



Leveraging family history data to disentangle time-varying effects on disease risk using lifecourse mendelian randomization

Tom G Richardson¹ · Helena Urquijo¹ · Michael V Holmes¹ · George Davey Smith¹

Received: 16 July 2022 / Accepted: 31 March 2023 / Published online: 8 May 2023
© The Author(s) 2023

Abstract

Lifecourse Mendelian randomization is a causal inference technique which harnesses genetic variants with time-varying effects to develop insight into the influence of age-dependent lifestyle factors on disease risk. Here, we apply this approach to evaluate whether childhood body size has a direct consequence on 8 major disease endpoints by analysing parental history data from the UK Biobank study.

Our findings suggest that, whilst childhood body size increases later risk of outcomes such as heart disease (odds ratio (OR) = 1.15, 95% CI = 1.07 to 1.23, $P = 7.8 \times 10^{-5}$) and diabetes (OR = 1.43, 95% CI = 1.31 to 1.56, $P = 9.4 \times 10^{-15}$) based on parental history data, these findings are likely attributed to a sustained influence of being overweight for many years over the lifecourse. Likewise, we found evidence that remaining overweight throughout the lifecourse increases risk of lung cancer, which was partially mediated by lifetime smoking index. In contrast, using parental history data provided evidence that being overweight in childhood may have a protective effect on risk of breast cancer (OR = 0.87, 95% CI = 0.78 to 0.97, $P = 0.01$), corroborating findings from observational studies and large-scale genetic consortia.

Large-scale family disease history data can provide a complementary source of evidence for epidemiological studies to exploit, particularly given that they are likely more robust to sources of selection bias (e.g. survival bias) compared to conventional case control studies. Leveraging these data using approaches such as lifecourse Mendelian randomization can help elucidate additional layers of evidence to dissect age-dependent effects on disease risk.

Keywords Childhood body size · Mendelian randomization · Lifecourse epidemiology · Family disease history · UK Biobank

Introduction

Disentangling causal from correlated risk factors which can vary over the lifecourse is a challenging and arduous task in a conventional epidemiological setting. Overcoming these obstacles is central to the conception and implementation of an approach known as Mendelian randomization (MR), a causal inference method which harnesses genetic variants as instrumental variables to estimate the effect of risk factors on disease outcomes [1, 2]. MR exploits the properties

of naturally occurring genetic variants which are typically fixed at conception, meaning that findings derived from this approach are more robust to confounding factors and reverse causation than findings from conventional observational epidemiological studies.

Recent findings emerging from the literature suggest that types of selection bias can hinder MR investigations, including survival bias which may distort findings when an outcome is measured in a nonrandom subset of the population who have survived long enough to be recruited into a study [3]. In this short communication, we propose the use of parental disease history data to help alleviate this source of bias in MR studies, given that the parents of individuals who have been diagnosed with a given disease will be considered a case regardless of their age at death. Furthermore, a recent study reported comparable results using case definitions based on family disease history in the UK Biobank (UKB) as when defining cases based on combined hospital

✉ Tom G Richardson
Tom.G.Richardson@bristol.ac.uk

¹ MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK

records and questionnaire data, as well as increased statistical power for certain endpoints when using family history information [4].

As an exemplar to demonstrate the value of analysing disease outcome data from first-degree relatives, we have investigated the genetically predicted effects of childhood body size on 8 major disease endpoints recorded for the parents of participants in the UKB (Supplementary Table 1). In doing so, we exploit the predictable genetic association between generations as a proxy for genotype-outcome estimates in measured cases, previously referred to as ‘proxy-genotype Mendelian randomization’ [5]. Findings were initially evaluated with univariable MR (Fig. 1A) and subsequently using a multivariable framework to estimate the direct and indirect effects of childhood body size on disease endpoints whilst accounting for the effect of adulthood body size (referred to as ‘lifecourse MR’ [6](Fig. 1B, C)).

Methods

Childhood and adult body size instrumental variables

Genetic instruments for childhood and adult body size were derived from a large-scale GWAS in the UKB conducted previously [7]. Full details of the GWAS protocol can be found in **Supplementary Note**. Linkage disequilibrium (LD) clumping was applied to identify our instruments using parameters of $P < 5 \times 10^{-8}$ and $r^2 < 0.001$ based on a reference panel based on 10,000 unrelated participants of European descent from UKB [8]. The final sets of genetic instruments can be found in Supplementary Table 2. These instruments have been validated in three independent populations which demonstrate their capability to reliably separate measured body mass index from childhood and adult timepoints as discussed in **Supplementary Note**.

