HHS Public Access

Author manuscript

BMJ Public Health. Author manuscript; available in PMC 2024 January 18.

Published in final edited form as:

BMJ Public Health. 2023 ; 1(1): . doi:10.1136/bmjph-2023-000655.

Sick individuals, sick populations revisited: a test of the Rose hypothesis for type 2 diabetes disparities

Sonali Gupta[1], **I King Jordan**[2], **Leonardo Mariño-Ramírez**[1]

[1]National Institute on Minority Health and Health Disparities, National Institutes of Health, Rockville, Maryland, USA

[2]School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

Abstract

Introduction—The Rose hypothesis predicts that since genetic variation is greater within than between populations, genetic risk factors will be associated with individuals' risk of disease but not population disparities, and since socioenvironmental variation is greater between than within populations, socioenvironmental risk factors will be associated with population disparities but not individuals' disease risk.

Methods—We used the UK Biobank to test the Rose hypothesis for type 2 diabetes (T2D) ethnic disparities in the UK. Our cohort consists of 26 912 participants from Asian, black and white ethnic groups. Participants were characterised as T2D cases or controls based on the presence or absence of T2D diagnosis codes in electronic health records. T2D genetic risk was measured using a polygenic risk score (PRS), and socioeconomic deprivation was measured with the Townsend

Correspondence to Dr I King Jordan; king.jordan@biology.gatech.edu and Dr Leonardo Mariño-Ramírez; marino@nih.gov, **Twitter** | I King Jordan @ikingjordan.

Contributors IKJ and LM-R conceived of the study, and provided supervision and funding. SG performed all data analyses. SG, IKJ and LM-R prepared figures, wrote and edited the manuscript. All authors read and approved the final manuscript. IKJ and LM-R are study guarantors.

Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjph-2023-000655).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Ethics approval Ethics approval for UKB was obtained from the Community Health Index Advisory Group (CHIAG) for Scotland, the Patient Information Advisory Group (PIAG) for England and Wales, and the North West Multi-center Research Ethics Committee (MREC) for the UK (project ID 299116). Written informed consent was obtained from all participants. The UK Biobank data were accessed under application number 65206. All data available to researchers had direct identifiers removed. Because the UK Biobank data were not collected specifically for this study, and no one on our study team has access to the subject identifiers linked to the data, this study is not considered human subjects research according to the NIH Revised Common Rule for the Protection of Human Subjects: https://grants.nih.gov/policy/humansubjects/hs-decision.htm.

Index (TI). The variation of genetic (PRS) and socioeconomic (TI) risk factors within and between ethnic groups was calculated using analysis of variance. Multivariable logistic regression was used to associate PRS and TI with T2D cases, and mediation analysis was used to analyse the effect of PRS and TI on T2D ethnic group disparities.

Results—T2D prevalence differs for Asian 23.34% (OR=5.14, CI=4.68 to 5.65), black 16.64% (OR=3.81, CI=3.44 to 4.22) and white 7.35% (reference) ethnic groups in the UK. Both genetic and socioenvironmental T2D risk factors show greater within (w) than between (b) ethnic group variation: PRS w=64.60%, b=35.40%; TI w=71.18%, b=28.19%. Nevertheless, both genetic risk (PRS OR=1.96, CI=1.87 to 2.07) and socioeconomic deprivation (TI OR=1.09, CI=1.08 to 1.10) are associated with T2D individual risk and mediate T2D ethnic disparities (Asian PRS=22.5%, TI=9.8%; black PRS=32.0%, TI=25.3%).

Conclusion—A relative excess of within-group versus between-group variation does not preclude T2D risk factors from contributing to T2D ethnic disparities. Our results support an integrative approach to health disparities research that includes both genetic and socioenvironmental risk factors.

INTRODUCTION

In his seminal 1985 paper ‘Sick individuals and sick populations’, epidemiologist Geoffrey Arthur Rose CBE considered whether and how the causes of individual disease cases relate to the causes of population disease disparities.[1] He regarded these as two distinct issues and fundamental to epidemiological inquiry. The concept of variation was central to Rose’s question. He pointed out that relative risk, that is, the relative disease risk of exposed versus non-exposed individuals, was a mainstay of epidemiological studies, and he stressed that the elucidation of relative risk requires variation in exposure levels. By way of example, Rose explained that if everyone smoked 20 cigarettes a day, neither case–control nor cohort studies would find an association between smoking and lung cancer. In other words, interindividual variation in smoking (the exposure) is prerequisite for an association with lung cancer (the outcome) via relative risk calculations.

Variation was equally important to the distinction that Rose made between risk factors for individual disease cases versus risk factors for population disease disparities. He hypothesised that exposures that varied more within than between populations should be associated with individual disease cases, whereas exposures that varied more between than within populations should be associated with population disease disparities. Rose’s hypothesis about the variance of exposures (ie, disease risk factors) made corollary predictions regarding the relative contributions of genetic (nature) versus environmental (nurture) risk factors to individual disease cases and population disparities. He predicted that (1) since genetic variation is much greater within than between populations, genetic risk factors should be associated with individual cases but not population disparities, and (2) since environmental exposure variation is much greater between than within populations, environmental risk factors should be associated with population disparities but not individual cases.

