STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies[1 2]

Item No.	Section	Checklist item	Page No.	Relevant text from manuscript
1	**TITLE and ABSTRACT**	Indicate Mendelian randomization (MR) as the study's design in the title and/or the abstract if that is a main purpose of the study	1	Causal association of gut microbiota with type 2 diabetes, type 1 diabetes and glycemic traits: a two-sample Mendelian randomization study.
	INTRODUCTION			
2	**Background**	Explain the scientific background and rationale for the reported study. What is the exposure? Is a potential causal relationship between exposure and outcome plausible? Justify why MR is a helpful method to address the study question	4-5	The exposure in this research was gut microbiota …as a means of reducing confounding in examining exposure–disease associations.
3	**Objectives**	State specific objectives clearly, including pre-specified causal hypotheses (if any). State that MR is a method that, under specific assumptions, intends to estimate causal effects	5	This research used the genetic variants for gut microbiota as an instrumental variable to conduct a two-sample bidirectional MR to explore the causality association of gut microbiota with T2D, T1D, fasting glucose (FG), fasting insulin (FI) and 2h-glucose (2h-g).
	METHODS			
4	**Study design and data sources**	Present key elements of the study design early in the article. Consider including a table listing sources of data for all phases of the study. For each data source contributing to the analysis, describe the following:		
	a)	Setting: Describe the study design and the underlying population, if possible. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection, when available.	6	This research used the data from the MiBioGen consortium, the DIAGRAM consortium, genome-wide association studies (GWAS) database and the Meta-Analyses of Glucose and Insulin-related traits Consortium(MAGIC). Tables were not used to present data information.
	b)	Participants: Give the eligibility criteria, and the sources and methods of selection of participants. Report the sample size, and whether any power or sample size calculations were carried out prior to the main analysis	6	The gut microbiota-related GWAS data were obtained from the MiBioGen consortium … from 133,010 (fasting glucose), 108,557 (fasting insulin) and 42,854 (2h-glucose) non-diabetic individuals of European ancestry.
	c)	Describe measurement, quality control and selection of genetic variants	6	Gut microbiota served as …we performed a harmonization process that referred to the human genome reference sequence (build 37) to remove duplicated alleles.

		d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases		Definitions of exposure, outcome, and other variables were not reported in this study.
		e)	Provide details of ethics committee approval and participant informed consent, if relevant	15	This research has stated in the Ethical Approval and Consent to Participate section that all data are publicly available through the original study, and no additional ethical approval is required.
5	**Assumptions**		Explicitly state the three core IV assumptions for the main analysis (relevance, independence and exclusion restriction) as well assumptions for any additional or sensitivity analysis	5	To allow unbiased estimation of the causal effect of the exposure … IVs only associated with the outcome through the exposure.
6	**Statistical methods: main analysis**		Describe statistical methods and statistics used		
		a)	Describe how quantitative variables were handled in the analyses (i.e., scale, units, model)		Scale, units and model were not included in the original data for this study, therefore, not reported.
		b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	6	We calculated the *F*-statistic to estimate the strength of instruments …The remaining SNPs were used for the subsequent MR analysis.
		c)	Describe the MR estimator (e.g. two-stage least squares, Wald ratio) and related statistics. Detail the included covariates and, in case of two-sample MR, whether the same covariate set was used for adjustment in the two samples	7	We conduct five methods, inverse-variance weighted (IVW), MR-Egger, weighted median, weighted mode and simple mode, to access the causal relationship between exposure and outcome. Covariates were not included in this study; furthermore, specific MR statistical methods were described in detail in the text.
		d)	Explain how missing data were addressed		The treatment of missing values was not addressed in this study.
		e)	If applicable, indicate how multiple testing was addressed	8	To avoid false positive errors as far as possible, the Benjamini–Hochberg (BH) method was used to calculate the false discovery rate *P* value (P_{FDR}) to adjust for multiple testing. The adjusted multiple testing significance threshold is $P_{FDR}<0.05/n$, n means the amount of independent bacterial taxa at the corresponding taxonomic level (phylum, class, order, family, and genus). The results were statistically significant at $P<0.05$ and $P_{FDR}<0.05/n$; meanwhile $P<0.05$ but $P_{FDR}>0.05/n$ was supposed to be nominally significant.