Furthermore, a recent study has found that the childhood genetic instruments have a much stronger effect on DXA-derived fat mass in early life compared to DXA-derived lean mass [9].

Genetic estimates of disease outcomes using data on first-degree relatives

Reported illnesses of mothers (field 20110) and fathers (field 20107) were recorded in the UKB study by the majority of participants ($n=492,986$ for maternal history and $n=488,077$ for paternal history). Amongst these endpoints were; bowel cancer, breast cancer (mothers only), diabetes, heart disease, high blood pressure, lung cancer, prostate cancer (fathers only) and stroke. All outcomes were coded as 0 = neither parent with reported disease, 1 = one parent with disease and 2 = both parents with disease, with the exception of breast cancer and prostate cancer which was encoded as binary outcomes depending on whether mothers or fathers respectively had reportedly had these diseases. These fields in the UKB study were for blood relatives only as adopted mothers and fathers had separate fields for reported disease history (fields 20112 and 20113). If participants were unsure about any answers they were encouraged to respond with ‘do not know’. A summary of final sample sizes can be found in Supplementary Table 1. GWAS were applied to these outcome variables using the same protocol found in **Supplementary Note** to derive estimates for subsequent MR analyses.

Statistical analysis

Mendelian randomization

Univariable MR analyses were initially undertaken to systematically estimate the total effect of genetically predicted exposures on each parentally proxied disease endpoint in

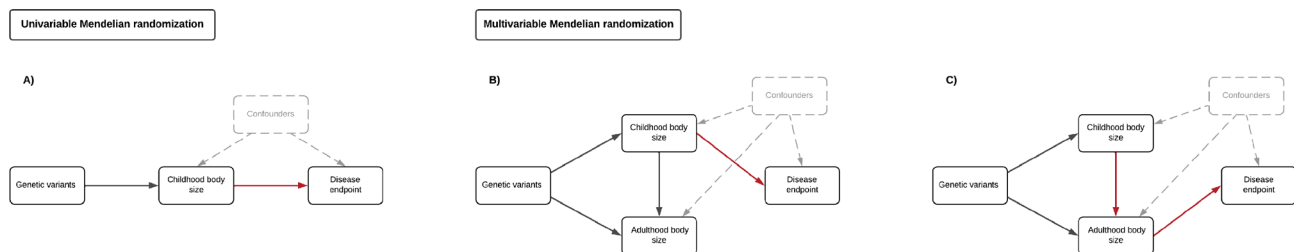


Fig. 1 Schematic representation of applying (A) Univariable Mendelian randomization to estimate the ‘total effect’ of childhood body size on disease risk and and multivariable Mendelian randomization to separately estimates the (B) ‘direct effect’ and (C) ‘indirect effect’ of childhood body size on disease risk whilst accounting for the effect of adulthood body size (known as ‘lifecourse Mendelian randomization’). For example, previous applications of this approach have suggested that childhood body size has a direct effect (B) on risk of type

1 diabetes [6], but an indirect effect (C) on risk of type 2 diabetes [7]. These findings can be interpreted as indicating that being overweight in childhood exerts an effect in early life on risk of type 1 diabetes, whereas its influence on risk of type 2 diabetes is likely attributed to a sustained effect of remaining overweight at later stages of the life-course. The red arrows represent the causal pathway being evaluated in each scenario

turn. This was firstly conducted using the inverse variance weighted (IVW) method, which takes the SNP-outcome estimates and regresses them on those for the SNP-exposure associations. We subsequently applied the weighted median and MR-Egger methods which are more robust to horizontal pleiotropy than the IVW approach [2].

We next conducted multivariable MR to estimate the direct and indirect effects of exposures on disease endpoints which provided evidence of an effect based on $FDR < 5\%$ from IVW univariable analyses. Multivariable MR involves obtaining estimates for all instruments on each exposure being evaluated, thus allowing each estimated effect to take into account the effect of all other exposures in the model. Although this approach has been conventionally applied to analyse different risk factors as exposures (where estimates are typically interpreted as ‘lifelong effects’), the novelty of analysing the same exposure measured at different time-points throughout the lifecourse (e.g. at age 10 and age 55 as conducted here) can facilitate inference in a lifecourse epidemiology setting. All analyses in this study were undertaken using R (version 3.5.1).