The objective of this study was to interrogate the Rose hypothesis for genetic versus environmental risk factors of individual disease cases versus population disease disparities for type 2 diabetes (T2D) in the UK. T2D is a paradigm of health disparities research; it is a complex disease with a high overall burden of morbidity and mortality, shows starkly disparate prevalence across ethnic groups, and has well-characterised genetic and socioenvironmental risk factors.[2-12] We tested the Rose hypothesis for T2D genetic versus socioenvironmental risk factors via analysis of Asian and black minority populations in the UK, compared with the majority white population, using a cohort from the UK Biobank (UKB). The UKB is a prospective cohort study of more than 500 000 participants, for whom demographic, genetic, socioenvironmental and clinical outcome data were collected. First, we confirmed previously observed T2D ethnic disparities using the UKB cohort. Then, we asked three specific questions: (1) how much of the variation for genetic and socioenvironmental T2D risk factors is found within versus between UK ethnic groups?, (2) are genetic and socioenvironmental risk factors associated with individual T2D cases?, and (3) do genetic and socioeconomic risk factors mediate T2D ethnic disparities?

METHODS

Study design, setting and participants

We used an observational study design with the UKB, a prospective cohort study of more than 500 000 volunteer participants.[13] Participants were recruited via 9.2 million mailed invitations, with a ~5.5% response rate.[14] Participant inclusion criteria included adults aged 40–69 years at recruitment, the capacity to consent and living within 20–25 miles of 1 of the 22 UKB assessment centres, located in England, Scotland and Wales. Participants were enrolled from 2006 to 2010 with baseline data and biological samples collected upon enrolment. Baseline data included questionnaires on sociodemographic data. Biological samples included blood samples that were used for DNA isolation and genetic data characterisation.[15 16] Detailed follow-up on participants' health is made possible by linkage to electronic health record (EHR) data from the UK National Health Service (NHS). UKB participant health, demographic, genetic and socioeconomic data were accessed and downloaded via registration with the Access Management System (application number 65206).

Patient and public involvement

UKB volunteer participants were not involved in any stage of the research process.

Variables, data sources and measurement

The outcome was a clinical diagnosis of T2D by the NHS, recorded by International Classification of Diseases Version 10 (ICD-10) codes taken from participants' EHRs (UKB Showcase data-field 41202), which were subsequently converted to phecodes. Phecodes are widely used to define disease phenotypes from EHRs.[17] Phecodes serve to group granular ICD-10 codes into a unified diagnosis code to better delineate cases, to define exclusion criteria for closely related conditions to define controls and to scale to biobank-size datasets. Participants with ICD-10 codes corresponding to T2D diagnoses were converted to the phecode 250.2 to define T2D cases; participants who did not have any ICD-10 codes

corresponding to the T2D exclusion phecodes 249–250.99 were defined as controls (online supplemental figure 1). T2D genetic risk and socioeconomic deprivation were used as predictors. Age (data-field 21003) and sex (data-field 31) were used as covariates in all models.

T2D genetic risk: genotype-imputed whole-genome variant data were accessed from data-field 21008. A T2D polygenic risk score (PRS) was created using 359 T2D-associated genetic variants taken from the Type 2 Diabetes Knowledge Portal (T2KDP).[3] The T2KDP provides a list of T2D-associated genetic variants (p<5e-8)—including variant positions, alleles, effect sizes and p values—based on a meta-analysis of 382 T2D genome-wide association studies (GWAS). The program PRSice-2 was used to calculate PRS based on analysis of participants' genomic variant data.[18] PRSice-2 uses the standard clumping and thresholding method to calculate PRS as the weighted sum of T2D-associated effect alleles found in the genome of any individual i: $PRS_i = \sum_j^M \hat{\beta}_j \times dosage_{ij}$, where M is the total number of variants, j is the jth variant, β_j is the GWAS-estimated allelic effect size for variant j and $dosage_{ij}$ is the number of effect alleles (0, 1 or 2) for variant j in the genome of individual i.[19] PRSice-2 was run with a minor allele frequency threshold of 0.1.

Two composite metrics were used to measure place-based socioeconomic deprivation: the Townsend Index (TI; data-field 189)[20] and the Index of Multiple Deprivation (IMD; data-field 26410).[21 22] The TI combines four domains of deprivation: unemployment, non-car ownership, non-home ownership and household values. The IMD combines seven domains of deprivation: income, employment, education, health, crime, barriers to housing service and living environment. TI and IMD scores are assigned to participants corresponding to their postcode output area. Lower values of TI and IMD correspond to less socioeconomic deprivation and relative affluence, and higher values correspond to more socioeconomic deprivation.

