

Do these genes make me look diabetic? Using a polygenic risk score to predict age at diabetes onset in the Health and Retirement Study

Carmen D. Ng, Hubert Department of Global Health at Emory University

carmen.ng@emory.edu

Jordan Weiss, University of Pennsylvania, Graduate Group in Demography

jordanwe@sas.upenn.edu

BACKGROUND

Diabetes is the seventh leading cause of death in the United States (US) and has been implicated in the etiology of several other leading causes of death as well (1). Diabetes was responsible for more than 250,000 deaths in 2015 (2) and, in 2012, imposed an economic burden of approximately \$245 billion stemming from direct medical costs and loss of productivity (3). The expected rise in diabetes prevalence among the US adult population, from 14% in 2010 to an estimated 21% in 2050 (4), will impose even greater burdens on the nation's economic and healthcare systems, as well as patients and their families.

Age, smoking behavior, body mass index (BMI), and levels of physical activity have all been implicated as risk factors for diabetes (5). Family history and genetic variants have also been linked to increased diabetes risk (6), but it has been suggested that their influence on diabetes is greatest for middle-aged individuals between the ages of 35 and 60 (7), plausibly suggesting an increased importance of behavioral or lifestyle characteristics in later life for

diabetes onset at older ages. Thus, exploring how modifiable risk factors and genetic risk influence diabetes onset in later life may aid in our understanding of the progression of diabetes as well as the utility of targeting specific modifiable risk factors for intervention among individuals who vary in their genetic predisposition.

Two prior studies, one using the Framingham Offspring Study (8) and one using cohorts of Swedish and Finnish subjects (9), found that genetic makeup plays a modest but significant role in predicting new cases of diabetes, even after accounting for common risk factors. These studies highlight the utility of incorporating a genetic component into analyses looking at the associations between risk factors and diabetes onset. However, these studies were each limited to fewer than 20 single nucleotide polymorphisms (SNPs). Increased collection of genetic material over the past decade has led to advances in genome-wide association studies (GWASs) and the construction of polygenic risk scores (PGSs) which can elucidate a better understanding of the genetic risk for diabetes.

Our aim in this study was to examine the effects of genetic risk of diabetes and later-life behavioral and lifestyle characteristics associated with diabetes. We used data from the national population-based and longitudinal US Health and Retirement Study (HRS). We conducted a time-to-event analysis with time-varying covariates to better understand how genetic endowment combines with changing behavioral characteristics to shape risk of diabetes onset for non-Hispanic whites and non-Hispanic blacks. We hypothesized that a higher genetic predisposition for diabetes would be associated with a higher risk of diabetes onset for both race groups. Furthermore, we analyzed which behavioral and lifestyle characteristics would still have statistically significant relationships with diabetes onset even after controlling for the genetic component, as those with persisting significant associations may be the most critical in terms of

clinical recommendations for diabetes management.

METHODS

Study population

The HRS is a nationally representative and longitudinal study that has biennially assessed the financial, physical, and mental well-being of community-dwelling adults aged 50+ and their spouses since 1992. Since the conception of the HRS, new participants have been added to the survey. The HRS is sponsored by the National Institute on Aging (NIA U01AG009740) and is conducted by the University of Michigan (10).

From 2006 to 2012, the HRS collected genetic data from a sub-sample of non-Hispanic white and non-Hispanic black respondents who consented and provided salivary deoxyribonucleic acid (DNA). Details on the sample selection and consent procedures are available elsewhere (11). We restricted our analysis to the non-Hispanic white and black respondents with available genetic information, and followed these respondents from 1992 to 2014. We linked the HRS data files compiled by RAND Corporation (12) with the HRS genetic data containing a PGS for diabetes (11). Descriptions of the assay and calculation procedures are detailed elsewhere (11).

Measures

Outcome. Incident diabetes was determined by a respondent's affirmative response to the question: "Since we last talked to you, that is since [last interview date], has a doctor ever told you that you have diabetes or high blood sugar?" Age was censored for individuals who did not

report diabetes by the last wave in 2014 or who died without ever reporting diabetes.