Results

Applying univariable MR to parentally proxied outcomes provided evidence that childhood body size increases risk of disease endpoints such as heart disease ($OR = 1.15$, $95\% CI = 1.07$ to 1.23 , $P = 7.8 \times 10^{-5}$) and diabetes ($OR = 1.43$, $95\% CI = 1.31$ to 1.56 , $P = 9.4 \times 10^{-15}$) (Supplementary Table 3). However, effect estimates attenuated to be close to the null upon accounting for adulthood body size in a multivariable MR setting. This is in line with previous investigations, which suggest that childhood body size has no direct influence on these disease outcomes conditional on adulthood body size [7] (Supplementary Tables 4 & Fig. 2). Similarly, our results suggest that the genetically predicted effect of childhood body size on risk of parentally proxied lung cancer is likely attributed to individuals remaining overweight into adulthood (Fig. 2). We further investigated lifetime smoking as an additional exposure in our model, which we hypothesised likely resides along the causal pathway between body size and lung cancer risk as previously proposed [10]. Results supported this hypothesis as the effect of adulthood body size additionally attenuated upon accounting for the effect of lifetime smoking ($OR = 1.11$, $95\% CI = 0.99$ to 1.25 , $P = 0.08$). Conversely, there was strong evidence of an effect of lifetime smoking on lung cancer risk whilst accounting for both childhood and adult body size ($OR = 2.85$, $95\% CI = 2.42$ to 3.35 , $P = 2.7 \times 10^{-36}$), suggesting that smoking mediates some of the effect of body size on lung cancer risk (Supplementary Table 5).

In contrast, we found evidence of a direct effect of childhood body size on risk of maternally proxied breast cancer ($OR = 0.87$, $95\% CI = 0.78$ to 0.97 , $P = 0.01$) after accounting for the genetically predicted effect of adulthood body size as has been reported previously using findings from a large-scale consortium [7] (Fig. 2). We also found evidence of an indirect effect of childhood body size on paternally proxied prostate cancer risk via the pathway involving adulthood body size ($OR = 0.82$, $95\% CI = 0.74$ to 0.91 , $P = 2.1 \times 10^{-4}$). However, this finding requires further evaluation given that it has not been validated using data from the largest available prostate cancer consortium [7], which may potentially be explained by the paternal cases analysed in this study having a comparatively older age distribution compared to the consortium cases.

Discussion

Our systematic evaluation of 8 major disease outcomes based on family disease history data using a lifecourse MR approach provides corroborating evidence into the long-term consequences of childhood body size. Such investigations would be challenging to undertake without the use of time-varying genetic variants harnessed as instrumental variables given the propensity of observational studies to be biased by confounding factors and reverse causation over the lifecourse. This study design using parental data also mitigates the influence of survival bias, which in particular emphasises the importance of developing insight into the aetiological relationship between childhood body size and breast cancer [11]. Furthermore, this approach may pave the way for mechanistic understanding into epidemiological relationships such as the effect of lifelong adiposity on lung cancer risk, which our findings suggest may be partly mediated by smoking.

There are however caveats to using disease history data in first-degree relatives with MR, such as the interpretation of effect estimates which in theory should be halved given that participants will on average share 50% of their DNA with individuals for whom outcomes occur [12]. For example, the multivariable MR estimate for adulthood body size on risk of diabetes using parental history data had a central effect estimate of $OR = 1.97$, which is approximately half the estimate reported previously using large-scale case control data ($OR = 3.90$) [6]. Supplementary Fig. 1 illustrates a side-by-side comparison of estimates derived in this study on parental endpoints with those from large-scale consortia.