Study size

The study cohort of 26 912 participants was created by taking all UKB participants who self-identified as belonging to the minority Asian or black ethnic groups (online supplemental figure 2). Since participants who self-identified as belonging to the white ethnic group make up the vast majority of the UKB (506 136; >94%), white participants were randomly downsampled to 10 000 participants to create a more balanced cohort. Random downsampling of white participants was repeated 100 times to ensure representativeness of this group. Participants who either did not have ICD-10 code data, or had excluded ICD-10 codes according to the phecode criteria, were excluded from the cohort. For specific analyses that included PRS, TI or IMD variables, participants with missing data were excluded from each analysis.

Statistical methods

All statistical analyses were performed using the R programming language V.4.2.1,[23] and visualisations were created using ggplot2.[24] Unadjusted T2D per cent prevalence was calculated as—[*cases*/ (*cases* + *controls*)] × 100—where cases are participants with a T2D diagnosis recorded in EHRs and controls are participants with no T2D diagnosis and no

excluded diagnoses. Analysis of variance (ANOVA) and t-tests were performed using the aov and t.test functions from the R Stats Package. ANOVA was used with the method of moments to partition the variance of genetic (PRS) and socioenvironmental (TI) T2D risk factors within and between ethnic groups. Multivariable logistic regression was used to model T2D (outcome) by ethnicity (predictor), genetic risk (predictor) and socioeconomic deprivation (predictor), with age and sex as covariates. Model specifications are provided in the corresponding table and figure legends. Logistic regression was run using the glm function for fitting generalised linear models, with the binomial family parameter, from the R Stats Package. Multiple mediation analysis was performed using the mma R package.[25] For multiple mediation analysis, T2D was the outcome, ethnicity was the predictor, genetic risk (PRS) and socioeconomic deprivation (TI) were mediators, and age and sex were covariates. The mma package was run with alpha=0.4, alpha2=0.4 and n2=10 parameter settings. Complete case analysis (listwise deletion), using only participants for which there are no missing data, was used for all statistical analyses.

RESULTS

T2D ethnic disparities

The UKB was used to build a T2D case–control study cohort made up of volunteer participants who self-identified as belonging to one of the two largest minority ethnic groups in the UK, Asian and black, or the majority white ethnic group (table 1). The full cohort includes 26 912 participants. Asian participants make up 34.8% (9372) of the cohort and black participants make up 28.0% (7540) of the cohort. Since white participants make up >94% of all UKB participants, this group was downsampled to include 10 000 participants for the study cohort (37.2%). The study cohort is 51.9% male and 48.1% female, with an average age of 54 years.

Clinical diagnoses gleaned from participant EHRs were used to define T2D cases and controls for the purpose of calculating T2D prevalence and for use as outcomes for logistic regression modelling. Minority Asian (23.34%) and black (16.64%) groups show higher unadjusted T2D prevalence than the majority white group (7.35%, table 1), where T2D prevalence for the white group is calculated as the average prevalence value across 100 random samples of 10 000 participants each (CI=7.18% to 7.52%). Multivariable logistic regression was used to model the association of ethnicity with T2D, using the majority white as the reference ethnic group and controlling for age and sex. Asian (OR=5.02, CI=4.57 to 5.52) and black (OR=3.93, CI=3.54 to 4.37) ethnicity are positively associated with T2D (table 2). The T2D ethnic disparities observed for the UKB are similar to previously reported T2D disparities for UK minority ethnic groups.[2 4 5 10-12]

Variation of genetic and socioenvironmental risk factors

Participants' genetic risk of T2D was estimated using a PRS, which is calculated as the weighted sum of T2D risk-increasing alleles at 359 loci genome-wide.[3] Participants' socioeconomic deprivation was measured by the TI and the IMD, composite place-based metrics that incorporate multiple dimensions of deprivation.[20] Average PRS values differ significantly among ethnic groups (ANOVA F=4904, P≈0; Kruskal-Wallis χ^2=7331, P≈0;

figure 1A). The black group shows the highest average PRS (6.29×10^{-4}), followed by the Asian (3.31×10^{-4}) and white (-3.09×10^{-4}) groups. Average PRS values also differ significantly among all pairs of ethnic groups (Tukey honest significant difference test Asian–white $d=6.23\times10^{-4}$, P≈0; black–white $d=9.39\times10^{-4}$, P≈0; Asian–black $d=3.16\times10^{-4}$, P≈0). Most of the variation for PRS values is found within (64.6%) rather than between (35.4%) ethnic groups (figure 1B and online supplemental table 1).

Average TI values also differ among groups (ANOVA F=3515, P≈0; Kruskal-Wallis χ^2=5514, P≈0; figure 1C). The black group shows the highest average TI (2.65), followed by the Asian (0.33) and white (−1.44) groups. Average TI values also differ significantly among all pairs of ethnic groups (Tukey HSD test Asian–white d=1.77, P≈0; black–white d=4.09, P≈0; Asian–black d=2.32, P≈0). Most of the variation in TI values is also found within (71.8%) rather than between (28.2%) ethnic groups (figure 1D and online supplemental table 1). Similar results are seen when socioeconomic deprivation is measured by IMD; there are significant differences among groups, with most of the variation found within compared with between groups (online supplemental table 1).

