American Journal of Epidemiology

Vol. 188, No. 8
DOI: 10.1093/aje/kwz125
Advance Access publication:
June 7, 2019

Practice of Epidemiology

Family History–Wide Association Study to Identify Clinical and Environmental Risk Factors for Common Chronic Diseases

Danielle Rasooly, John P. A. Ioannidis, Muin J. Khoury, and Chirag J. Patel*

* Correspondence to Dr. Chirag J. Patel, Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115 (e-mail: Chirag_Patel@hms.harvard.edu).

Initially submitted August 30, 2018; accepted for publication May 9, 2019.

Family history is a strong risk factor for many common chronic diseases and summarizes shared environmental and genetic risk, but how this increased risk is mediated is unknown. We developed a "family history–wide association study" (FamWAS) to systematically and comprehensively test clinical and environmental quantitative traits (CEQTs) for their association with family history of disease. We implemented our method on 457 CEQTs for association with family history of diabetes, asthma, and coronary heart disease (CHD) in 42,940 adults spanning 8 waves of the 1999–2014 US National Health and Nutrition Examination Survey. We conducted pooled analyses of the 8 survey waves and analyzed trait associations using survey-weighted logistic regression. We identified 172 (37.6% of total), 32 (7.0%), and 78 (17.1%) CEQTs associated with family history of diabetes, asthma, and CHD, respectively, in subcohorts of individuals without the respective disease. Twenty associated CEQTs were shared across family history of diabetes, asthma, and CHD, far more than expected by chance. FamWAS can examine traits not previously studied in association with family history and uncover trait overlap, highlighting a putative shared mechanism by which family history influences disease risk.

chronic disease; family history; family history–wide association study; NHANES

Abbreviations: CEQT, clinical and environmental quantitative traits; CHD, coronary heart disease; CHF, congestive heart failure; CI, confidence interval; FamWAS, family history–wide association study; FDR, false discovery rate; NHANES, National Health and Nutrition Examination Survey.

Family history is a well-known risk factor for developing many common chronic diseases, such as diabetes, asthma, and coronary heart disease (CHD), and reflects inherited genetic and shared environmental contribution in disease. Methods to delineate the mechanisms by which family history of disease influences inherited traits and environmental variables can be valuable in identifying how disease risk is conferred and distinguishing possible target areas amenable to intervention.

While previous efforts have studied the association between several specific candidate factors of disease and a family history, there might be as yet many undiscovered traits associated with a positive family history. We present here a "family history–wide association study" (FamWAS) to comprehensively identify clinical and environmental quantitative traits (CEQT) associated with family histories of chronic disease, focusing here on diabetes, asthma, and CHD. FamWAS extends previous studies that assess a few traits at a time with a single family history by systematically evaluating 457 CEQTs, including anthropometric and laboratory measurements as well as environmental pollutants, nutrients, and organic substances in association with family histories among participants in the Continuous National Health and Nutrition Examination Survey (NHANES). Complementary to the methodologies for genome-wide association studies or phenome-wide association studies (1, 2), FamWAS scans for traits associated with a family history in an unbiased approach on a broad scale while controlling for multiple testing, potentially uncovering novel associations that could enhance our understanding of the influence of family history. Some of the associated factors might be the ones that eventually mediate the increase in disease risk that family history is known to confer.

METHODS

NHANES cohort construction

We derived the cross-sectional study cohort from questionnaire and laboratory examinations of 8 independent waves of the 1999–2014 NHANES (United States). Individuals selected

to participate in NHANES were identified via a random sampling method and administered questionnaires on health status, as well as clinical phenotypic measurements (e.g., body mass index) and laboratory tests (e.g., blood and urine analyses) (3).

