Supplementary Material

STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies

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	Investigation of the association between air pollution and non-alcoholic fatty liver disease in the European population: a Mendelian randomization study.
INTRODUCTION				
2	Background	Explain the scientific background and rationale for the reported study. Is causality between exposure and outcome plausible? Justify why MR is a helpful method to address the study question.	1,2,10	Introduction, Paragraph 1-4; Discussion, Paragraph 1.
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	2.3	Introduction, Paragraph 4.
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.	3,4	Materials and Methods, Section “Study design", Paragraph 1; Section “Data sources", Paragraph 1.
	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	4	Materials and Methods, Section “Data sources", Paragraph 1,2; Supplemental Material, Table S1.
	c)	Describe measurement, quality control and selection of genetic variants	4,5	Materials and Methods, Section “Selection of instrumental variables".
	d)	For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases	4	Materials and Methods, Section “Data sources", Paragraph 1,2; Supplemental Material, Table S1
	e)	Provide details of ethics committee approval and participant informed consent, if relevant	4	Materials and Methods, Section “Data sources", Paragraph 2.
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	3	Materials and Methods, Section “Study design".
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)	4	Materials and Methods, Sections “Data sources" Paragraph 1,2; Supplemental Material, Table S1.
	b)	Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected	4-5	Materials and Methods, Section "Selection of instrumental variables"; Materials and Methods, Section "Mendelian randomization 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	4-5	Materials and Methods, Section “Selection of instrumental variables”, "Mendelian randomization analysis ".
	d)	Explain how missing data were addressed	—	Not applicable to our study.
	e)	If applicable, indicate how multiple testing was addressed	3,6	Materials and Methods, Section “Study design"; Materials and Methods, Section " Statistical analysis".
7	Assessment of assumptions	Describe any methods or prior knowledge used to assess the assumptions or justify their validity	4-6	We selected variants that were associated with the exposures at different GWAS significance. The PhenoScanner and GWASCatalog tools were used to ensure whether the IVs were significantly correlated with the risk factors for NAFLD. We further performed separate UVMR and MVMR analyses to estimate the direct causal effect of air pollution on the risk of NAFLD. The relevant

				information is all reported in the Materials and Methods.
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)	3-6	Materials and Methods, Section “Study design", Paragraph 2; Materials and Methods, Section "Mendelian randomization analysis"; Materials and Methods, Section "Sensitivity Analysis".
9	Software and pre registration			
	a)	Name statistical software and package(s), including version and settings used	6	Materials and Methods, Section "Statistical Analysis".
	b)	State whether the study protocol and details were pre-registered (as well as when and where)	—	No pre-registration.
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	supplementary material	The information about our dataset is given in Table 1 and Table S1. For the exposure datasets and outcome satasets, the summary genetic data could be obtained from the IEU Open GWAS project open to public.
	b)	Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions)		
	c)	If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies	—	Not applicable to our study.
	d)	For two-sample MR:		

		i. Provide justification of the similarity of the genetic variant-exposure associations between the exposure and outcome samples	supplementary material	The information is given in S2.
		ii. Provide information on the number of individuals who overlap between the exposure and outcome studies	supplementary material	The information is given in S4 sheet7.
11	Main results			
	a)	Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale	6-10, supplementary material	
	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	6-10, supplementary material	Results, Sections "UVMR results in the derivation dataset", “MVMR results in the derivation dataset”, “UVMR results in the validation datasets”.
	c)	If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	6-10, supplementary material	The associations between all IVs used in our analyses and our exposures and outcomes are reported in Table 2, 3, S2, S3, S4 and Figure 2, 3, 4.
	d)	Consider plots to visualize results (e.g. forest plot, scatterplot of associations between genetic variants and outcome versus between genetic variants and exposure)	Figure 2, 3, 4	
12	Assessment of assumptions			
	a)	Report the assessment of the validity of the assumptions	6-10, supplementary material	Results, Sections "UVMR results in derivation dataset", “MVMR results in derivation dataset”, “UVMR results in

				validation dataset", Figure 3.
	b)	Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as I2, Q statistic or E-value)	6-10, supplementary material	Other statisitics are reported in Table 2, 3, S4.
13	Sensitivity analyses and additional analyses			
	a)	Use sensitivity analyses to assess the robustness of the main results to violations of the assumptions	6-10	Relevant information is given in the results section.
	b)	Report results from other sensitivity analyses (e.g., replication study with different dataset, analyses of subgroups, validation of instrument(s), simulations, etc)		In addition we performed MVMR model including adjustment for alcohol intake frequency and air pollution and UVMR model with different datasets.
	c)	Report any assessment of direction of causality (e.g., bidirectional MR)		We conducted MR Steiger filtering to monitor the direction of causation.
	d)	When relevant, report and compare with estimates from non-MR analyses		Discussion, Paragraph 1.
	e)	Consider any additional plots to visualize results (e.g., leave-one-out analyses)		A leave-one-out analysis was used for the analysis of the IVW results (Figure 2). We deleted each SNP individually and obtained $P < 0.05$, which is consistent with the results of the IVW method in the analysis of the causal effects, indicating that there were no non-specific SNPs that could have influenced the causal

				estimation results.
DISCUSSION				
14	Key results	Summarize key results with reference to study objectives	13	Conclusion
15		Discuss limitations of the study, taking into account the validity of the MR assumptions, other sources of potential bias, and imprecision. Discuss both direction and magnitude of any potential bias, and any efforts to address them	13	Discussion, Paragraph 7.
16	Interpretation			
	a)	a) Give a cautious overall interpretation of results considering objectives and limitations Compare with results from other relevant studies.	10,11	Discussion, Paragraph 1,2,3.
	b)	b) Discuss underlying biological mechanisms that could be modelled by using the genetic variants to assess the relationship between the exposure and the outcome.	11,12,13	Discussion, Paragraph 3,4,7.
	c)	c) Discuss whether the results have clinical or policy relevance, and whether interventions could have the same size effect	10,11	Discussion, Paragraph 1,2.
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	10-13	Discussion, Paragraph 1, 2,7.

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	13	Section “Funding”.
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	13	Section "Data Availability Statement".
20	Conflicts of Interest	All authors should declare all potential conflicts of interest	13	Section “Conflict of Interest”.