Individual-level T2D risk factor associations

Genetic and socioenvironmental risk factors are associated with T2D risk among individual UKB participants. The observed prevalence of T2D increases across increasing values of the PRS; T2D prevalence is 33.46% for the highest PRS percentile bin compared with 4.83% for the lowest bin (figure 2A). Participants in the top 10% of the PRS have 3.31 greater odds of T2D compared with the remaining 90% of participants. The average PRS value of cases is higher than the average PRS value of controls (t-test t=29.15, p=6.59e-175; figure 2B).

The observed prevalence of T2D also increases across increasing values of the TI, although not as steeply as seen for the PRS. T2D prevalence is 25.65% for the highest TI percentile bin compared with 6.30% for the lowest bin (figure 2C). Participants in the top 10% of the TI have 1.75 greater odds of T2D compared with the remaining 90% of participants. The average TI value of cases is higher than the average TI value of controls (t-test t=19.73, p=6.22e-84; figure 2D). Similar results are seen when socioeconomic deprivation is measured by IMD (online supplemental figure 3).

Logistic regression was used to model the association of genetics (PRS) and socioeconomic deprivation (TI) with T2D, controlling for age and sex. The PRS (OR=1.96, CI=1.87 to 2.07) and the TI (OR=1.09, CI=1.08 to 1.10) are both positively associated with T2D (figure 3 and online supplemental table 2). The effect size estimates for the PRS and the TI change only slightly between unadjusted models, where both variables are modelled separately, and adjusted models, where they are modelled together. The IMD also shows positive associations with T2D when modelled separately or together with the PRS (online supplemental table 2).

Ethnic group T2D disparity mediation

Multiple mediation analysis was used to evaluate the extent to which genetic risk factors (PRS) and socioeconomic deprivation (TI and IMD) mediate the observed T2D ethnic disparities (figure 4A).[25 26] Considered together, the PRS and the TI mediate 32.3% of

the Asian group effect on T2D and 57.3% of the black group effect (figure 4B). PRS mediates more of the ethnicity effect than TI for both the Asian and the black groups (Asian PRS=22.5%, TI=9.8%; black PRS=32.0%, TI=25.3%).

Risk factor and disease simulation

We considered that our seemingly paradoxical results could be attributed to the fact that greater within-group versus between-group variation of disease risk factors should not necessarily preclude their association with group differences in disease prevalence. To test this idea, we simulated normal risk factor distributions for two groups of 1000 individuals, where the group means (100 and 115) differ by 1 SD (15), and we considered a liability threshold model, with individuals in the top 10% risk stratum to be disease cases (online supplemental figure 4). Risk factors for the simulated groups show 64.4% within-group variation and 35.6% between-group variation. Nevertheless, the group risk factor means are different (t-test t=23.52, p=3.36e-108) and group membership is associated with disease (OR=8.95, CI=5.59 to 15.25).

DISCUSSION

Among health disparities researchers, there is a broad sense that individual-level health outcomes result from genetics, socioenvironmental factors and their interactions, whereas population-level disparities are primarily linked to structural inequities resulting in social, economic and environmental disadvantage.[27 28] The intellectual origins of this idea can be traced to Rose's central insights about the distinction between individual-level versus population-level disease risk factors and the relationship of risk factor variance components to this difference. It has long been appreciated that most human genetic variation falls within, rather than between, population groups,[29] a fact which Rose took to argue against a role for genetics in population-level disparities.[1] Rose also held that environmental variation was greater between, rather than within, populations, and therefore environmental risk factors are more likely to be associated with population-level disparities.

Contrary to Rose's predictions, both genetic and socioenvironmental risk factors show far greater variation within rather than between UK ethnic groups. In other words, genetic or socioenvironmental differences between any two individuals within a UK ethnic group are likely to be greater than the corresponding average differences between groups. Nevertheless, both classes of risk factors are associated with individual-level T2D cases and mediate the observed T2D prevalence differences between ethnic groups, indicating that their similar patterns of within-group versus between-group variation do not preclude disease associations at both the individual and population levels. Furthermore, genetic risk factors (PRS) show stronger associations with T2D than socioeconomic deprivation (TI) at both the individual and ethnic group levels, even though TI shows a relatively greater amount of within-group versus between-group variation. In addition to these empirical findings, we used simulation to show that an excess of within-group variation does not theoretically preclude a risk factor from contributing to between-group differences in disease risk.

This study may have a limitation related to volunteer participant sampling bias. UKB participants tend to be in better health and have higher socioeconomic status, on average, compared with the broader UK population. This 'healthy volunteer' bias could impact the disease prevalence and ethnic disparities reported in this study.[30] As a result, the extent to which our findings correspond to the general UK population might differ depending on the specific disease and population group being considered. However, it is worth highlighting that despite this potential limitation, the T2D ethnic disparities observed here align with findings from epidemiological studies of the broader UK population.[2 4 5 10-12]

We hope that our findings may support a reconciliation between health disparities researchers focused on the role of genetics and biological pathways with researchers dedicated to the study of social determinants of health. Indeed, our results support an approach to health disparities research that is cause-agnostic, data-driven, and includes both genetic and socioenvironmental risk factors. The decreasing cost and increasing availability of genomic data, together with the growth of population biobanks worldwide, make this approach to health disparities research far more feasible than ever before.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by the Division of Intramural Research of the National Institute on Minority Health and Health Disparities at the National Institutes of Health to LM-R (award numbers: 1ZIAMD000016 and 1ZIAMD000018); the National Institutes of Health Distinguished Scholars Program to LM-R; and the IHRC-Georgia Tech Applied Bioinformatics Laboratory to IKJ (award number: RF383).