For each respondent, we obtained demographic information, including age, sex, and race/ethnicity (covariates), as well as family history information and current disease status for diabetes, asthma, and CHD (Web Figure 1A). These conditions were chosen because of the availability of questionnaire family history data in all 8 waves of NHANES. Family history of each disease was ascertained from the Medical Conditions Questionnaire with an affirmative response to the questions "were any of your biological that is, blood relatives including grandparents, parents, brothers, sisters ever told by a health professional that they had" for diabetes, asthma, and heart attack or angina (CHD). Current diabetes status was ascertained in 2 ways: 1) individuals diagnosed with diabetes were ascertained using an affirmative response to the question "have you ever been told by a doctor or health professional that you have diabetes or sugar diabetes," and 2) individuals with undiagnosed diabetes were ascertained from fasting glucose levels greater than 7.0 mmol/L (126 mg/dL) following at least a 6-hour fast or glycated hemoglobin greater than 6.5%, in accordance with the American Diabetes Association guidelines (4). Current status for asthma was ascertained using an affirmative response to "has a doctor or other health professional ever told you that you have asthma," or a ratio of forced expiratory volume in 1 second to forced vital capacity (FEV1/FVC) of less than 0.70, indicative of airflow obstruction (5). We identified individuals diagnosed for the condition CHD by an affirmative response to diagnosis of congestive heart failure (CHF), CHD, angina, or heart attack. To ensure consistent reporting of family history and disease status, individuals with no information on family history or current disease status (a response of "refused" or "don't know") were removed from further analysis. Pregnant women and participants under 18 years of age were also removed.

Clinical and environmental quantitative traits selection

We collected a total of 457 CEQTs that represented a range of anthropometric, laboratory, and environmental attributes from the categories in Web Table 1 (see Web Table 2). We removed noncontinuous traits (e.g., languages spoken at home), and for measurements represented multiple times in different units (e.g., triglycerides measured in mmol/L and mg/dL), we removed all but 1 measurement. To increase statistical power, traits with measurements present in less than 5% of the total population were also removed (e.g., osteoporosis-related measures). We investigated the summary statistics for each CEQT and plotted distributions using the raw values for each CEQT, and we additionally scaled and obtained log (base 10) transformations of each CEQT. We then fitted transformations on a case-by-case basis appropriate for the distributions of each trait. CEQTs with approximately normal distributions were scaled and centered in analyses (mean-subtracted and divided by the standard deviation) to allow comparison of association sizes, which reflect a change per 1 standard deviation of the CEQT. CEQTs with right-skewed distributions were additionally log (base 10)-transformed and scaled.

Statistical analyses

To test for association between family history of disease and prevalent disease, we used survey-weighted logistic regression, adjusting for covariates. To account for the stratified and weighted design of NHANES, all regression models were fitted using the *svyglm* function from the *survey* package in R (R Foundation for Statistical Computing, Vienna, Austria) (6).

For our main analyses, we tested each of the 457 CEQTs for association with family history of diabetes, asthma, and CHD in the pooled NHANES survey data, using survey-weighted linear regression (Web Figure 1B and 1C). We constructed weights appropriate for combining 16 years of data that include 1999 through 2014. Notably, in the main analysis, we evaluated family history–CEQT associations in individuals who were not diagnosed with the disease (also removing undiagnosed individuals with disease, as ascertained above). We did this in order to avoid the situation where disease might have affected some of these CEQTs. We wanted to use the FamWAS approach to reveal correlates of family history, some of which might then act also as risk factors for developing disease, rather than vice versa (disease being a risk factor for these correlates). Following the regression analyses, we used the false discovery rate (FDR) method to correct for multiple testing (7). We also evaluated for comparison family history–CEQT associations on the entire cohort of individuals, regardless of participant disease status. We computed the number of identified traits shared between the 3 types of family history and compared that with the number of expected traits. The expectation was derived by taking the product of the 3 proportions of CEQTs identified for each disease, such that each event is assumed to be independent of the other 2 events, and calculating the product of that probability and the total number of CEQTs.