Recent methodological developments to integrate individual-level case-control and family history data, such as the application of liability threshold modeling [13], may help improve the statistical power of downstream analyses

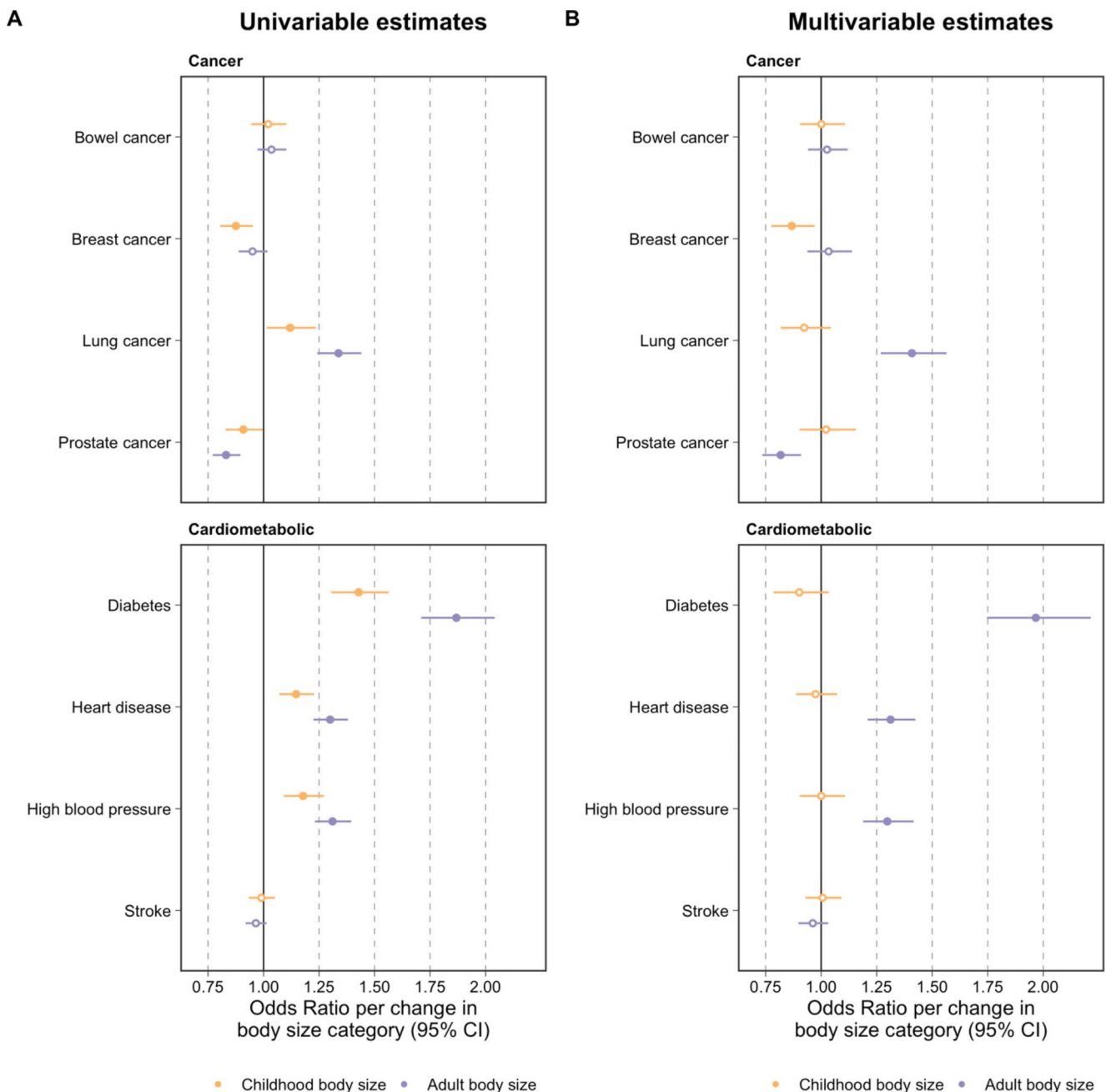


Fig. 2 Univariable and multivariable Mendelian randomization estimates for childhood (yellow) and adult (purple) body size on risk of 8 major disease endpoints using parental history as proxy outcomes in the UK Biobank study

such as MR. This is particularly attractive given that case numbers may be higher for disease outcomes in parents compared to individuals enrolled in a cohort, which has been exploited by genetic consortia for endpoints such as Alzheimer's disease [14]. Future research is required to investigate the most appropriate manner to derive estimates using MR when outcomes are based on self- and parental reported endpoints. Lastly, these methods and the approach taken in this study rely on large-scale biobanks collecting data on family history data as pioneered by the UK

Biobank. Where available these data provide a compelling source of evidence to triangulate findings from conventional MR investigations and therefore improve the robustness of investigations into lifecourse epidemiological relationships.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10654-023-01001-8>.