Exposure. GWASs have identified a large number of genetic variants, typically SNPs, associated with a wide range of health outcomes and behaviors. However, the majority of these variants have a small effect and typically correspond to a small fraction of truly associated variants, meaning that they have limited predictive power. A PGS aggregates and weights this information into a single measure linked to a phenotype of interest (13). We used a PGS for diabetes which was constructed by the HRS researchers using the GWAS meta-analysis conducted by Morris and colleagues (14). The GWASs in the meta-analysis used to estimate SNP weights were derived from analyses based on European ancestry groups; thus, the predictive power of the PGSs for African ancestry groups may vary (13, 15). To account for this, PGSs were standardized by the HRS for each ethnicity to a standard normal curve (mean = 0, standard deviation [SD] = 1). This PGS z-score allows for a simple interpretation – a one SD increase in the PGS versus the change of one risk allele within a race group. In our primary analysis, PGS was included as a continuous standardized score. We also performed sensitivity analyses with the PGS as a dichotomous variable ($z\text{-score} < 0$, $z\text{-score} \geq 0$) and as a categorical variable splitting the PGS into tertiles.

Covariates. We selected covariates based on their anticipated association with diabetes. Sociodemographic covariates included sex (male, female), race (non-Hispanic white, non-Hispanic black), foreign born (yes, no), level of education (less than high school, high school/GED, some college, college or above), and partnership status (married/partnered, not married/partnered). Measures of economic well-being included employment status (employed,

unemployed, retired, disabled, not in labor force), household income (log-transformed),¹ household wealth (log-transformed),² and whether the respondent had Medicare (yes, no), Medicaid (yes, no) or another form of health insurance (yes, no). We assessed behavioral and lifestyle characteristics by including respondent's self-report of BMI (continuous), exercise (waves 1-6: report of vigorous activity at least three times per week; waves 7-12: report of vigorous activity more than once per week), smoking status (never smoker, current smoker, former smoker), and alcohol consumption (report of consuming 3+ alcoholic drinks on days they drank). Extreme values of BMI ($BMI < 10$, $BMI > 75$), were recoded as missing values. We also included self-reported binary indicators of whether the respondent had been diagnosed between waves with high blood pressure, cardiovascular disease, and arthritis, which are important health comorbidities for diabetes (16-18). For the purpose of this analysis, our main interest was in the behavioral and lifestyle variables, and how they were modified with the inclusion of our exposure. By adjusting for all these sociodemographic covariates, measures of economic well-being, and health comorbidities, we attained better estimates of our behavioral and lifestyle variables.

Additionally, we adjusted for birth cohort to account for the structured sampling design of the HRS which introduces new birth cohorts approximately every six years. We also included ancestry-specific principal components to account for possible confounding from population stratification and possible ancestry differences in genetic makeup that could bias estimates, as

¹ Household income is the sum of all income in a household, including respondent's and spouse's income from wages, pension and annuity, social security, disability, and retirement, unemployment and workers compensation, other government income, as well as household capital and other income.

² Household wealth is the net value of total household wealth including primary residence, other real estate, transportation, businesses, stocks and bonds, checking and savings accounts, bonds, total mortgage, other home loans, debt, and individual retirement accounts.

recommended in the literature (11, 15). See Ware et al. (11) for detailed information on the construction of the ancestry-specific principal components. Their estimates are not displayed in our tables for brevity.

Statistical analysis

Our analytic sample consisted of 15,190 respondents, of which 12,090 were non-Hispanic white and 3,100 were non-Hispanic black. Over the course of the study period, this resulted in 103,059 person-years of follow-up.

Kaplan-Meir survival curves and multivariate Cox regression models (19) were used to estimate the contribution of the diabetes PGS to diabetes onset after adjusting for time-varying measures of behavioral and lifestyle characteristics. First, models were run as a function of all covariates except for the diabetes PGS and ancestry-specific principal components, both on the analytic sample and stratified by race to account for ancestral differences between whites and blacks (20, 21). Most GWASs, including the one conducted by Morris and colleagues (14), are done predominantly on observations of European descent, so the predictive ability of the PGS might differ by race. These models were then run with the addition of the genetic variables as independent variables, again, both on the analytic sample and stratified by race. This second set of models demonstrated how the relationships changed with the inclusion of the genetic components. Concordance values (i.e., the proportion of pairs of cases in which the subject with higher risk had the event before the subject with lower risk) were used as our goodness of fit measures (22). Analyses of deviance, using log likelihoods, were run between corresponding models in the first and second sets (23). Because of the nested nature of these models, these analyses were able to determine how the inclusion of the genetic component altered model fit.

In all our survival models, we included cluster-robust standard errors to account for household stratification in the HRS and to address potential within-household spillover effects (24). We used age as the time unit in all analyses with an individual's age at study entry as the baseline measure. All statistical analyses were performed in R version 3.5.0 (25) with the “survival” package for our primary analyses (26). In all cases, significance was reported at the five-percent level.