In addition to the pooled analysis, we conducted a meta-analysis of trait associations across survey years in order to ascertain heterogeneity of CEQTs across the different survey years due to potential year-by-year variation in disease prevalence. We tested each of the 457 CEQTs for association with family history of diabetes, asthma, and CHD in each wave of the NHANES survey using survey-weighted logistic regression, and we then meta-analyzed the associations for each trait and family history of disease across all survey years (Web Figure 2). All meta-analyses were conducted using the *metafor* package in R (R Foundation for Statistical Computing) using a random effects meta-analysis (8). We conducted a statistical test of heterogeneity to determine the variation among the association sizes observed for each trait across survey years.

RESULTS

Participant demographic characteristics

The initial cohort size was 82,091 participants. We excluded individuals who were under age 18 years (34,735 individuals) or pregnant (another 1,182 individuals). To obtain the size for each disease cohort, we excluded participants with missing information on current disease status for diabetes, asthma, and CHD (32 adults for diabetes, 44 for asthma, and 3,916 for CHD), and family history (another 4,401, 3,190, and 1,234 adults, respectively),

which resulted in final cohort sizes for study of 41,741 eligible participants for diabetes, 42,940 for asthma, and 41,024 for CHD (Table 1). Participants who responded "don't know" or "refused" are counted as missing information on family history and current disease status. The weight- and stratification-adjusted demographic characteristics of each cohort are shown in Table 1.

Family history as a risk factor for diabetes, asthma, and CHD

Family history is a well-known risk factor for diabetes, asthma, and CHD (9–11); as expected, we observed a substantially increased risk of disease associated with a self-reported family history that was consistent across all survey years (Web Figure 3). The odds ratio for disease in association with a family history was 3.60 (95% confidence interval (CI): 3.28, 3.96), 2.50 (95% CI: 2.33, 2.67), and 2.68 (95% CI: 2.37, 3.02) for diabetes, asthma, and CHD, respectively. The proportion of individuals with family history was 42.2%, 21.3%, and 13.8% in the diabetes, asthma, and CHD cohorts, respectively (Table 1).

Comprehensive search of CEQTs associated with a family history

We examined the summary statistics (as shown in Web Table 3) and distributions of each CEQT (as shown in Web Figure 4) and applied case-by-case statistical transformations of each CEQT (listed in Web Table 4). Web Figure 5 shows volcano plots illustrating the distribution of the regression coefficients for the CEQTs for family history of diabetes (Web Figure 5A), asthma (Web Figure 5B), and CHD (Web Figure 5C) in individuals without the respective disease. The traits at an FDR less than 5% are annotated on the plots. In the pooled analysis, of 457 tested CEQTs, 172 (37.6% of total), 32 (7.0%), and 78 (17.1%) CEQTs achieved an FDR threshold of 5% for association with family history of diabetes, asthma, and CHD, respectively. The majority of the CEQTs exhibited little variation in study outcomes between NHANES waves, with 38.3%, 40.6%, and 33.3% of the total number of traits at FDR of 5% having an I^2 estimate greater than 25% for diabetes, asthma, and CHD, respectively (Web Figure 6). The meta-analysis of CEQT–family history associations across survey years indicated relatively little heterogeneity (Web Figures 7–9).

High-density lipoprotein cholesterol and adiposity-related traits such as body mass index and waist circumference achieved the highest magnitude of association size for a family history of diabetes and CHD (Web Figures 5A and 5C). Cotinine ($\beta = 0.11$, 95% CI: 0.09, 0.13; $P = 2.6 \times 10^{-7}$), urinary thiocyanate ($\beta = 0.13$, 95% CI: 0.11, 0.16; $P = 9.0 \times 10^{-6}$), and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol ($\beta = 0.13$, 95% CI: 0.11, 0.16; $P = 1.2 \times 10^{-5}$), biomarkers of smoking, were identified in the asthma analyses, indicating that individuals not diagnosed with asthma or with airway obstruction but with a family history of asthma exhibited higher levels of tobacco smoke biomarkers (Web Figure 5B). Respiratory measurements including baseline forced expiratory volume from 0.5 seconds ($\beta = -0.071$, 95% CI: −0.089, −0.053; $P = 2.7 \times 10^{-4}$) to 6 seconds ($\beta = -0.061$, 95% CI: −0.077, −0.044; $P = 6.1 \times 10^{-4}$) and baseline forced vital capacity ($\beta = -0.062$, 95% CI: −0.078, −0.045; $P = 5.7 \times 10^{-4}$) were also associated with a family history of asthma (Web Figure 5B).