Data availability statement

Data are available in a public, open access repository. The UK Biobank is an open-access study. Data are publicly available via registration with the Access Management System.

REFERENCES

1. Rose G. Sick individuals and sick populations. Int J Epidemiol 2001;30:427–32; [PubMed: 11416056]
2. Barnett AH, Dixon AN, Bellary S, et al. Type 2 diabetes and cardiovascular risk in the UK South Asian community. Diabetologia 2006;49:2234–46. [PubMed: 16847701]
3. Costanzo MC, von Grotthuss M, Massung J, et al. The type 2 diabetes knowledge portal: an open access genetic resource dedicated to type 2 diabetes and related traits. Cell Metab 2023;35:695–710. [PubMed: 36963395]
4. Espelt A, Arriola L, Borrell C, et al. Socioeconomic position and type 2 diabetes mellitus in Europe 1999-2009: a panorama of inequalities. Curr Diabetes Rev 2011;7:148–58. [PubMed: 21418003]
5. Goff LM. Ethnicity and type 2 diabetes in the UK. Diabet Med 2019;36:927–38. [PubMed: 30614072]
6. Mahajan A, Spracklen CN, Zhang W, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat Genet 2022;54:560–72. [PubMed: 35551307]

7. Mahajan A, Wessel J, Willems SM, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet 2018;50:559–71. [PubMed: 29632382]
8. Nagar SD, Jordan IK, Mariño-Ramírez L. The landscape of health disparities in the UK Biobank. Database (Oxford) 2023;2023:baad026. [PubMed: 37114803]
9. Nagar SD, Nápoles AM, Jordan IK, et al. Socioeconomic deprivation and genetic ancestry interact to modify type 2 diabetes ethnic disparities in the United Kingdom. EClinicalMedicine 2021;37:100960. [PubMed: 34386746]
10. Pham TM, Carpenter JR, Morris TP, et al. Ethnic differences in the prevalence of type 2 diabetes diagnoses in the UK: cross-sectional analysis of the health improvement network primary care database
11. Thomas C, Nightingale CM, Donin AS, et al. Socio-economic position and type 2 diabetes risk factors: patterns in UK children of South Asian, black African-Caribbean and white European origin. PLoS One 2012;7:e32619. [PubMed: 22412897]
12. Whincup PH, Nightingale CM, Owen CG, et al. Early emergence of ethnic differences in type 2 diabetes precursors in the UK: the child heart and health study in England (CHASE study). PLoS Med 2010;7:e1000263. [PubMed: 20421924]
13. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep Phenotyping and Genomic data. Nature 2018;562:203–9. [PubMed: 30305743]
14. Allen N, Sudlow C, Downey P, et al. UK Biobank: Current status and what it means for epidemiology. Health Policy Technol 2012;1:123–6.
15. Welsh S, Peakman T, Sheard S, et al. Comparison of DNA Quantification methodology used in the DNA extraction protocol for the UK Biobank cohort. bMc Genomics 2017;18:26. [PubMed: 28056765]
16. Elliott P Peakman TC, Biobank UK. The UK Biobank sample handling and storage protocol for the collection, processing and Archiving of human blood and urine. Int J Epidemiol 2008;37:234–44. [PubMed: 18381398]
17. Bastarache L. Using Phecodes for research with the electronic health record: from Phewas to Phers. Annu Rev Biomed Data Sci 2021;4:1–19. [PubMed: 34465180]
18. Choi SW, O'Reilly PF. Prsice-2: Polygenic risk score software for Biobank-scale data. Gigascience 2019;8:giz082. [PubMed: 31307061]
19. Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing Polygenic risk score analyses. Nat Protoc 2020;15:2759–72. [PubMed: 32709988]
20. Townsend P Phillimore P Beattie A. Health and deprivation: inequality and the North. London: Routledge, 1988.
21. Abel GA, Barclay ME, Payne RA. Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. BMJ Open 2016;6:e012750.
22. Noble M, Wright G, Smith G, et al. Measuring multiple deprivation at the small-area level. Environ Plan A 2006;38:169–85.
23. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013.
24. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer-Verlag, 2016.
25. Yu Q, Li B. Mma: an R package for mediation analysis with multiple mediators. JORS 2017;5:11.
26. Naimi AI, Schnitzer ME, Moodie EEM, et al. Mediation analysis for health disparities research. Am J Epidemiol 2016;184:315–24. [PubMed: 27489089]
27. West KM, Blacksher E, Burke W. Genomics, health disparities, and missed opportunities for the nation's research agenda. JAMA 2017;317:1831–2. [PubMed: 28346599]
28. Sankar P, Cho MK, Condit CM, et al. Genetic research and health disparities. JAMA 2004;291:2985–9. [PubMed: 15213210]
29. Lewontin RC. The apportionment of human diversity. Evol Biol 1972;6:381–98.

