**STROBE-MR checklist of recommended items to address in reports of Mendelian randomization studies**12

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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 | TITLE: Relationship between serum vitamin D levels and thyroid and parathyroid-related diseases: a Mendelian randomization study  Abstract: Traditional observational studies have supported an association between vitamin D and thyroid and parathyroid-related diseases. However, this association may be influenced by several confounding factors that can lead to significant differences in the results between studies. Mendelian randomization is strongly associated with the exposure and strengthens the causal inference by controlling for nonheritable environmental confounders and reverse causation. In this study, two-sample bidirectional mendelian randomization analysis was performed to investigate the causal relationship between serum vitamin D levels and thyroid and parathyroid-related diseases. Three main MR estimation methods were utilized, and the IVW method was the primarily used to obtain the results. Heterogeneity was evaluated using the Cochran Q test. The possibility of pleiotropy was assessed using the MR-PRESSO and MR-Egger intercepts. The Bonferroni corrected significance threshold was 0.0038. At the Bonferroni-corrected significance level, we found that serum vitamin D levels suggestively decreased the risk of benign parathyroid adenoma (OR=0.244, 95% CI=0.074-0.802, P=0.0202) in the MR analyses. According to the reverse Mendelian randomization study, a genetically predicted risk of thyroid cancer suggestively increased the risk of serum vitamin D levels (OR=1.007, 95% CI=1.010-1.013, P=0.0284), chronic thyroiditis significantly increased the risk of serum vitamin D levels (OR=1.007, 95% CI=1.002-1.011, P=0.0030), and thyroid nodules was significantly decreased the risk of serum vitamin D levels (OR=0.991, 95% CI=0.985-0.997, P=0.0034). The findings might be less susceptible to horizontal pleiotropy and heterogeneity. Chronic thyroiditis and thyroid nodules may play a causal role in the development of vitamin D, but the underlying mechanisms require further investigation. |
|  | **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 | 3-6 | The majority of published studies have focused on the impact of vitamin D on autoimmune thyroid disease and thyroid cancer. Few studies have explored the effects of vitamin D on other thyroid and parathyroid-related diseases.  Data on vitamin D and thyroid- and parathyroid-related diseases derived from observational studies will inevitably be affected by factors such as sample size, ethnicity and other confounding variables, making causal inference challenging. Mendelian randomization (MR) is a highly effective methodology for investigating the causal relationship between exposure and disease since disease status typically does not alter the germline DNA sequences. [17] This is accomplished by utilizing genetic variation as an instrumental variable.[18] Genetic variants are randomly allocated during meiosis, much like a random assignment in a randomized controlled trial, additionally the genetic variants undergoes minimal changes throughout an individual's lifetime,[19] which minimizes unmeasured confounding factors and biases caused by reverse causation.[20] Previous MR studies have demonstrated that there is no causal relationship between serum vitamin D levels and the development of Graves’ disease or thyroid cancer. [21,22] However, the causal relationships between vitamin D and other thyroid- and parathyroid -related diseases are unclear. Furthermore, there is currently a lack of MR analyses examining whether the thyroid- and parathyroid-related diseases influence vitamin D levels. |
| 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 | Observational studies have reported correlations between vitamin D levels and thyroid-and parathyroid-related disease. However, traditional observational studies are susceptible to biases such as reverse causality and residual confounders, which hinder our comprehension of vitamin D levels' effect on the likelihood of thyroid- and parathyroid-related diseases. Therefore, we utilized genetic data to investigate the causal associations between serum vitamin D levels and thyroid- and parathyroid-related diseases. |
|  | **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: | Table 1 | Methods |
|  | 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. | 6-7 | Methods: Exposure and outcome data sources section |
|  | 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 | 6-7 | Methods: Exposure and outcome data sources section |
|  | c) | Describe measurement, quality control and selection of genetic variants | 6-7 | Methods: Exposure and outcome data sources section |
|  | d) | For each exposure, outcome, and other relevant variables, describe methods of assessment and diagnostic criteria for diseases | NA | Used GWAS summary statistics |