RESULTS

Table 1 shows summary characteristics for some basic demographics and the behavioral and lifestyle characteristics of the analytic (i.e., genetic) sample at baseline. The analytic sample was 42.04% male and 79.59% of respondents were non-Hispanic white. The mean age was 56.53 years. The non-Hispanic white sample was slightly older and more male than the non-Hispanic black sample. The mean BMI of the non-Hispanic white sample was about 27 kg/m², which would be classified as overweight, while the mean BMI of the non-Hispanic black sample was about 30 kg/m², which is the threshold for obese. There were fewer regular exercises among non-Hispanic blacks, but fewer current smokers and heavy drinkers among non-Hispanic whites.

A total of 4,217 (27.76%) individuals reported being diagnosed with diabetes over the survey period. In Figure 1, we display the unadjusted cumulative hazard of diabetes onset for non-Hispanic white and non-Hispanic black respondents. As expected, the cumulative hazard increased with advancing age. However, the curve for blacks rose more quickly than that for whites. By the end of the age range, the hazard of diabetes onset was clearly more likely among non-Hispanic blacks than non-Hispanic whites.

Table 2 shows the results from three separate multivariate Cox regression models for diabetes onset as a function of all covariates except the diabetes PGS and ancestry-specific principal components. The first used the analytic sample, the second used only non-Hispanic white respondents, and the third used only non-Hispanic black respondents. In the model using the analytic sample, non-Hispanic whites had a lower risk of diabetes onset relative to non-Hispanic blacks, which we demonstrated in Figure 1. Respondents who reported being married/partnered, being disabled (compared to employed), having higher BMI, being a current smoker (compared to a never smoker), having high blood pressure, and having a cardiovascular disease were significantly associated with increased risk of diabetes onset whereas those who reported being retired (compared to employed), having higher levels of income or wealth, being a Medicare recipient, being a heavy drinker, and having arthritis were significantly associated with reduced risk of diabetes onset. One's sex, foreign-born status, educational attainment, participation in Medicaid, use of other health insurance, or physical activity were not found to be significant.

The results from the model of non-Hispanic whites were the same, likely due to the overwhelming proportion of non-Hispanic whites in the analytic sample. While partnership status, employment status, wealth, being a Medicare recipient, BMI, having high blood pressure, and having arthritis registered significance in the model for non-Hispanic blacks as well, there were some discrepancies in other variables. Income, alcohol consumption, and having a cardiovascular disease were no longer significant at the five-percent level. Former smokers (in addition to current smokers) were found to be at an increased risk of diabetes onset compared to never smokers among non-Hispanic blacks. Additionally, regular exercise was associated with a decreased risk of diabetes onset among non-Hispanic blacks.

We re-estimated the three models in Table 2, but included the diabetes PGS and ancestry-specific principal components. The results from these runs are in Table 3. For the analytic sample, the estimated hazard of the diabetes PGS was 1.16 (95% confidence interval: [1.12, 1.20], $P < 0.001$), suggesting that a one SD increase in the diabetes PGS increased the risk of diabetes by 16% while holding adjusted covariates constant. In these analyses, the diabetes PGS was statistically significant for both non-Hispanic whites (HR = 1.38, 95% confidence interval: [1.30, 1.46]) and non-Hispanic blacks (HR = 1.22, 95% confidence interval: [1.06, 1.40]). The hazard ratio for the analytic sample did not fall between those obtained from the stratified analyses for non-Hispanic whites and non-Hispanic blacks. The stratified models implicitly allowed for interactions between race and all other covariates in the model. Thus, it is possible that allowing for these interactions affected the coefficient estimates for the diabetes PGS.

As before, the variables that were significant in the overall model were also significant in the model for non-Hispanic whites, but these did not necessarily line up with the variables that were significant in the model for non-Hispanic blacks. Again, the differences were that income, alcohol consumption, and having a cardiovascular disease were not significant for diabetes onset among non-Hispanic blacks, while being a former smoker (compared to being a never smoker) and being a regular exerciser were significant. However, arthritis was no longer a significant health comorbidity for non-Hispanic blacks once the diabetes PGS was included.

The variables that were significant before, for the most part, remained significant in the corresponding models that included genetic information, so the inclusion of the diabetes PGS and ancestry-specific principal components generally did not change the significance of any of the associations between the other characteristics and diabetes onset. However, these models informed us that PGS was also a significant variable for diabetes onset and that its relationship

should not be ignored.