30. 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;186:1026–34. [PubMed: 28641372]

WHAT IS ALREADY KNOWN ON THIS TOPIC

- Geoffrey Rose famously predicted that distinct classes of risk factors will explain individual-level disease cases (genetic risk factors) versus group-level disparities (environmental risk factors), owing to their within-group versus between-group variance components.

WHAT THIS STUDY ADDS

- We tested the Rose hypothesis for type 2 diabetes (T2D) ethnic disparities in the UK Biobank.
- T2D genetic and socioeconomic risk factors show similar variance components, with more variation seen within than between Asian, black and white UK ethnic groups.
- Nevertheless, genetic and socioeconomic risk factors are associated with both individual-level T2D cases and T2D ethnic disparities.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- This study upends the conceptual foundations of a health disparities research paradigm focused primarily on social determinants of health and underscores the need for the inclusion of both genetic and socioenvironmental risk factors.

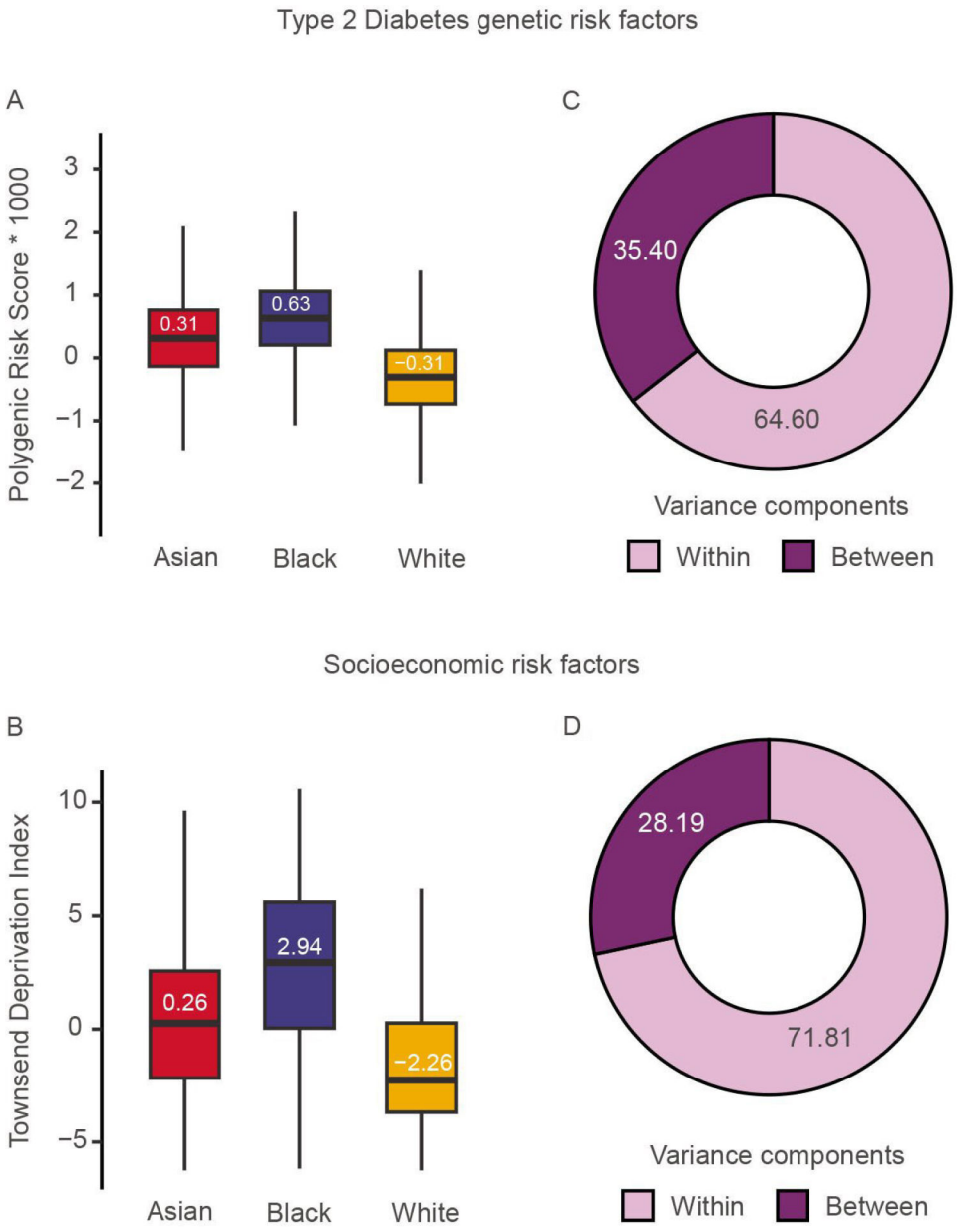

Figure 1.