|  | e) | Provide details of ethics committee approval and participant informed consent, if relevant | 2 | The data used in this study were obtained from publicly available GWAS. Ethical approval is not required for this study. |
| 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 | 8-9 | Method: Study design & Mendelian randomization analysis section |
| 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) |  | The effect size of this study was presented as odds ratios (ORs), indicating an increased risk of outcome for each increase in exposure log ratio. |
|  | b) | Describe how genetic variants were handled in the analyses and, if applicable, how their weights were selected | 8-9 | Instrumental variables were determined through various processes. First, to meet the association assumption, a threshold of P<5\*10-8 was established, and the threshold was relaxed to 5\*10-6 or 1\*10-5 when adequate single nucleotide polymorphisms (SNPs, which primarily DNA sequence polymorphisms that result from variation in a single nucleotide at the genomic level [30]) were not available for analysis at the P<5\*10-8 threshold. Second, to minimize the occurrence of multiple results due to linkage disequilibrium (LD), LD analyses were conducted (r2< 0.001, kb = 10,000) in accordance with the required independence assumptions. The LD levels were then obtained from the European samples of the 1000 Genomes Project.[31] To ensure that the impact of selected SNPs on both exposure and outcome aligned with the corresponding alleles(an allele is one of two or more versions of the DNA sequence at a given genomic location[32]), we eliminated the palindromic structure and utilized surrogate SNPs when relevant SNPs were absent from the GWAS dataset for outcome. Confounders not associated with SNPs (including other types of thyroid disease, other vitamin levels, smoking, alcohol consumption, obesity, etc.) were manually removed by PhenoScanner.[33] The direction of causality between exposure and outcome was evaluated using Steiger filtering. If the instrumental variable met the criteria, the instrument's direction was "TRUE", and SNPs with a direction of "FALSE" were excluded.[34] To assess the strength of the instrumental variables, we calculated the F-statistic (F reflects the bias of an IV or a set of IVs[35]), using the formula[36], with R2 representing the proportion of variance explained by the instrumental variable, n representing sample size, and k representing the number of SNPs. [37] An F-statistic greater than 10 indicates that the instrumental variable may be resistant to the effects of weak instrumental bias. This analysis adheres to the conventional academic structure and employs clear, objective language with precise technical terms.[38] |
|  | 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 | 8-9 | MR methods applied include Inverse variance, IVW, Weighted median, MR Egger. IVW assumes that all instrumental variables (IVs) meet the validity criteria to obtain unbiased estimates,[28] specifically, it assumes that all single nucleotide polymorphisms (SNPs) are not related to the pleiotropic effect of serum vitamin D levels (known as the InSIDE hypothesis). [29] which has the highest test efficacy. While the Weighted Median method assumes that only more than half of the IVs are unbiased,[30] the IVW method is the primary reference for obtaining results. Weighted Median and MR Egger methods are used to complement the IVW analysis. Causal effects were assessed as odds ratios (OR), indicating an increased risk of outcome for each increase in exposure log ratio. Our sensitivity analyses used four main methods, including Cochran's Q test, MR-Egger intercept analysis, MR-PRESSO, and leave-one-out sensitivity test. Horizontal pleiotropy was evaluated through the MR-Egger test's intercept, with a significance threshold of P<0.05 denoting horizontal pleiotropy's existence and the intercept term's value distance from 0 indicating horizontal pleiotropy's magnitude. [31] Heterogeneity was evaluated using the Cochran Q statistic, whereby a P-value of less than 0.05 indicates the existence of heterogeneity.[32] When heterogeneity was detected, we utilized MRPRESSO to identify potential outliers, then removed them before re-evaluating causality with the remaining SNPs.[33] The "Leave-one-out" technique was utilized to evaluate the impact of a single SNP on the analysis, which enhances the robustness of the outcomes.[16] We performed RadialMR analysis using modified second-order weights to identify outliers. [34] Radial plots offer advantages in identifying peripheral studies, detecting small study biases, and pinpointing outliers more directly than traditional scatter plots. Furthermore, the outcomes of Mendelian randomization analyses were presented via forest plots, scatter plots, funnel plots, and "Leave-one-out" plots. |
|  | d) | Explain how missing data were addressed | NA | Used GWAS summary statistics |
|  | e) | If applicable, indicate how multiple testing was addressed | NA |  |