7	**Assessment of assumptions**	Describe any methods or prior knowledge used to assess the assumptions or justify their validity	6	In this study, the *F*-statistic was calculated to estimate statistical validity, and no other methods were used.
8	**Sensitivity analyses and additional analyses**	Describe any sensitivity analyses or additional analyses performed (e.g. comparison of effect estimates from different approaches, independent replication, bias analytic techniques, validation of instruments, simulations)	7	Additionally, we also performed sensitivity analysis to evaluate the robustness of the results. MR-Egger intercept and MR-PRESSO were applied to identify the horizontal pleiotropy. We regard both MR-PRESSO global test *P*>0.05 and MR-Egger regression test *P*>0.05 as the absence of pleiotropy. Besides, leave-one-out was performed to verify whether the causal signal was driven by a single SNP.
9	**Software and pre-registration**			
	a)	Name statistical software and package(s), including version and settings used	8	All statistical analyses were completed with R software (version 4.2.1), including “TwoSampleMR” and “MR-PRESSO” packages.
	b)	State whether the study protocol and details were pre-registered (as well as when and where)	8	The study weren’t pre-registered at any platforms.
	RESULTS			
10	**Descriptive data**			
	a)	Report the numbers of individuals at each stage of included studies and reasons for exclusion. Consider use of a flow diagram	8	After rigorous filtering processes, 140 SNPs were …153 SNPs in T2D and 31 SNPs in FI were retained.
	b)	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions)	5-6	The original data for exposures and outcomes did not contain summary statistics for the variables of interest, and only the number of inclusions was reported. The largest-scale genome-wide meta-analysis contains 16S ribosomal RNA (rRNA)… non-diabetic individuals of European ancestry.
	c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	Supplementary Table 1-2	It is not provided in the original research. However, Supplementary Table 1-2 reports detailed information on each SNP, such as locus, allele frequency, etc.
	d)	For two-sample MR:		No information on overlapping populations was reported in this study.

			i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples ii. Provide information on the number of individuals who overlap between the exposure and outcome studies		
11	**Main results**				
		a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale	8, Fig 2	After rigorous filtering processes, 140 SNPs were selected…MR-Egger regression (both MR-PRESSO global test *P*>0.05 and MR-Egger regression test *P*>0.05).
		b)	Report MR estimates of the relationship between exposure and outcome, and the measures of uncertainty from the MR analysis, on an interpretable scale, such as odds ratio or relative risk per SD difference	9-11, Table 1 Supplementary Table 3-8	Based on MR analysis using IVW method, we found a protective effect of the host-genetic driven increase in phylum *Verrucomicrobia* (*OR*=0.91, 95%*CI*, 0.86-0.96, $P = 4.76\times10^{-3}$), whereas genus *Actinomyces* (*OR*=1.13, 95%*CI*, 1.06-1.19, $P=4.21\times10^{-4}$) had a hazardous effect. We found family *Veillonellaceae* (*OR*=0.72, 95%*CI*, 0.60-0.86, $P=1.01\times10^{-3}$) was a significant protective factor for T1D in IVW method. Out of the 3 glycemic traits, the genetically predicted level of gut microbiota was considered…phylum *Proteobacteria*(IVW: *OR*=1.17,95%*CI*,1.06-1.30, $P=9.63\times10^{-3}$) and 2h-g.
		c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period		No relavant.
		d)	Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)	Fig 3-4	
12	**Assessment of assumptions**				
		a)	Report the assessment of the validity of the assumptions	9-11, Table 1 Supplementary Table 3-7	This study reports the validity of the hypotheses in the Results section. For example, the *F*-value for each genetic variant and the heterogeneity of each association were reported.
		b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I^2, Q statistic or E-value)	Table 1 Supplementary Table 3-7	The value of heterogeneity (Q statistic) was reported in this study.
13	**Sensitivity analyses and**				