Ethnic group distributions and variance components for type 2 diabetes genetic risk factors and socioeconomic deprivation. (A) Box-plot distributions of polygenic risk scores (PRS) for Asian (red), black (blue) and white (orange) ethnic groups. (B) PRS variance components within and between ethnic groups. (C) Box-plot distributions of Townsend index (TI) values for Asian (red), black (blue) and white (orange) ethnic groups. (D) TI variance components within and between ethnic groups.

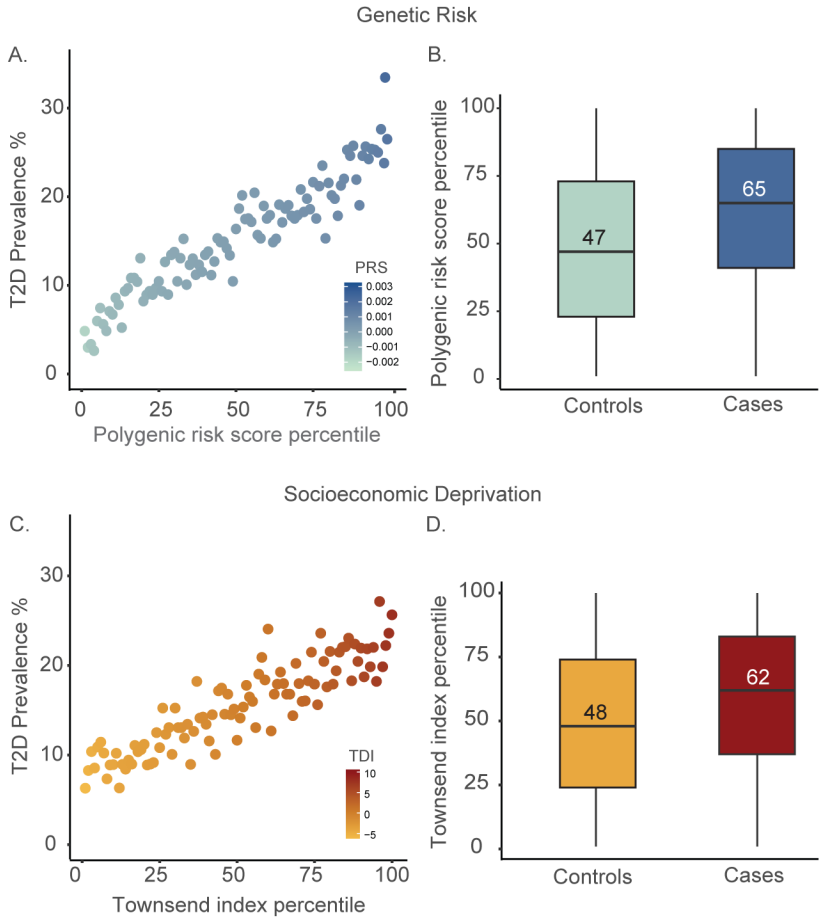

Figure 2.

Type 2 diabetes (T2D), genetic risk and socioeconomic deprivation. (A) T2D per cent prevalence (y-axis) is shown for 100 polygenic risk score (PRS) percentile bins (x-axis). (B) Box-plot distributions of T2D PRS percentiles, with median values shown, for T2D controls and cases. (C) T2D per cent prevalence (y-axis) is shown for 100 Townsend Index (TI) percentile bins (x-axis). (D) Box-plot distributions of TI percentiles, with median values shown, for T2D controls and cases. PRS, polygenic risk score; TI, Townsend Index.

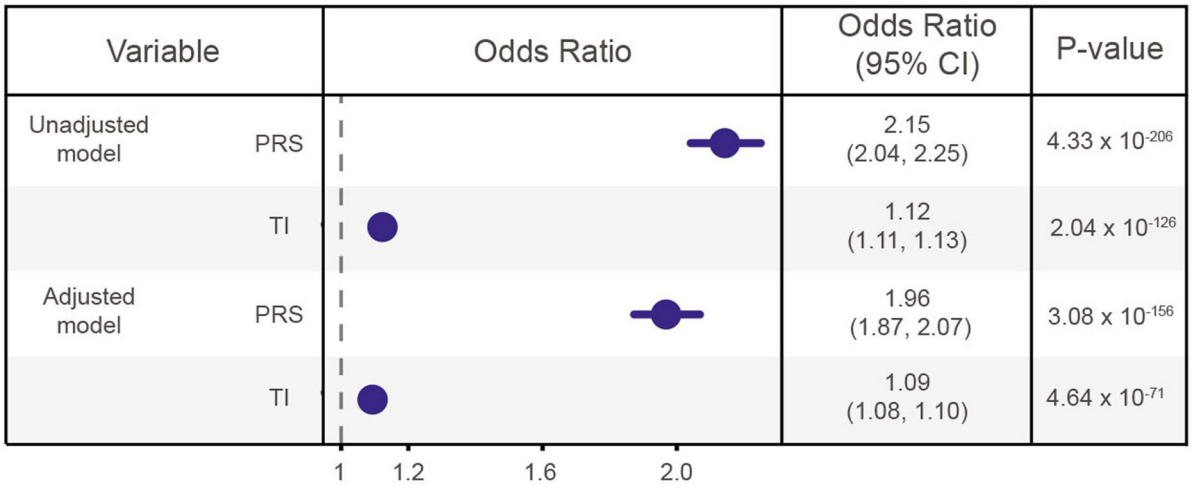

Figure 3.