Acknowledgements We would like to thank the participants of the UK Biobank study for making this research possible. GDS conducts research at the NIHR Biomedical Research Centre at the University

Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Tom G Richardson and Helena Urquijo. The first draft of the manuscript was written by Tom G Richardson and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Integrative Epidemiology Unit which receives funding from the UK Medical Research Council and the University of Bristol (MC_UU_00011/1). HU is supported by a grant from the British Heart Foundation (BHF) (grant FS/17/60/33474).

Data Availability All data on genetic instruments used in this study are located in Supplementary Tables 2 and the full genome-wide study summary statistics on parental outcomes will be made available via the GWAS catalog upon acceptance of publication.

Declarations

Competing interests TGR is employed by GlaxoSmithKline outside of this work. MVH is employed by 23andMe outside of this work and holds stock in the company. All other authors declare no conflicts of interest.

Ethics approval Ethical approval for the UK Biobank was obtained from the Research Ethics Committee (REC; approval number: 11/NW/0382). All analyses were undertaken under UKB application #15825.

Consent to participate Informed consent was collected from all participants whose data was analysed in this study.

Consent to publish All study participants consent to having their data published in journal articles.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Davey Smith G, Ebrahim S. Mendelian randomization: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22.
- Richmond RC, Davey Smith G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harb Perspect Med.* 2022;12(1).
- Smit RAJ, Trompet S, Dekkers OM, Jukema JW, le Cessie S. Survival Bias in mendelian randomization studies: a threat to causal inference. *Epidemiology.* 2019;30(6):813–6.
- DeBoever C, Tanigawa Y, Aguirre M, McInnes G, Lavertu A, Rivas MA. Assessing Digital phenotyping to Enhance Genetic Studies of Human Diseases. *Am J Hum Genet.* 2020;106(5):611–22.
- Barry CJ, Carslake D, Wade KH, Sanderson E, Davey Smith G. Comparison of intergenerational instrumental variable analyses of body mass index and mortality in UK Biobank. *Int J Epidemiol.* 2022.
- Richardson TG, Crouch DJM, Power GM, Morales-Berstein F, Hazelwood E, Fang S, et al. Childhood body size directly increases type 1 diabetes risk based on a lifecourse mendelian randomization approach. *Nat Commun.* 2022;13(1):2337.
- Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. *BMJ.* 2020;369:m1203.
- Kibinge NK, Relton CL, Gaunt TR, Richardson TG. Characterizing the Causal Pathway for Genetic Variants Associated with Neurological Phenotypes Using Human Brain-Derived Proteome Data. *Am J Hum Genet.* 2020.
- Waterfield S, Richardson TG, Davey Smith G, O'Keefe LM, Bell JA. Life course effects of genetic susceptibility to higher body size on body fat and lean mass: prospective cohort study. *Int J Epidemiol.* 2023; <https://doi.org/10.1093/ije/dyad029>
- Taylor AE, Richmond RC, Palviainen T, Loukola A, Wootton RE, Kaprio J, et al. The effect of body mass index on smoking behaviour and nicotine metabolism: a mendelian randomization study. *Hum Mol Genet.* 2019;28(8):1322–30.
- Vabistsevi M, Davey Smith G, Sanderson E, Richardson TG, Lloyd-Lewis B, Richmond RC. Deciphering how early life adiposity influences breast cancer risk using mendelian randomization. *Commun Biol.* 2022;5(1):337.
- Richardson TG, Wang Q, Sanderson E, Mahajan A, McCarthy MI, Frayling TM, et al. Effects of apolipoprotein B on lifespan and risks of major diseases including type 2 diabetes: a mendelian randomisation analysis using outcomes in first-degree relatives. *Lancet Healthy Longev.* 2021;2(6):e317–e26.
- Hujoel MLA, Gazal S, Loh PR, Patterson N, Price AL. Liability threshold modeling of case-control status and family history of disease increases association power. *Nat Genet.* 2020;52(5):541–7.
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet.* 2019;51(3):404–13.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.