	additional analyses				
		a)	Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions	9-11, Table 1 Supplementary Table 3-7	This study reports the results of sensitivity analyses in the Results section, including heterogeneity, pleiotropy, and the results of other MR methods (MR-Egger, weighted median, weighted mode and simple mode).
		b)	Report results from other sensitivity analyses or additional analyses	9-11, Table 1 Supplementary Table 3-7	This study reports the false discovery rate (FDR) in the Results section.
		c)	Report any assessment of direction of causal relationship (e.g., bidirectional MR)	11, Supplementary Table 8	To evaluate reverse causality... The detailed statistical results will be shown in Supplementary Table 8.
		d)	When relevant, report and compare with estimates from non-MR analyses		No relavant.
		e)	Consider additional plots to visualize results (e.g., leave-one-out analyses)	Fig 4	Circular Heatmap of MR results with significant and nominally significant causal associations for gut microbiota and outcomes.
	DISCUSSION				
14	**Key results**		Summarize key results with reference to study objectives	11	In this research, we aimed to use two-sample MR to evaluate … achieve a new breakthrough in prevention and treatment field as earlier risk identification of T2D, T1D and glycemic traits in the future.
15	**Limitations**		Discuss limitations of the study, taking into account the validity of the IV assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias and any efforts to address them	13-14	Also, some limitations of our research should be fully considered. First ….other nominally significant estimates should be treated with caution.
16	**Interpretation**				
		a)	Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies	11-13	This study interpreted the MR results in the Discussion section by comparing them with several published studies. Recently, researches are being conducted on changes in gut microbiota so that new insights can be gained into…these experimental or observational studies greatly add credibility to our results

		b) Mechanism: Discuss underlying biological mechanisms that could drive a potential causal relationship between the investigated exposure and the outcome, and whether the gene-environment equivalence assumption is reasonable. Use causal language carefully, clarifying that IV estimates may provide causal effects only under certain assumptions	11-13	This study also complemented the MR results in terms of mechanism in the Discussion section. In this field, family *Veillonellaceae* have beneficial effects in the maintenance of gut barriers thanks to their production of short-chain fatty acids (SCFA), which could reshape the gut ecology and induce immune modulation. Studies have shown that SCFA reduces epithelial permeability by regulating interleukin (IL)-10 receptor. It is interesting to note that the growth of phylum *Proteobacteria* drives immune dysregulation and promotes intestinal inflammation.
		c) Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions	11-13	This study also complemented the MR results in terms of clinical trials in the Discussion section. Recently, researches are being conducted on changes in gut microbiota so that new insights can be gained into…these experimental or observational studies greatly add credibility to our results
17	**Generalizability**	Discuss the generalizability of the study results (a) to other populations, (b) across other exposure periods/timings, and (c) across other levels of exposure	13	First, the vast majority of the samples are restricted to individuals of European ancestry, restricting the generalizability of the results to this ethnic group.
	OTHER INFORMATION			
18	**Funding**	Describe sources of funding and the role of funders in the present study and, if applicable, sources of funding for the databases and original study or studies on which the present study is based	16	This research was funded by Key Research and Development Program and Promotion project in Henan Province (grant No: 222102310268). And the funding sources had no involvement in study design, collection, analysis, interpretation of data, writing of the report, or the decision to submit the article for publication.
19	**Data and data sharing**	Provide the data used to perform all analyses or report where and how the data can be accessed, and reference these sources in the article. Provide the statistical code needed to reproduce the results in the article, or report whether the code is publicly accessible and if so, where	16	The datasets analyzed in the current study can be downloaded from the website https://mibiogen.gcc.rug.nl, http://diagram-consortium.org/downloads.html, https://www.ebi.ac.uk/gwas/publications/32005708 http://magicinvestigators.org/downloads/
20	**Conflicts of Interest**	All authors should declare all potential conflicts of interest	16	No potential of interest disclosures were reported by the author(s).

1. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) Statement. JAMA. 2021;under review.

2. Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA, VanderWeele TJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomisation (STROBE-MR): Explanation and Elaboration. BMJ. 2021;375:n2233.