Furthermore, models with the diabetes PGS and ancestry-specific principal components performed better than those without them. Concordance was consistently higher for the models in Table 3 than the corresponding models in Table 2. Analyses of deviance were computed for corresponding models in Tables 2 and 3, and these results are presented in Table 4. For the analytic sample, the non-Hispanic white subset, and the non-Hispanic black subset, the tests were statistically significant. That is, the inclusion of these genetic components significantly improved model fit in the explanation of diabetes onset.

DISCUSSION

In the current study, we utilized a national population-based sample of the US to explore diabetes onset and better understand the effects of genetic endowment and time-varying behavioral characteristics commonly associated with diabetes. Models with the genetic component performed significantly better than models without it. The diabetes PGS was consistently statistically significant with diabetes onset after testing different operationalizations and adjusting for a range of characteristics. Respondents with a higher genetic propensity for diabetes were at higher risk of diabetes, irrespective of the other characteristics we included in our model. We found a number of these other characteristics to be statistically significantly associated with diabetes onset, including sociodemographics, economic well-being, behavior and lifestyle, and health comorbidities. Similar behavioral variables were also found to be significant in a population-based study in Turkey (27).

Our cumulative hazard curves demonstrated that diabetes onset differed by whites and

blacks, both in rate of onset as well as overall levels of onset, which has been found previously in the literature (28, 29). In stratified models by race, we found that the association between the diabetes PGS and onset of diabetes to be statistically significant among both non-Hispanic white and black respondents, but the relationship was stronger among whites. Although PGSs were calculated separately for European and African ancestry groups, the GWAS meta-analysis used to estimate SNP weights were derived from analyses based on European ancestry groups; thus, the predictive power of the PGSs for African ancestry groups may vary (13, 21). Therefore, the weaker relationship of the diabetes PGS among the non-Hispanic black sample could be due to a myriad of factors. For example, it could be an artifact of how the PGS was calculated, it could be due to the smaller sample size for non-Hispanic black respondents, or possibly a true weaker association between the diabetes PGS and diabetes onset. Extending GWASs to other ancestry groups is essential for a better understanding of how well these PGSs can actually perform for groups that are not non-Hispanic white.

Regardless of race, genetics were associated with diabetes onset. However, this should not downplay the role of behavioral or lifestyle characteristics. These behavioral and lifestyle characteristics “ultimately interact with risk alleles in susceptibility genes to initiate common forms of [diabetes]” (30). For both non-Hispanic whites and blacks, BMI was significantly associated with a higher propensity of diabetes onset, as was being a current smoker (compared to being a never smoker). Interestingly, being a former smoker (compared to being a never smoker) was also associated with a higher propensity of diabetes onset for blacks. Heavy drinking was associated with a decreased risk of diabetes onset for whites but not blacks, while exercising was associated with a decreased risk of diabetes onset for blacks but not whites. These results demonstrate the potential ability of behavioral characteristics as a mechanism for delaying

or preventing diabetes onset, and lifestyle interventions have indeed been useful in prevention of type II diabetes (31, 32). However, differences between non-Hispanic whites and non-Hispanic blacks in stratified models demonstrate the potential need for targeted interventions, as well as the need to expand this line of research to other race and ethnic population segments.

There are a few limitations to note. In our study, the analytic sample comprised of respondents who consented and provided DNA samples for genotyping. The Appendix presents the summary characteristics of our analytic sample, the summary characteristics of the complete HRS sample, and t-tests or χ^2 -tests as appropriate. Our sample differed significantly when compared to the complete HRS sample, which perhaps is not surprising, as there was selection into the genetic sample. For example, respondents in the analytic sample were more likely to be younger (mean [SD] age, 56.53 [8.44] years vs. 62.76 [12.27] years, $P < 0.001$) and have lower BMI (mean [SD] BMI, 27.82 [5.69] vs. 27.05 [5.52], $P < 0.001$). This is a caveat that has been noted in several prior studies (33-35), and unfortunately cannot be rectified with the use of weights. These differences should be taken into account when considering the results and interpretations of our findings.

Mortality selection could also be a concern, as respondents had to survive to age 50 (or be the spouse of someone who survived to age 50) in order to be in the HRS sampling frame. While this is an issue with all studies using the HRS, respondents in studies also using the genetic component had to survive to 2006-2012 in order to be included for potential genetic sampling.

Another caveat is that the survey question used in the HRS to assess regular physical activity changed after wave 6. In our analysis, a respondent was considered a regular exerciser during waves 1-6 if they reported vigorous physical activity 3+ times per week or, for waves 7-

12, if they reported vigorous physical activity at least once per week. We opted to classify vigorous physical activity based on how it was defined in the wave by HRS.