Table 1. Demographic Breakdown of Diabetes, Asthma, and Coronary Heart Disease Cohorts Presented as Weighted Percentages, National Health and Nutrition Examination Survey, United States, 1999–2014

Characteristic	Diabetes Cohort			Asthma Cohort			CHD Cohort		
	All (n = 41,741)	Has Diabetes[a] (n = 6,324)	No Diabetes (n = 35,417)	All (n = 42,940)	Has Asthma[b] (n = 6,886)	No Asthma (n = 36,054)	All (n = 41,024)	Has CHD[c] (n = 3,527)	No CHD (n = 37,497)
Age, years[d]	46.24	58.41	44.77	45.68	46.30	45.56	46.09	63.95	44.89
Female sex	51.54	49.02	51.81	51.23	54.01	50.76	51.55	43.29	52.08
Race/ethnicity									
White	69.18	61.15	70.16	68.87	73.44	67.90	69.20	77.70	68.60
Black	11.35	16.37	10.71	11.46	11.89	11.36	11.33	10.59	11.39
Mexican	7.95	8.98	7.82	8.08	4.33	8.86	7.97	4.00	8.24
Other Hispanic	5.45	6.18	5.36	5.48	4.92	5.90	5.43	3.18	5.58
Other	6.07	7.32	5.92	6.11	5.41	6.26	6.01	4.55	6.17
Positive family history	42.20	67.00	39.02	21.26	35.90	18.03	13.79	25.37	13.04

Abbreviation: CHD, coronary heart disease.

[a] Participants with diabetes were classified as diagnosed (self-reported a diagnosis by a doctor or other health professional) or undiagnosed (fasting glucose value greater than 126 mg/dL or glycated hemoglobin value greater than 6.5% in the laboratory testing panels for participants who did not self-report a diabetes diagnosis).

[b] Participants with asthma were classified according to self-reported diagnosis or ratio of forced expiratory volume in 1 second to forced vital capacity (FEV1/FVC) of <0.70.

[c] Participants with coronary heart disease were classified according to self-reported diagnosis for congestive heart failure, coronary heart disease, angina/angina pectoris, or heart attack.

[d] Expressed as mean values.

We identified an inverse association of pyridoxal 5′-phosphate ($\beta = -0.144$, 95% CI: −0.166, −0.121; $P = 1.6 \times 10^{-7}$), the active form of vitamin B6, and a positive association of 2-fluorene ($\beta = 0.151$, 95% CI: 0.120, 0.181; $P = 3.1 \times 10^{-6}$), a polycyclic aromatic hydrocarbon, with a family history of diabetes (in individuals without diabetes) (Web Figure 5A). We also identified combined lutein/zeaxanthin associated with a family history of diabetes ($\beta = -0.122$, 95% CI: −0.14, −0.099; $P = 4.1 \times 10^{-6}$) and CHD ($\beta = -0.132$, 95% CI: −0.164, −0.100; $P = 1.9 \times 10^{-4}$) in individuals without the respective disease (Web Figure 5A and 5C). We found the volatile compounds blood tetrachloroethene ($\beta = -0.151$, 95% CI: −0.192, −0.101; $P = 5.8 \times 10^{-4}$) and blood trichloroethene ($\beta = -0.093$, 95% CI: −0.122, −0.063; $P = 2.5 \times 10^{-3}$) negatively associated with a family history of asthma in individuals without asthma (Web Figure 5B). We found cadmium ($\beta = 0.092$, 95% CI: 0.062, 0.123; $P = 2.9 \times 10^{-3}$), a heavy metal, positively associated with a family history of asthma in individuals without asthma, as well as white blood cell count, measured by eosinophil number ($\beta = 0.075$, 95% CI: 0.055, 0.096; $P = 4.0 \times 10^{-4}$), and monocyte number ($\beta = 0.067$, 95% CI: 0.046, 0.087; $P = 1.4 \times 10^{-3}$), (Web Figure 5B). Body mass index, cotinine, and high-density lipoprotein cholesterol were identified as the traits with the lowest FDR in association with a family history of diabetes, asthma, and CHD, respectively (Web Figure 5).