| 7 | **Assessment of assumptions** | Describe any methods or prior knowledge used to assess the assumptions or justify their validity |  | Methods: Mendelian randomization analysis section |
| 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) | 8-9 | The MR methods applied included the inverse variance weighted (IVW), weighted median and MR Egger methods. The IVW method assumed that all IVs met the validity criteria to obtain unbiased estimates,[39] specifically, it assumed that all SNPs were not related to the pleiotropic effect of exposure (known as the InSIDE hypothesis). [40] which had the highest test efficacy. When the weighted median method assumed that only more than half of the IVs were unbiased,[41] the IVW method was the primary reference for obtaining results. The weighted median and MR Egger methods were used to complement the IVW analysis. Causal effects were assessed as odds ratios (ORs), indicating an increased risk of outcome for each increase in the exposure log ratio. Our sensitivity analyses used four main methods, including Cochran's Q test, MR-Egger intercept analysis, MR-PRESSO, and the leave-one-out sensitivity test. Horizontal pleiotropy was evaluated through the MR-Egger test's intercept, with a significance threshold of P<0.05 denoting the existence of horizontal pleiotropy and the intercept term's value distance from 0 indicating the magnitude of horizontal pleiotropy. [42] Heterogeneity was evaluated using the Cochran Q statistic, whereby a P\_value of less than 0.05 indicated the existence of heterogeneity.[43] When heterogeneity was detected, we utilized MRPRESSO to identify potential outliers and then removed them before re-evaluating causality with the remaining SNPs.[44] The "Leave-one-out" test was utilized to evaluate the impact of a single SNP on the analysis, and this test enhanced the robustness of the outcomes.[19] We performed RadialMR analysis using modified second-order weights to identify outliers.[45] Radial plots offered advantages in identifying peripheral studies, detecting small study biases, and identifying outliers more directly than traditional scatter plots.  The outcomes of the Mendelian randomization analyses are presented as scatter plots, forest plots, "leave-one-out" plots, and funnel plots. Each point in the scatter plots represents an SNP, thereby demonstrating the association of that SNP with exposure and outcome. The forest plot comprised horizontal lines and points, with each line representing the effect size of a SNP and its 95% confidence interval. Given the lack of robustness of the results for individual SNPs, it was necessary to combine them, and this is represented by the bottom red line (All − Inverse variance weighted). The leave-one-out forest plot was used to calculate the meta effect of the remaining SNPs after removing each SNP one by one. If all the error lines were consistent to the right or left of 0, the results were deemed reliable. The funnel plot was generated to determine whether the points situated on either side of the IVW line were approximately symmetrical. The presence of any outlying points indicated the potential for outliers, which could be removed, and the analysis process was repeated. |
| 9 | **Software and pre-registration** |  |  |  |
|  | a) | Name statistical software and package(s), including version and settings used | 8 | Causal estimation was mainly conducted using the "TwoSampleMR" (version 0.5.7), "MRPRESSO" (version 1.0), "ggplot2" (version 3.4.3), "plyr" (version 1.8.8), and "phenoscanner" (version 1.0) packages. All MR analyses were conducted using R version 4.3.1 (The R Foundation for Statistical Computing). |
|  | b) | State whether the study protocol and details were pre-registered (as well as when and where) | NA |  |
|  | **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 | NA | Used publicly available GWAS summary statistics |
|  | b) | Report summary statistics for phenotypic exposure(s), outcome(s), and other relevant variables (e.g. means, SDs, proportions) | Table 1 | Methods: Exposure and outcome data sources section |
|  | c) | If the data sources include meta-analyses of previous studies, provide the assessments of heterogeneity across these studies | Table 2 |  |
|  | d) | For two-sample MR:  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 | Table 1 | The selected GWASs included participants of predominantly European ancestry. We used the summary statistics of GWAS conducted in exclusively European ancestry participants.  SNPs that met the genome-wide significance level (P <5\*10-8) were initially selected as genetic instruments. If no SNPs met the genome-wide significance level, a lower threshold of P <5\*10-6或1\*10-5 was used.  Provided data sources (ethnicity) in Table 1 |
| 11 | **Main results** |  |  |  |
|  | a) | Report the associations between genetic variant and exposure, and between genetic variant and outcome, preferably on an interpretable scale | Supplemental file TableS1-S4 | The F statistic are in Supplemental file Table S1-4 |
|  | 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 | Figure 2、Figure 4 |  |