Despite the limitations, this paper has shown the importance of looking at the effects of genetic and behavioral characteristics together, and that both are necessary in understanding the etiology of diabetes. Although previous papers have examined them together, the advantage of this paper is that we studied their relationship in both non-Hispanic white and black respondents using a national population-based study. Our findings suggest that although genetic variants are associated with diabetes onset, behavioral and lifestyle characteristics remain an important part of diabetes management. BMI, smoking, alcohol, and exercise were all found to be significant in various specifications of our models. Thus, despite the statistically significant role genetic endowment plays in diabetes onset, individuals might still be able to reduce their risk by engaging in protective behaviors, which has substantial clinical relevance.

Diabetes is a multifaceted trait that has both a heritable and lifestyle component. A 2015 review by Prasad and Groop (36) reported that the heritability of type two diabetes mellitus varied between 25% and 80%, depending on the length of follow-up, which may indicate a change in heritability with age and thus the changing importance of modifiable risk factors for diabetes onset. In the context of our findings that both lifestyle factors and genetic risk play a role in diabetes onset, it is important to target lifestyle factors that may mitigate the role of genetic endowment. Thus, future studies should examine gene-environment interactions in the onset of diabetes. Understanding the contribution of lifestyle factors over the lifespan to epigenetic changes in the expression of genetic risk for diabetes would be a valuable contribution to this line of work.

