

OPEN

A causal relationship between cigarette smoking and type 2 diabetes mellitus: A Mendelian randomization study

Shuai Yuan^{1,2} & Susanna C. Larsson^{1,2*}

The causality between smoking and type 2 diabetes is unclear. We conducted a two-sample Mendelian randomization study to explore the causal relationship between smoking initiation and type 2 diabetes. Summary-level data for type 2 diabetes were obtained from a meta-analysis of 32 genome-wide association studies (DIAbetes Genetics Replication And Meta-analysis consortium), which included 898 130 individuals of European ancestry. Totally, 377 single-nucleotide polymorphisms associated with smoking initiation at genome wide significance threshold ($p < 5 \times 10^{-8}$) were identified from the hitherto largest genome-wide association study on smoking. The inverse-variance weighted, weighted median, MR-Egger regression, and MR-PRESSO approaches were used to analyze the data. Genetically predicted smoking initiation was associated with type 2 diabetes with an odds ratio of 1.28 (95% confidence interval, 1.20, 1.37; $p = 2.35 \times 10^{-12}$). Results were consistent across sensitivity analyses and there was no evidence of horizontal pleiotropy. This study provides genetic evidence supporting a causal association between the smoking initiation and type 2 diabetes. Reducing cigarette smoking initiation can now be even more strongly recommended for type 2 diabetes prevention.

Globally, type 2 diabetes mellitus is a public health issue, affecting around 1 in 11 adults in 2015¹. Cigarette smoking has been proposed as an independent and modifiable risk factor for type 2 diabetes in observational studies among both men and women²⁻⁴. Results of a recent meta-analysis of 88 observational prospective studies, including 5 898 795 participants and 295 446 incident cases of type 2 diabetes, showed that current smoking was associated with a 37% increased risk of developing type 2 diabetes compared with non-smoking³. A dose-response relationship was identified across light, moderate and heavy smokers³. Nevertheless, whether the association between smoking and type 2 diabetes is causal remains unclear due to potential residual confounding and reverse causality that could bias the results from observational studies. In addition, the causality is not feasible to be tested in the experimental settings considering the ethical issues.

Mendelian randomization (MR) is a technique that minimizes unobserved confounding and reverse causality by proposing genetic variants as instrumental variables of an exposure⁵. Genetic variants have no relevance with self-adapted lifestyle factors and behaviors as they are randomly assorted at meiosis. Reverse causality can be eliminated as allelic randomization precedes disease onset. Therefore, we conducted a MR study to assess the associations between smoking initiation and risk of type 2 diabetes.

Methods

Study design overview. We employed an MR design to investigate the association between smoking initiation and type 2 diabetes using summary-level data from two large genome-wide association studies for tobacco use and type 2 diabetes. Details of used datasets are shown in Supplementary Table 1. There are three key assumptions for an MR study⁶. These include: (1) the genetic variants selected as instrumental variables should be strongly associated with the exposure; (2) the instrumental variables should not be associated with confounders of the exposure-outcome association; (3) the instrumental variables should affect the risk of the outcome merely through the risk factor, not via alternative pathways. Ethical approval for this study was obtained from the

¹Unit of Cardiovascular and Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. ²Department of Surgical Sciences, Uppsala University, Uppsala, Sweden. *email: susanna.larsson@ki.se

Swedish Ethical Review Authority. This MR study only uses published or publicly available summary-level data (i.e., beta coefficients and standard errors). Ethical approval (and informed consent from each participant) for each study included in the genome-wide association studies of tobacco use and type 2 diabetes can be found in the original articles^{7,8}. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional or national research committee and with the 1964 Helsinki declaration.

Outcome sources and SNP selection. Summary-level data for type 2 diabetes were obtained from a publicly available genome-wide association study (GWAS) of 32 studies (DIAbetes Genetics Replication And Meta-analysis consortium), which included 898 130 individuals (74 124 cases and 824 006 controls) of European ancestry⁷. A harmonized protocol was developed to improve the quality of the genotype scaffold in each study. The HRC reference panel was used in the imputation stage and adjustments were made for population structure (e.g., through principal components), relatedness and study-specific covariates. Data without body-mass index adjustment were used in the main analysis, and data adjusted for body-mass index were used in the sensitivity analysis in the present study.