Shared and distinct family-history associated traits between a cohort of individuals without the respective disease and the entire cohort

We examined the overlap of family history-associated traits in individuals without the respective disease (including diagnosed and undiagnosed individuals) and the entire cohort (individuals with and without disease). Web Figure 10 shows volcano plots for CEQT associations in the entire cohort. A majority of the traits identified in the cohort of individuals without disease overlapped with the traits identified in the entire cohort, with 161 of 172 (93.6%) traits overlapping in the diabetes analyses, 30 of 32 (93.8%) in asthma, and 74 of 78 (94.9%) in CHD. Notably, we observed that many of the traits exhibited a strong positive linear relationship, and this was consistent among all 3 types of family history (Web Figure 11), as well as among all 457 traits (Web Figure 12). We identified 46, 23, and 12 traits that demonstrated discordant results for the cohort of individuals without the respective disease and the entire cohort analyses for diabetes, asthma, and CHD, respectively (Web Figure 13).

Shared traits among family histories of diabetes, asthma, and CHD

We examined traits shared between the family histories as well as traits that were associated with a single type of family history (FDR of 5%) and not with the others (Web Figures 14 and 15). Web Figure 16 shows the 20 shared CEQTs associated with family histories of diabetes, asthma, and CHD. Of the 20 shared traits, 13 traits were not highly correlated (Pearson correlation coefficient (ρ) < 0.50) with each other, which is more than the expected 1.2 noncorrelated traits shared across all 3 diseases. For all 20 shared traits, each trait exhibited either a positive association or a negative association consistent among all 3 types of family history (Web Figure 16). Of the shared traits that exhibited a positive association, 5 were adiposity-related measures (e.g., arm circumference, trunk fat, sagittal abdominal diameter), indicating a shared relationship between the different family disease histories and obesity. Smoking biomarkers (e.g., cotinine and urinary thiocyanate), vitamin-related compounds (e.g., γ-tocopherol), and liver-related compounds (C-reactive protein and bilirubin) were also shared in association with the 3 types of family history. The association sizes were almost always larger for diabetes and CHD than for asthma.

DISCUSSION

In this study, using a FamWAS approach, we comprehensively scanned 457 clinical and environmental quantitative traits for their association with family history of diabetes, asthma, and CHD. By conducting a systematic search, we studied many phenotypes that have not been previously studied for their association with a family history, allowing for discovery of candidate phenotypes or environmental biomarkers. For example, we identified an apparently novel association between decreased levels of pyridoxal 5′-phosphate, the biologically active form of vitamin B6, and family history of diabetes in individuals without diabetes, the implication being that even in individuals with controlled levels of blood glucose and hemoglobin A1C, a positive family history can contribute to decreased levels of vitamin B6. We also identified a novel inverse association between lutein and zeaxanthin, carotenoids with antioxidant properties commonly found in egg yolks and green leafy vegetables, and a family history of diabetes and CHD in individuals without the respective diseases. We further show that our method can lead to identification of traits associated among multiple disease family-history indicators (e.g., family history associations shared between asthma, diabetes, and CHD), providing possible insight into shared underlying biological similarities in the diseases (12).