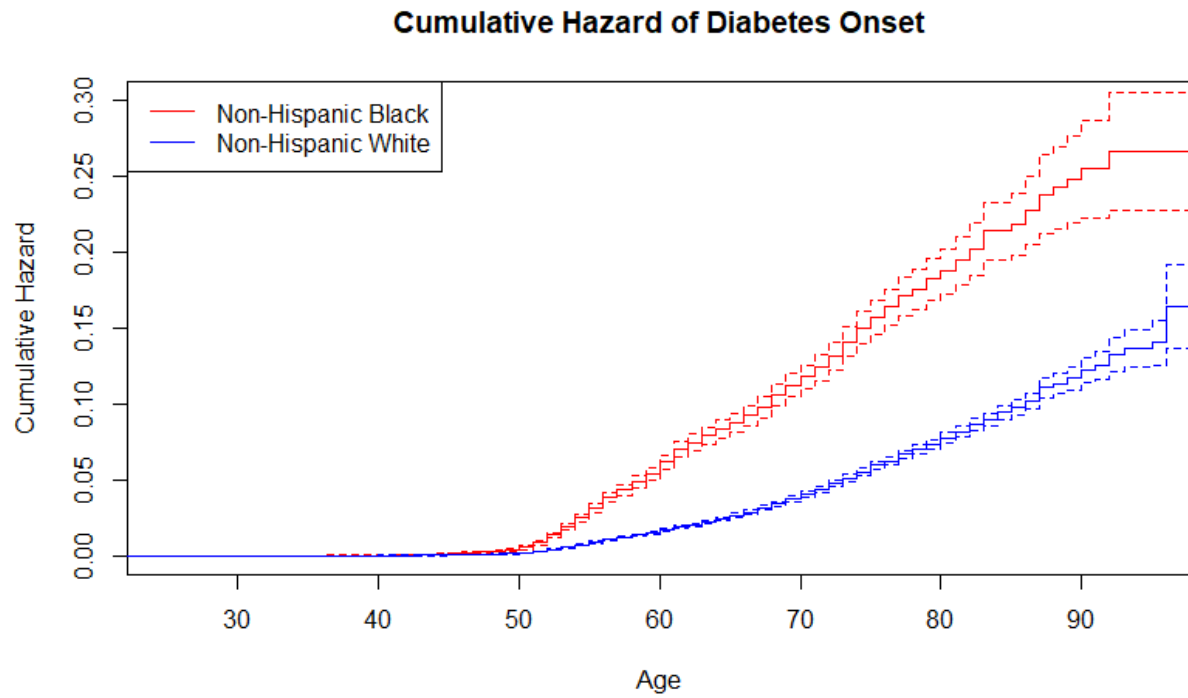
REFERENCES

1. Centers for Disease Control Prevention. National Diabetes Statistics Report. Atlanta, GA: Centers for Disease Control and Prevention. 2017.
2. Centers for Disease Control and Prevention. About Underlying Cause of Death 1999–2015. CDC WONDER Database. <http://wonder.cdc.gov/ucd-icd10.html>. Updated December 2016. Accessed September 22, 2018.
3. American Diabetes Association. Economic Costs of Diabetes in the U.S. in 2017. *Diabetes Care*. 2018;41(5):917-28.
4. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr*. 2010;8:29.
5. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An exposure-wide umbrella review of meta-analyses. *PLoS One*. 2018;13(3):e0194127.
6. Valdez R, Yoon PW, Liu T, Khoury MJ. Family history and prevalence of diabetes in the US population: the 6-year results from the National Health and Nutrition Examination Survey (1999–2004). *Diabetes care*. 2007;30(10):2517-22.
7. Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen MR, Lysenko V, et al. Heritability and familiarity of type 2 diabetes and related quantitative traits in the Botnia Study. *Diabetologia*. 2011;54(11):2811-9.
8. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med*. 2008;359(21):2208-19.
9. Lysenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, et al. Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med*. 2008;359(21):2220-32.
10. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: The Health and Retirement Study (HRS). *Int J Epidemiol*. 2014;43(2):576-85.
11. Ware EB SL, Gard AM, Faul JD. HRS Polygenic Scores – Release 2. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. 2018.
12. RAND. RAND HRS Longitudinal File 2014 (V2) public use dataset. Produced by the RAND Center for the Study of Aging, with funding from the National Institute on Aging and the Social Security Administration. 2018.
13. Ware EB, Schmitz LL, Faul JD, Gard A, Mitchell C, Smith JA, et al. Heterogeneity in polygenic scores for common human traits. *bioRxiv*. 2017:106062.
14. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segre AV, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*. 2012;44(9):981.
15. Walter S, Mejía-Guevara I, Estrada K, Liu SY, Glymour MM. Association of a genetic risk score with body mass index across different birth cohorts. *JAMA*. 2016;316(1):63-9.
16. Ramos AL, Redeker I, Hoffmann F, Callhoff J, Zink A, Albrecht K. Comorbidities in patients with rheumatoid arthritis and their association with patient-reported outcomes: Results of claims data linked to questionnaire survey. *The Journal of rheumatology*. 2019:jrheum. 180668.
17. Fillenbaum GG, Pieper CF, Cohen HJ, Cornoni-Huntley JC, Guralnik JM. Comorbidity of five chronic health conditions in elderly community residents: determinants and impact on mortality. *J Gerontol A Biol Sci Med Sci*. 2000;55(2):M84-M9.
18. Stamler J, Vaccaro O, Neaton JD, Wentworth D. Diabetes, other risk-factors, and 12-year cardiovascular mortality for men screened in the multiple risk factor Intervention trial. *Diabetes Care*. 1993;16(2):434-44.

19. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*. 1958;53(282):457-81.
20. Oetjens MT, Brown-Gentry K, Goodloe R, Dilks HH, Crawford DC. Population stratification in the context of diverse epidemiologic surveys sans genome-wide data. *Front Genet*. 2016;7:76.
21. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100(4):635-49.
22. Stare J, Perme MP, Henderson R. A measure of explained variation for event history data. *Biometrics*. 2011;67(3):750-9.
23. Dalgaard P. *Introductory statistics with R*. Springer-Verlag New York Inc; 2002.
24. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society Series B (Methodological)*. 1972;34(2):187-220.
25. R Core Team. *R: A language and environment for statistical computing*. 3.5.1 ed. Vienna, Austria: R Foundation for Statistical Computing; 2018.
26. Therneau TM, Lumley T. Package 'survival'. *R Top Doc*. 2015;128.
27. Satman I, Yilmaz T, Sengül A, Salman S, Salman F, Uygur S, et al. Population-based study of diabetes and risk characteristics in Turkey: results of the Turkish diabetes epidemiology study (TURDEP). *Diabetes care*. 2002;25(9):1551-6.
28. Narayan KMV. Public health challenges for the 21st century: Convergence of demography, economics, environment and biology: Nalanda distinguished lecture. *Natl Med J India*. 2017;30(4):219-23.
29. Park Y-W, Zhu S, Palaniappan L, Heshka S, Carnethon MR, Heymsfield SB. The metabolic syndrome: Prevalence and associated risk factor findings in the US population from the Third National Health and Nutrition Examination Survey, 1988-1994. *Archives of internal medicine*. 2003;163(4):427-36.
30. Murea M, Ma L, Freedman BI. Genetic and environmental factors associated with type 2 diabetes and diabetic vascular complications. *The review of diabetic studies: RDS*. 2012;9(1):6.
31. Tuomilehto J, Lindström J, Eriksson JG, Valle TT, Hämäläinen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *New England Journal of Medicine*. 2001;344(18):1343-50.
32. Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *The New England journal of medicine*. 2002;346(6):393-403.
33. Domingue BW, Belsky DW, Harrati A, Conley D, Weir DR, Boardman JD. Mortality selection in a genetic sample and implications for association studies. *Int J Epidemiol*. 2017;46(4):1285-94.
34. Zajacova A, Sarah AB. Healthier, wealthier, and wiser: a demonstration of compositional changes in aging cohorts due to selective mortality. *Population Research and Policy Review*. 2013;32(3).
35. Boef A, le Cessie S, Dekkers OM. Mendelian randomization studies in the elderly. *Epidemiology*. 2015;26(2):e15-e6.
36. Prasad RB, Groop L. Genetics of type 2 diabetes—Pitfalls and possibilities. *Genes*. 2015;6(1):87-123.