|  | c) | If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | NA |  |
|  | 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 3&5-7 |  |
| 12 | **Assessment of assumptions** |  |  |  |
|  | a) | Report the assessment of the validity of the assumptions | Table 2,  Supplemental file Table S1-4 | This study used publicly available GWAS summary statistics which limits the assessment for testing the assumptions .The F statistic are in Supplemental file Table S1-4. |
|  | b) | Report any additional statistics (e.g., assessments of heterogeneity across genetic variants, such as *I2*, Q statistic or E-value) | Table 2 |  |
| 13 | **Sensitivity analyses and additional analyses** |  |  |  |
|  | a) | Report any sensitivity analyses to assess the robustness of the main results to violations of the assumptions | Table 2 |  |
|  | b) | Report results from other sensitivity analyses or additional analyses |  |  |
|  | c) | Report any assessment of direction of causal relationship (e.g., bidirectional MR) | Figure 2 & Figure 4 |  |
|  | d) | When relevant, report and compare with estimates from non-MR analyses | NA |  |
|  | e) | Consider additional plots to visualize results (e.g., leave-one-out analyses) | NA |  |
|  | **DISCUSSION** |  |  |  |
| 14 | **Key results** | Summarize key results with reference to study objectives |  | At the Bonferroni-corrected significance level, we found that serum vitamin D levels suggestively decreased the risk of benign parathyroid adenoma (OR=0.244, 95% CI=0.074-0.802, P=0.0202) in the MR analyses. According to the reverse Mendelian randomization study, a genetically predicted risk of thyroid cancer suggestively increased the risk of serum vitamin D levels (OR=1.007, 95% CI=1.010-1.013, P=0.0284), chronic thyroiditis significantly increased the risk of serum vitamin D levels (OR=1.007, 95% CI=1.002-1.011, P=0.0030), and thyroid nodules was significantly decreased the risk of serum vitamin D levels (OR=0.991, 95% CI=0.985-0.997, P=0.0034). The findings may be less susceptible to horizontal pleiotropy (with P-values for MR-Egger intercept and MRPRESSO P-values greater than 0.05). No significant causal relationship was discovered between vitamin D and either thyroid or parathyroid-related diseases. |
| 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 | 34 | First, the data from our study were obtained from summary GWAS data, and specific information necessary for further analysing age, gender, and time of blood collection among the study population are lacking. Second, the data in the FinnGen database were derived from the primary diagnosis of the study population. However, patients might have additional comorbidities, which could lead to biased results. Third, because the number of SNPs included in some diseases was insufficient for MR analysis, the thresholds for selecting SNPs were appropriately lowered. Additionally, some of the available SNPs for the exposure-disease associations were low, which may have affected the study conclusions. Fourth, while we have employed various approaches to minimize pleiotropy, potential unidentified pathways and confounders between exposure and outcome might still lead to inaccuracies in our findings. Fifth, the study subjects were primarily of European origin (the study's participants were primarily from FinnGen and the UK database), which limits the generalizability of the study findings to the broader European population, and there may be genetic variations between different races. Whether the findings from this research can be extended to other racial groups remains uncertain, and additional corroboration of the outcomes is necessary in the future. |
| 16 | **Interpretation** |  |  |  |
|  | a) | Meaning: Give a cautious overall interpretation of results in the context of their limitations and in comparison with other studies | 14-17 | Discussion section |
|  | 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 | 15-16 | Discussion 2-3 |
|  | c) | Clinical relevance: Discuss whether the results have clinical or public policy relevance, and to what extent they inform effect sizes of possible interventions | 15-17 | Discussion 2-3 |
| 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 | 17 | Discussion 5 |
|  | **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 | 21 | This study was supported by the Natural Science Foundation of Fujian,China(No.2022J011004), the Natural Science Foundation of Fujian,China(No.2023J011207), the Natural Science Foundation of Fujian,China(No.2021J01397), the Fujian provincial health technology project(No.2022GGA010) |
| 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 | NA | Used publicly available GWAS summary statistics |
| 20 | **Conflicts of Interest** | All authors should declare all potential conflicts of interest | 1 | Disclosures: None |

This checklist is copyrighted by the Equator Network under the Creative Commons Attribution 3.0 Unported (CC BY 3.0) license.

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.