While we have demonstrated the feasibility of a comprehensive search for traits associated with a family history of disease, we acknowledge that there are some limitations to our methodology. First, current disease diagnosis and family history of disease were ascertained using surveys, and self-reported measurements are prone to measurement errors and recall biases. For example, participants might be underreporting family history of disease status, which could affect the estimates of CEQT–family history associations. Second, the NHANES family history questions pose several limitations. The blood relatives listing in the wording of the question groups first-degree relatives (i.e., parent, sibling) with second-degree relatives (i.e., grandparents); however, it does not list all possible second-degree relatives, such as uncles, aunts, nephews, nieces, and grandchildren. This could result in a smaller population of individuals who reported a positive family history and potentially an underestimation of CEQT–family history associations. Furthermore, the number of available family members and thus also the number of potentially affected family members will vary across participants and is not captured by the survey question.

Also, participants might have partial knowledge of their family history. Third, we excluded a number of participants due to missing information about family history and current disease

status. However, even with the exclusion of participants, we obtained a large sample size of 42,940 eligible participants, and the specificity of self-reported diabetes, asthma, myocardial infarction, and CHD ranged from 95% to 99%, while the sensitivity was 96% for diabetes, 91% for asthma, 90% for myocardial infarction, and 78% for CHD (13, 14). Furthermore, we estimated associations in individuals who were not, to the best of our knowledge, diagnosed with disease. For diabetes, we accounted for undiagnosed diabetes and mistaken reporting by marking participants without a reported diagnosis, but who fit ADA diagnostic criteria for diabetes, as individuals with diabetes for the purposes of all analyses. For asthma, we accounted for individuals with abnormal airway obstruction by marking participants with a ratio of forced expiratory volume in 1 second to forced vital capacity (FEV1/FVC) of less than 0.70. For CHD, we marked participants who self-reported any of 4 cardiovascular events (CHF, CHD, angina, or heart attack) in order to ensure we marked individuals who had a condition with a symptomatic result of angina. Some 14.8% of the participants marked with a cardiovascular event had CHF and did not have any of the other 3 conditions. While CHF is not necessarily always due to CHD, we included these individuals because CHF is often caused by coronary artery disease, heart attack, and other conditions that damage the heart muscle. In studies examining the accuracy of reported family history, a self-reported family history of diabetes, when compared with physician-assessed diabetes status of close relatives, had a sensitivity of 78.5% and specificity of 94.9%; a self-reported family history of CHD, compared with status reported by parents, had 85% sensitivity and 93% specificity; and a self-reported family history of asthma had 53% sensitivity and 99% specificity (15, 16).

Another limitation relates to a segment of the surveyed population with missing data on family history. However, only 2.9%–9.5% of individuals across the 3 cohorts met our selection criteria yet had missing data on family history. We speculate that because most of the missing segment of the sample was of younger age, the magnitude of the CEQT associations might be attenuated. Moreover, missingness was substantial for many traits, and in particular for environmental variables, which can bias CEQT–family history associations and create larger uncertainty about these associations.

Family history could have several advantages over other analytical methods to find risk factors prior to disease onset because it reflects the complex interaction of shared genetic, environmental, lifestyle, and behavioral factors (10, 17). First, family history information is easy to capture and is commonly collected in population-based studies. A recent approach termed genome-wide association study by proxy (GWAX) leveraged family history of disease information along with the genotypes of undiagnosed relatives to identify common genotypes in 12 common diseases, reconfirming known and identifying novel risk loci (18). Their findings demonstrate the utility of family history to conduct association mapping without direct case genotyping. Similarly, we leverage family history information in FamWAS to identify modifiable risk factors in addition to genetic risk factors prior to disease onset. A major strength reflected in FamWAS is in cases where a disease endpoint is unknown; family history information can be leveraged as a substitute in identifying modifiable risk factors shared in households. Our approach adds to the data-driven tools to identify phenotypes and exposures associated with family history. We show the feasibility of our method by reidentifying previously known traits associated with a family history of 3 prevalent chronic diseases.