Selection of instrumental variables was based on a recent published meta-analysis of GWASs for smoking initiation, which included data from up to 1 232 091 individuals of European ancestry⁸. In total, 378 conditionally independent single-nucleotide polymorphisms (SNPs) associated with smoking initiation at the genome wide significance threshold ($p < 5 \times 10^{-8}$) were identified⁸ of which all but one SNP was available in the type 2 diabetes dataset. Adjustments had been made for genetic principal components⁸. Detailed information for each SNP is shown in Supplementary Table 2. Smoking initiation was defined as having smoked >100 cigarettes over the course of your life, smoked every day for at least a month or ever smoked regularly. The GWAS for smoking initiation reported the effect sizes in the unit of standard deviation, which was calculated from the weighted average prevalence across all included studies⁸.

Statistical analyses. The association between genetically predicted smoking initiation and type 2 diabetes attributable to each SNP was estimated with the Wald method, which computes the ratio between the SNP-diabetes and SNP-smoking estimates. In the main analysis, the ratio estimates for individual SNPs were combined by using the multiplicative random-effects inverse-variance weighted meta-analysis method⁹. We also performed supplementary analysis based on the weighted median, MR-Egger (with or without adjustment via Simulation Extrapolation [SIMEX] method), MR-PRESSO approaches, to examine the robustness of the association and assess whether the MR assumption of no pleiotropy is met. The inverse-variance weighted method provides the most precise estimates but could be influenced by invalid instrumental variables and pleiotropic effects⁹. The weighted median approach provides a consistent estimate under the requirement that more than half of the weight in the analysis comes from valid instrumental variables¹⁰. The MR-Egger method identifies and corrects for directional pleiotropy, albeit with low power¹¹. MR-Egger regression with SIMEX adjustment was used as a sensitivity analysis as regression dilution bias in the SNP-smoking estimates was observed in the standard MR-Egger analysis ($I^2_{GX} < 90\%$)¹². The MR pleiotropy residual sum and outlier (MR-PRESSO) test aims at detecting possible outliers and results obtained from the MR-PRESSO analysis are corrected for horizontal pleiotropy via outlier removal¹³. Rucker's Q' value was additionally estimated to measure the heterogeneity in the MR-Egger analysis and as a comparison with Cochran's Q value. A Rucker's Q' value that is lower than Cochran's Q value indicates that the MR-Egger method provides a model with better fit for examining the particular association.

We harmonized the summary statistics data across datasets so that the effect allele reflected the allele associated with an increased probability of lifetime smoking initiation. When SNPs were palindromic (i.e., A/T or G/C), we used information on allele frequency to resolve strand ambiguity. Ten palindromic SNPs with minor allele frequency above 0.45 were retained because the SNPs were read from the same strand, and exclusion of those ten palindromic SNPs did not change the results. We estimated mean F-statistics to assess the strength of the instrumental variables¹⁴ and got an F-statistic of 76.7. The reported odds ratios (ORs) and confidence intervals (CIs) of type 2 diabetes correspond to the increase of one standard deviation in prevalence of smoking initiation. All estimates are reported with two-tailed P values. The statistical analyses were performed in Stata/SE 15.0 using the `mrr` package¹⁵, except the MR-PRESSO analysis which was conducted in R software 3.6.0.

Results

Genetically predicted smoking initiation was positively associated with type 2 diabetes. The ORs of type 2 diabetes were 1.28 (95% CI, 1.20, 1.37; $p = 2.35 \times 10^{-12}$) and 1.29 (95% CI, 1.20, 1.38; $p = 4.22 \times 10^{-13}$) in the inverse-variance weighted and weighted median models, respectively (Fig. 1). There was substantial heterogeneity across estimates of included SNPs with an I^2 value of 70% (95% CI 67%, 73%) and a Cochran's Q value of 1255 ($p < 0.001$). The MR-Egger estimate without SIMEX adjustment was directionally consistent with the other estimates, albeit non-significant with a wide CI. We observed an I^2_{GX} of 61%, indicating dilution of the MR-Egger estimate due to violation of the No Measurement Error assumption. In a sensitivity analysis to correct for dilution bias using SIMEX adjustment, the OR was 1.22 (95% CI, 0.72, 2.09; $p = 0.457$). There was no indication of horizontal pleiotropy in the MR-Egger analysis (intercept 0.001; 95% CI -0.004, 0.007; $p = 0.600$). We observed a Rucker's Q' value of 1254 ($p < 0.001$), indicating that the MR-Egger approach did not provide a model with better fit compared to the inverse-variance weighted method. In the MR-PRESSO analysis, we detected 19 possible outliers. After outlier correction, the OR of type 2 diabetes was 1.28 (1.20, 1.35; $p = 2.16 \times 10^{-14}$) (Fig. 1). A scatter plot for the associations of the smoking-related SNPs with smoking initiation and type 2 diabetes is shown in Supplementary Fig. 1. Results remained in the same pattern in the sensitivity analysis based on data with body mass index adjustment (Supplementary Table 3).