Genetic risk (PRS) and socioeconomic deprivation (TI) associations with type 2 diabetes (T2D). Forest plot showing ORs, CIs and p values for PRS and TI associations with T2D. Model specifications: unadjusted PRS model T2D~PRS+age+sex; unadjusted TI model T2D~TI+age+sex; adjusted model T2D~PRS+TI+age+sex. PRS, polygenic risk score; TI, Townsend Index.

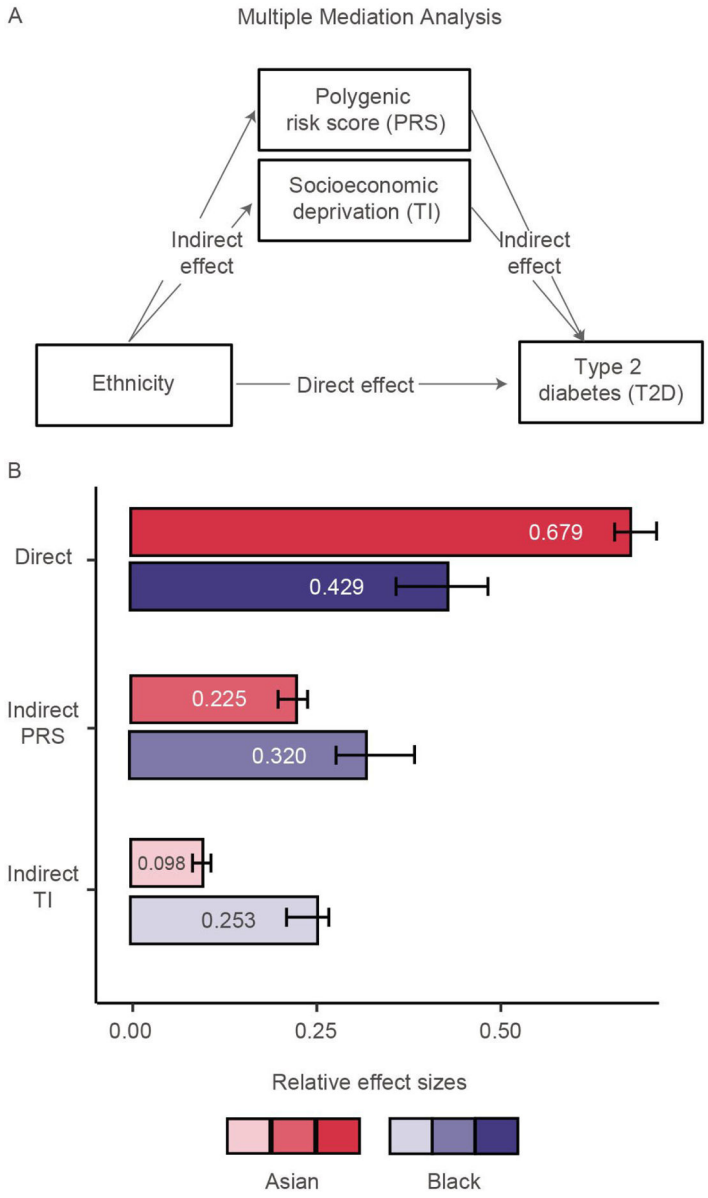

Figure 4.

Multiple mediation analysis of type 2 diabetes (T2D) ethnic disparities. (A) Directed acyclic graph indicating the outcome (T2D), predictor (ethnicity) and mediators (PRS and TI). Age and sex are included as covariates. (B) Relative effect sizes are shown for the direct effect and the indirect effects mediated by PRS and TI, for Asian (red) and black (blue) groups. PRS, polygenic risk score; TI, Townsend Index.

Table 1

UK Biobank type 2 diabetes (T2D) cohort used for this study

	Full cohort	Asian	Black	White
Number	26912	9372 (34.82%)	7540 (28.02%)	10 000 (37.16%)
Age	54.2 (8.42)	53.30 (8.45)	51.89 (8.05)	56.77 (7.99)
Male	12 952 (48.13%)	5064 (54.03%)	3233 (42.88%)	4655 (46.55%)
Female	13 960 (51.88%)	4308 (45.97%)	4307 (57.12%)	5345 (53.45%)
T2D prevalence	15.52%	23.34%	16.64%	7.35%

Table 2

Type 2 diabetes (T2D) ethnic disparities

	Asian	Black	White
Estimate	1.64	1.34	Reference
SE	0.05	0.05	Reference
Z-value	34.25	25.74	Reference
P value	4.57e-257	4.71e-146	Reference
OR (95% CI)	5.14 (4.68 to 5.65)	3.81 (3.44 to 4.22)	Reference

Model specification: T2D~ethnicity+age+sex, with the majority white ethnic group used as the reference group.