Future directions include incorporating genotype information to partition DNA-transmitted genetic versus environmental variance in phenotype in family history to decompose the various components of risk influenced by familial disease. Further, we show feasibility in a cross-sectional data set; FamWAS can also be performed with data from longitudinal cohorts including information on time of disease diagnosis for each participant in order to identify potential CEQTs that might mediate the association between family history and risk of disease. This could identify novel traits that could explain some of the remainder of the family history-associated disease risk, of which, for example, in diabetes, the known anthropometric and genetic risk factors currently explain a marginal ~20% of the association between family history and diabetes risk (19, 20). Comparison of FamWAS results across multiple data sets and cohorts with different settings and background could also further help investigators understand the consistency of these associations.

ACKNOWLEDGMENTS

Author affiliations: Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts (Danielle Rasooly, Chirag J. Patel); Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, California (John P. A. Ioannidis); and Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, Georgia (Muin J. Khoury).

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program (grant DGE1745303). C.J.P. is supported by the National Institutes of Health (grant R01AI127250), National Institute of Environmental Health Sciences (grants R00 ES023504 and R21 ES025052), and National Science Foundation Big Data Spoke (grant 1636870).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Conflict of interest: none declared.

REFERENCES

1. Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010; 26(9):1205–1210.
2. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101(1):5–22.
3. Centers for Disease Control and Prevention (CDC). National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey Data. Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention, 1999–2014.

https://www.cdc.gov/nchs/nhanes/. Updated January 7, 2019. Accessed February 12, 2019.
4. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2011;34(suppl 1):S62–S69.
5. Vogelmeier CF, Criner GJ, Martinez FJ, et al. Global Initiative for Chronic Obstructive Lung Disease Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease 2017 Report: GOLD executive summary. *Am J Respir Crit Care Med*. 2017;195(5):557–582.
6. Lumley T. Analysis of complex survey samples. J Stat Softw. 2004;9(1):1–19.
7. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995;57(1):289–300.
8. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36(3): 1–48.
9. Meigs JB, Cupples LA, Wilson PW. Parental transmission of type 2 diabetes: the Framingham Offspring Study. *Diabetes*. 2000;49(12): 2201–2207.
10. Harrison TA, Hindorff LA, Kim H, et al. Family history of diabetes as a potential public health tool. *Am J Prev Med*. 2003; 24(2):152–159.
11. Hariri S, Yoon PW, Qureshi N, et al. Family history of type 2 diabetes: a population-based screening tool for prevention? *Genet Med*. 2006;8(2):102–108.
12. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236–1241.
13. Eliassen BM, Melhus M, Tell GS, et al. Validity of self-reported myocardial infarction and stroke in regions with Sami and Norwegian populations: the SAMINOR 1 survey and the CVDNOR project. *BMJ Open*. 2016;6(11):e012717.
14. Oksanen T, Kivimäki M, Pentti J, et al. Self-report as an indicator of incident disease. *Ann Epidemiol*. 2010;20(7): 547–554.
15. Janssens AC, Henneman L, Detmar SB, et al. Accuracy of self-reported family history is strongly influenced by the accuracy of self-reported personal health status of relatives. *J Clin Epidemiol*. 2012;65(1):82–89.
16. Bensen JT, Liese AD, Rushing JT, et al. Accuracy of proband reported family history: the NHLBI Family Heart Study (FHS). *Genet Epidemiol*. 1999;17(2):141–150.
17. Valdez R, Yoon PW, Qureshi N, et al. Family history in public health practice: a genomic tool for disease prevention and health promotion. *Annu Rev Public Health*. 2010;31: 69–87.
18. Liu JZ, Erlich Y, Pickrell JK. Case-control association mapping by proxy using family history of disease. *Nat Genet*. 2017;49(3):325–331.
19. InterAct Consortium, Scott RA, Langenberg C, et al. The link between family history and risk of type 2 diabetes is not explained by anthropometric, lifestyle or genetic risk factors: the EPIC-InterAct study. *Diabetologia*. 2013;56(1): 60–69.
20. Raghavan S, Porneala B, McKeown N, et al. Metabolic factors and genetic risk mediate familial type 2 diabetes risk in the Framingham Heart Study. *Diabetologia*. 2015;58(5):988–996.