FIGURES

Figure 1. Unadjusted Cumulative Hazard of Diabetes Onset for the Analytic Sample Stratified by Race with 95% Confidence Interval.



TABLES

Table 1. Summary Characteristics for the Analytic Sample and Race Sub-Samples at Baseline.

Characteristic	Mean or %		
	Analytic sample (n = 15,190)	Non-Hispanic whites (n = 12,090)	Non-Hispanic blacks (n = 3,100)
Age, mean	56.53	56.90	55.09
Male, %	42.04	42.98	38.39
BMI, mean	27.82	27.25	30.04
Regular exerciser, %	32.12	33.76	25.75
Smoking status, %			
Current smoker	21.59	19.61	29.32
Former smoker	36.08	37.54	30.41
Never smoker	42.21	42.75	40.11
Heavy drinker, %	9.33	8.97	10.71

Note. Statistically significant differences between non-Hispanic white and non-Hispanic black respondents were observed for all characteristics at the $p = 0.05$ level.

Table 2. Hazard Ratios from a Multivariate Cox Regression Model Without Diabetes PGS
Included as a Covariate.

Characteristic	Hazard ratio (standard error)		
	Analytic sample (n = 15,190)	Non-Hispanic whites (n = 12,090)	Non-Hispanic blacks (n = 3,100)
Male	1.070 (0.044)	1.071 (0.050)	0.981 (0.092)
Non-Hispanic white	0.735*** (0.055)	NA	NA
Foreign-born	0.944 (0.099)	0.892 (0.119)	1.059 (0.176)
Education			
Less than high school (reference group)	1	1	1
High school/GED	1.018 (0.061)	0.986 (0.072)	1.023 (0.112)
Some college	1.003 (0.067)	0.986 (0.078)	0.980 (0.128)
College or above	0.892 (0.073)	0.875 (0.085)	0.955 (0.149)
Married/partnered	1.597*** (0.050)	1.594*** (0.059)	1.540*** (0.091)
Employment status			
Employed (reference group)	1	1	1
Unemployed	0.981 (0.122)	0.816 (0.163)	1.383 (0.190)
Retired	0.599*** (0.053)	0.568*** (0.062)	0.738** (0.102)
Disabled	1.932*** (0.130)	1.818** (0.183)	2.247*** (0.195)
Not in labor force	0.919 (0.090)	0.862 (0.102)	1.189 (0.212)
Income	0.907*** (0.023)	0.889*** (0.029)	0.953 (0.042)
Wealth	0.920*** (0.012)	0.912*** (0.015)	0.949** (0.020)
Medicare	0.190*** (0.062)	0.173*** (0.073)	0.253*** (0.121)
Medicaid	1.032 (0.091)	0.993 (0.134)	1.136 (0.123)

Health insurance	1.034 (0.044)	0.996 (0.050)	1.191 (0.090)
BMI	1.086*** (0.003)	1.098*** (0.004)	1.057*** (0.006)
Regular exerciser	1.002 (0.042)	1.076 (0.048)	0.833* (0.090)
Smoking status			
Never smoker (reference group)	1	1	1
Current smoker	1.370*** (0.063)	1.382*** (0.075)	1.320* (0.119)
Former smoker	1.076 (0.043)	1.034 (0.049)	1.228* (0.090)
Heavy drinker	0.808** (0.073)	0.779** (0.085)	0.940 (0.147)
High blood pressure	1.852*** (0.043)	1.898*** (0.049)	1.564*** (0.098)
Cardiovascular disease	1.112* (0.046)	1.142** (0.051)	1.053 (0.105)
Arthritis	0.789*** (0.041)	0.771*** (0.046)	0.845* (0.085)
Concordance	0.874	0.872	0.835

Notes. Significance is denoted by *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Model also controls for birth cohort.