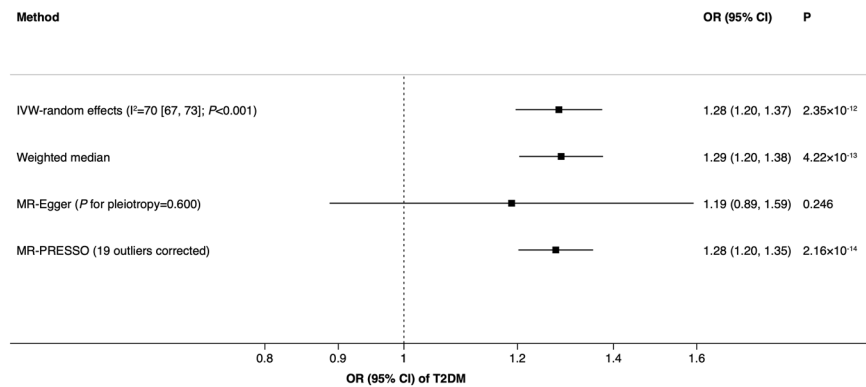


Figure 1. Association between smoking initiation and type 2 diabetes from Mendelian randomization. CI indicates confidence interval; OR, odds ratio; T2DM, type 2 diabetes mellitus.

Discussion

Findings of the present study support a causal association between smoking initiation and increased risk of type 2 diabetes. Most available studies found a consistent detrimental effect of current and past smoking on type 2 diabetes^{2–4}, which is in line with the present study. A systematic review including 1.2 million participants revealed a dose-response relationship between active smoking and risk of type 2 diabetes among the pooled whole population and all subgroups². Although several large-scale studies indicate that smoking cessation increases the short-term risk of type 2 diabetes, which is mediated by body mass index, the risk decreases substantially with time since quitting smoking^{3,16}.

The mechanisms behind the causality between smoking and type 2 diabetes are not fully understood. There are several plausible explanations, such as negative effects of cigarette smoking on the function¹⁷ and mass of islet β -cells¹⁸, gastrointestinal tract¹⁹, nervous system^{20,21}, obesity²² and inflammation²³. Nicotine, a major bioactive element of cigarette, has been proved to impair the function and mass of the islet β -cells^{17,18}, thereby disturbing its feedback regulation and interrupting glucose homeostasis, which plays an important role of type 2 diabetes onset²⁴. Smoking also negatively influences the function of gastrointestinal tract¹⁹, such as suppressing bile acids, which is of great importance in the regulation of glucose metabolism²⁴. Recently, smoking has been found to be associated with change of composition of intestinal microbiome²⁵ that potentially acts as a vital part in the pathophysiology of type 2 diabetes²⁴. In addition, cigarette smoking also influences the function of the nervous system, such as vagus²⁰, hypothalamus²¹, and circadian rhythmicity²⁶, which are important regulators of glucose metabolic processes²⁴. Inflammation induced by smoking also partly explains the causality²³. Proinflammatory factors and C-reactive protein overproduction is associated with insulin resistance, β -cells function impairment and metabolism-related neuronal injury²⁴. Even though these pathological pathways have been established to explain the causality between smoking and type 2 diabetes, more investigations on its etiology are needed, especially from genetics, epigenetics and omics²⁴, for type 2 diabetes prevention and treatment.

A major limitation is that the dose-response relation of smoking heaviness (e.g., the number of cigarettes smoked per day) with type 2 diabetes could not be assessed in the present study because we could not exclude never smokers based on summary-level data. Another limitation is that there was a large overlap of participants included in the datasets for smoking initiation and type 2 diabetes, potentially leading to bias in the causal estimate in the direction of the observational association between smoking initiation and risk of type 2 diabetes²⁷.

The validity of the results of an MR study depends on whether the MR assumptions are met. In this study, we only used SNPs that are strongly associated with smoking initiation at the genome-wide significance level, thereby reducing possible violation of the first assumption. A strong risk factor and thus a potential confounder in analyses of type 2 diabetes is body mass index. A positive association between smoking and type 2 diabetes remained when using data with body mass index adjustment. In addition, the consistency across sensitivity analyses indicated a negligible distortion by potential pleiotropy.

Conclusion

This study provides genetic evidence supporting a positive causal association between the smoking initiation and type 2 diabetes. Thus, by more firm scientific support, reducing cigarette smoking initiation for type 2 diabetes prevention can now be even more strongly recommended. The mechanisms behind the relationship warrant more investigations.

