up study of the METSIM cohort. *Diabetologia*. 2015;58(5):1109-1117. doi:10.1007/s00125-015-3528-5

8. Jones M, Tett S, Peeters GMEE, Mishra GD, Dobson A. New-Onset Diabetes After Statin Exposure in Elderly Women: The Australian Longitudinal Study on Women's Health. *Drugs Aging*. 2017;34(3):203-209. doi:10.1007/s40266-017-0435-0

9. Carmena R, Betteridge DJ. Diabetogenic Action of Statins: Mechanisms. *Curr Atheroscler Rep*. 2019;21(6):23. doi:10.1007/s11883-019-0780-z

10. Fuchsberger C, Flannick J, Teslovich TM, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41-47. doi:10.1038/nature18642

11. Zhang X, Belsky J. Three phases of Gene × Environment interaction research: Theoretical assumptions underlying gene selection. *Development and Psychopathology*. 2022;34(1):295-306. doi:10.1017/S0954579420000966

12. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z

13. Liu W, Zhuang Z, Wang W, Huang T, Liu Z. An Improved Genome-Wide Polygenic Score Model for Predicting the Risk of Type 2 Diabetes. *Frontiers in Genetics*. 2021;12. Accessed June 5, 2022. https://www.frontiersin.org/article/10.3389/fgene.2021.632385

14. Matloff WJ, Zhao L, Ning K, Conti DV, Toga AW. Interaction effect of alcohol consumption and Alzheimer disease polygenic risk score on the brain cortical thickness of cognitively normal subjects. *Alcohol*. 2020;85:1-12. doi:10.1016/j.alcohol.2019.11.002

15. Kim W, Moll M, Qiao D, et al. Interaction of Cigarette Smoking and Polygenic Risk Score on Reduced Lung Function. *JAMA Netw Open*. 2021;4(12):e2139525. doi:10.1001/jamanetworkopen.2021.39525

16. Damask A, Steg PG, Schwartz GG, et al. Patients With High Genome-Wide Polygenic Risk Scores for Coronary Artery Disease May Receive Greater Clinical Benefit From Alirocumab Treatment in the ODYSSEY OUTCOMES Trial. *Circulation*. 2020;141(8):624-636. doi:10.1161/CIRCULATIONAHA.119.044434

17. Li JH, Szczerbinski L, Dawed AY, et al. A Polygenic Score for Type 2 Diabetes Risk Is Associated With Both the Acute and Sustained Response to Sulfonylureas. *Diabetes*. 2020;70(1):293-300. doi:10.2337/db20-0530

18. Goodarzi MO, Li X, Krauss RM, Rotter JI, Chen YDI. Relationship of Sex to Diabetes Risk in Statin Trials. *Diabetes Care*. 2013;36(7):e100-e101. doi:10.2337/dc13-0490

19. Wiese CB, Agle ZW, Zhang P, Reue K. Chromosomal and gonadal sex drive sex differences in lipids and hepatic gene expression in response to hypercholesterolemia and statin treatment. *Biol Sex Differ*. 2022;13(1):63. doi:10.1186/s13293-022-00474-8

20. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. doi:10.1038/s41586-018-0579-z

21. Thompson DJ, Wells D, Selzam S, et al. UK Biobank release and systematic evaluation of optimised polygenic risk scores for 53 diseases and quantitative traits. Published online June 16, 2022:2022.06.16.22276246. doi:10.1101/2022.06.16.22276246

22. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing. Published online 2021. https://www.R-project.org/

23. Lüdecke D. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*. 2018;3(26):772. doi:10.21105/joss.00772

24. Li T, Quan H, Zhang H, et al. Type 2 diabetes is more predictable in women than men by multiple anthropometric and biochemical measures. *Sci Rep*. 2021;11(1):6062. doi:10.1038/s41598-021-85581-z

25. Ferguson LD, Ntuk UE, Celis-Morales C, et al. Men across a range of ethnicities have a higher prevalence of diabetes: findings from a cross-sectional study of 500 000 UK Biobank participants. *Diabet Med*. 2018;35(2):270-276. doi:10.1111/dme.13551

26. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017;186(9):1026-1034. doi:10.1093/aje/kwx246

27. Petrovski S, Goldstein DB. Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biology*. 2016;17(1):157. doi:10.1186/s13059-016-1016-y

28. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*. 2019;10(1):3328. doi:10.1038/s41467-019-11112-0

29. Wang Y, Namba S, Lopera E, et al. Global Biobank analyses provide lessons for developing polygenic risk scores across diverse cohorts. *Cell Genomics*. 2023;3(1):100241. doi:10.1016/j.xgen.2022.100241