Table 3. Hazard Ratios from a Multivariate Cox Regression Model With Diabetes PGS Included as a Covariate.

Characteristic	Hazard ratio (standard error)		
	Analytic sample (n = 15,190)	Non-Hispanic whites (n = 12,090)	Non-Hispanic blacks (n = 3,100)
Diabetes PGS	1.159*** (0.019)	1.380*** (0.030)	1.219** (0.072)
Male	1.075 (0.044)	1.096 (0.051)	0.964 (0.094)
Non-Hispanic white	0.735*** (0.056)	NA	NA
Foreign-born	0.912 (0.100)	0.905 (0.121)	1.076 (0.177)
Education			
Less than high school (reference group)	1	1	1
High school/GED	1.032 (0.061)	1.006 (0.072)	1.002 (0.113)
Some college	1.014 (0.067)	1.012 (0.079)	0.976 (0.129)
College or above	0.883 (0.074)	0.883 (0.087)	0.911 (0.150)
Married/partnered	1.607*** (0.050)	1.594*** (0.060)	1.536*** (0.092)
Employment status			
Employed (reference group)	1	1	1
Unemployed	0.979 (0.122)	0.818 (0.161)	1.339 (0.190)
Retired	0.602*** (0.053)	0.571*** (0.062)	0.726** (0.102)
Disabled	1.965*** (0.129)	1.829*** (0.180)	2.227*** (0.196)
Not in labor force	0.931 (0.090)	0.873 (0.101)	1.184 (0.208)
Income	0.904*** (0.023)	0.891*** (0.029)	0.946 (0.042)
Wealth	0.922*** (0.012)	0.913*** (0.015)	0.949** (0.020)
Medicare	0.190*** (0.062)	0.176*** (0.074)	0.251*** (0.121)

Medicaid	1.019 (0.091)	1.005 (0.134)	1.130 (0.122)
Health insurance	1.032 (0.043)	0.991 (0.050)	1.181 (0.090)
BMI	1.086*** (0.003)	1.098*** (0.004)	1.056*** (0.006)
Regular exerciser	1.000 (0.042)	1.079 (0.048)	0.833* (0.090)
Smoking status			
Never smoker (reference group)	1	1	1
Current smoker	1.370*** (0.063)	1.380*** (0.074)	1.330* (0.120)
Former smoker	1.070 (0.043)	1.021 (0.050)	1.239* (0.092)
Heavy drinker	0.824** (0.074)	0.797** (0.085)	0.943 (0.148)
High blood pressure	1.851*** (0.044)	1.871*** (0.049)	1.582*** (0.099)
Cardiovascular disease	1.116* (0.046)	1.157** (0.052)	1.054 (0.105)
Arthritis	0.796*** (0.040)	0.774*** (0.047)	0.853 (0.085)
Concordance	0.877	0.876	0.838

Notes. Significance is denoted by *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. Models also control for birth cohort and ancestry-specific principal components.

Table 4. Test Statistics from Analyses of Deviance Comparing Models Without and With Diabetes PGS Included as a Covariate.

	Analytic sample (n = 15,190)	Non-Hispanic whites (n = 12,090)	Non-Hispanic blacks (n = 3,100)
Without diabetes PGS	Log lik = -33196	Log lik = -24940	Log lik = -6368
With diabetes PGS	Log lik = -33159	Log lik = -24865	Log lik = -6364
Test statistic	$\chi^2 = 74.41^{***}$	$\chi^2 = 149.41^{***}$	$\chi^2 = 9.09^{**}$

Note. Significance is denoted by *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$.

APPENDIX

Summary Characteristics for the Analytic and Complete Samples.

Characteristic	Mean or %	
	Analytic sample (n = 15,190)	Complete HRS sample (n = 37,495)
Age, mean	56.53	60.24
Male, %	42.04	43.82
Non-Hispanic white, %	79.59	68.20
BMI, mean	27.82	27.36
Regular exerciser, %	32.12	28.83
Smoking status, %		
Current smoker	21.59	21.73
Former smoker	36.08	36.47
Never smoker	42.21	41.63
Heavy drinker, %	9.33	8.39

Note. Statistically significant differences between the analytic sample and excluded respondents were observed for all characteristics at the $p = 0.05$ level.