Data availability

All summary-level data necessary to conduct this MR analysis are included in Supplementary Table 2.

Received: 20 May 2019; Accepted: 5 December 2019;

Published online: 18 December 2019

References

- Zheng, Y., Ley, S. H. & Hu, F. B. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol.* **14**, 88–98 (2018).
- Willi, C., Bodenmann, P., Ghali, W. A., Faris, P. D. & Cornuz, J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. *JAMA.* **298**, 2654–2664 (2007).

3. Pan, A., Wang, Y., Talaei, M., Hu, F. B. & Wu, T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol.* **3**, 958–967 (2015).
4. Yuan, S. *et al.* Cigarette smoking as a risk factor for type 2 diabetes in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *J Public Health.* **41**, e169–e176 (2018).
5. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey-Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med.* **27**, 1133–1163 (2008).
6. Davies, N. M., Holmes, M. V. & George, D. S. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ.* **362**, k601 (2018).
7. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* **50**, 1505–1513 (2018).
8. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet.* **51**(2), 237–244 (2019).
9. Burgess, S., Bowden, J., Fall, T., Ingelsson, E. & Thompson, S. G. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. *Epidemiology.* **28**, 30–42 (2017).
10. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol.* **40**, 304–314 (2016).
11. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol.* **44**, 512–525 (2015).
12. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol.* **45**, 1961–1974 (2016).
13. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol.* **32**, 377–389 (2017).
14. Burgess, S., Thompson, S. G. & CHD, C. R. P. Genetics Collaboration. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol.* **40**, 755–64 (2011).
15. Spiller, W., Davies, N. M. & Palmer, T. M. Software application profile: mrrobust—a tool for performing two-sample summary Mendelian randomization analyses. *Int J Epidemiol.* **48**, 684–690 (2019).
16. Hu, Y. *et al.* Smoking Cessation, Weight Change, Type 2 Diabetes, and Mortality. *N Engl J Med.* **379**, 623–632 (2018).
17. Yoshikawa, H., Hellstrom-Lindahl, E. & Grill, V. Evidence for functional nicotinic receptors on pancreatic beta cells. *Metabolism.* **54**, 247–254 (2005).
18. Somm, E. *et al.* Prenatal nicotine exposure alters early pancreatic islet and adipose tissue development with consequences on the control of body weight and glucose metabolism later in life. *Endocrinology.* **149**, 6289–6299 (2008).
19. Thomas, G. A., Rhodes, J. & Ingram, J. R. Mechanisms of disease: nicotine—a review of its actions in the context of gastrointestinal disease. *Nat Clin Pract Gastroenterol Hepatol.* **2**, 536–544 (2005).
20. Floto, R. A. & Smith, K. G. The vagus nerve, macrophages, and nicotine. *Lancet.* **361**, 1069–1070 (2003).
21. Kroemer, N. B., Guevara, A., Vollstadt-Klein, S. & Smolka, M. N. Nicotine alters food-cue reactivity via networks extending from the hypothalamus. *Neuropsychopharmacology.* **38**, 2307–2314 (2013).
22. Carreras-Torres, R. *et al.* Role of obesity in smoking behaviour: Mendelian randomisation study in UK Biobank. *BMJ.* **361**, k1767 (2018).
23. Goncalves, R. B. *et al.* Impact of smoking on inflammation: overview of molecular mechanisms. *Inflamm Res.* **60**, 409–424 (2011).
24. Kahn, S. E., Cooper, M. E. & Del Prato, S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet.* **383**, 1068–1083 (2014).
25. Savin, Z., Kivity, S., Yonath, H. & Yehuda, S. Smoking and the intestinal microbiome. *Arch Microbiol.* **200**, 677–684 (2018).
26. Bartel, K. A., Gradisar, M. & Williamson, P. Protective and risk factors for adolescent sleep: a meta-analytic review. *Sleep Med Rev.* **21**, 72–85 (2015).
27. Burgess, S., Davies, N. M. & Thompson, S. G. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol.* **40**(7), 597–608 (2016).

Acknowledgements

The authors thank the DIAGRAM consortium for providing summary-level data for type 2 diabetes. Open access funding provided by Karolinska Institute.

Author contributions

S.Y. analyzed the data and wrote the manuscript. S.C.L. designed the study, managed and interpreted the data and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-56014-9>.

Correspondence and requests for materials should be addressed to S.C.L.

Reprints and permissions information is available at www.nature.com/reprints.

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



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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019