89254_Auto_Edited.docx

WORD COUNT	3966	TIME SUBMITTED	29-JAN-2024 06:13PM
		PAPER ID	106500595

Name of Journal: *World Journal of Clinical Cases*

Manuscript NO: 89254

Manuscript Type: ORIGINAL ARTICLE

Clinical and Translational Research

Investigating the Causal Associations between Five Anthropometric Indicators and Nonalcoholic Fatty Liver Disease: Mendelian Randomization Study

Causal associations between anthropometric indicators and NAFLD

Xian-Pei Xiao, Yu Zhang, Yong-Jun Dai, Meng Yang, Jian Xie, Guo Chen, Zheng-Jun Yang

Abstract

Purpose: To evaluate the causal relationships between five anthropometric indicators and nonalcoholic fatty liver disease (NAFLD) employing Mendelian randomization (MR) design.

Methods: The Anthropometric Consortium provided genetic exposure data for five anthropometric indicators, including hip circumference (HC), waist circumference (WC), waist-to-hip ratio (WHR), body mass index (BMI), and body fat percentage (BF). Genetic outcome data for NAFLD were obtained from the UK Biobank and FinnGen Consortium. Genome-wide significant single nucleotide polymorphisms (SNPs) were chosen as instrumental variables (IVs). Univariable MR (UVMR) and multivariable MR (MVMR) designs with analytical approaches, including inverse variance weighted (IVW), MR-Egger, weighted median (WM), and weighted mode methods, were used to assess the causal relationships between anthropometric indicators and NAFLD.

Results: Causal relationships were revealed by UVMR, indicating that a higher risk of NAFLD was

associated with a per-unit increase in WC (IVW: OR = 2.67, 95%CI: 1.42, 5.02, $P = 2.25\times10^{-3}$), and BF was causally associated with an increased risk of NAFLD (WM: OR = 2.23, 95%CI: 1.07, 4.66, $P = 0.033$). The presence of causal effects of WC on the decreased risk of NAFLD was supported by MVMR after adjusting for BMI and smoking. However, no causal association between BF and NAFLD was observed. In addition, other causal relationships of HC, WHR (BMI adjusted), and BMI with the risk of NAFLD were not retained after FDR correction.

Conclusion: This study establishes a causal relationship, indicating that an increase in WC is associated with a higher risk of NAFLD. This demonstrates that a suitable decrease in WC is advantageous for preventing NAFLD.

Keywords: Anthropometric indicator; Waist circumference; Nonalcoholic fatty liver disease; Mendelian randomization

Abbreviations: MR, Mendelian randomization; IVW, inverse variance weighting method; MR-PRESSO, MR pleiotropy residual sum and outlier; TSMR, Two-sample Mendelian randomization; MVMR, Multivariable Mendelian randomization; SD, standard deviation; SNPs, single nucleotide polymorphisms; WM, weighted median; NAFLD, Nonalcoholic fatty liver disease; GWAS, genome-wide association study.

Core Tip: Previous studies have demonstrated the potential significance of anthropometric indicators in the development of nonalcoholic fatty liver disease (NAFLD). Nevertheless, inconsistencies exist in the results of these studies, and the causal association remains unclear. Abdominal obesity (AO), measured by waist circumference (WC), is a risk factor for NAFLD, as demonstrated by previous studies. Nevertheless, many of these studies were cross-sectional or considered only a single measurement, neglecting a comprehensive evaluation of changes in WC over time and the effect of long-term development and lifestyle changes.

Consequently, establishing a causal relationship between anthropometric indicators and NAFLD requires further robust evidence.

1. Introduction

Excessive fat deposition in liver cells characterizes nonalcoholic fatty liver disease (NAFLD). Currently, due to lifestyle changes and insufficient daily physical activity, NAFLD has emerged as the most severe chronic disease in society[1]. Reports indicate that NAFLD affects a quarter of the global population, with a prevalence rate of 24% and a tendency toward younger age groups[2,3]. Although the etiology of NAFLD has not been thoroughly understood, the emerging roles of anthropometric indicators in assessing and predicting the risk of NAFLD have been highlighted by accumulating evidence[4]. Numerous previous observational studies have reported relationships between the risk of NAFLD and noninvasive quantitative measurements of the body, such as anthropometric indicators, which comprise height, weight, hip circumference (HC), waist circumference (WC), waist-to-hip ratio (WHR), body mass index (BMI), and body fat percentage (BF)[4,5]. Contradictory findings have been obtained in some other studies, indicating no relationship between anthropometric indicators and the risk of NAFLD[6,7]. The causal relationships between anthropometric indicators and NAFLD risk remain undetermined, considering these inconsistent findings and the absence of randomized controlled studies.

Mendelian randomization (MR) is a novel epidemiological tool that employs genetic data to investigate the causal relationship between exposure and outcome[8]. Generally, genetic variants are independent of disease state and are randomly assigned to offspring through the maternal generation. Consequently, the limitations of conventional observational design can be overcome, and biases such as potential confounders and reverse causality can be minimized[9,10]. Previous studies have demonstrated the use of MR to investigate causal relationships between NAFLD and numerous diseases, including cardiovascular disease and psoriasis[11,12].

In this study, the causality of five anthropometric indicators, including HC, WC, WHR (BMI adjusted), BMI, and BF, with the risk of NAFLD, was investigated using genetic summary statistics with UVMR and MVMR frameworks to establish a foundation for the prevention and management of this high-burden disease.

Methods

Study Design

In this MR analysis, the causal effects of various modifiable exposures on outcomes were estimated using single nucleotide polymorphisms (SNPs) as instrumental variables (IVs). However, three basic assumptions must be satisfied[13]. First, IVs must exhibit a high correlation with the exposure factor (correlation assumption) to minimize weak instrumental variable bias[14]. Second, the outcome should only be influenced by the identified IVs through exposure and not through other factors, expressed as "no horizontal pleiotropy" (exclusionary assumption)[15]. Third, to identify genuine causal associations, IVs should be independent of confounders in exposure-outcome associations (independence assumption)[16].

A two-stage design was used in this study, with the first phase employing bidirectional UVMR analyses to analyze the causal relationship between the five anthropometric measures and NAFLD. In the second phase, MVMR analyses were conducted to establish causality after adjusting for potential confounders, such as BMI and smoking. These findings were validated using databases from various sources. **Figure 1** shows the study design.

Exposure Data Sources and IVs Selection

Five anthropometric measures, including HC, WC, WHR (BMI adjusted), BMI, and BF, were chosen as exposure points. Shungin *et al.* reported a genome-wide association study (GWAS) that included up to

224,459 individuals, from which summary statistics for HC, WC, and WHR (BMI adjusted) were extracted[17]. The summary data for BMI were obtained from another GWAS, which involved 700,000 participants[18]. Genetic summary data for BF were obtained from GWAS data published by Neale Lab in 2017, which included 331,117 participants (http://www.nealelab.is/uk-biobank). All exposure datasets were conducted in European ancestry.

The threshold was set at $P < 5 \times 10^{-08}$ to choose SNPs related to anthropometric metrics as IVs. Furthermore, SNPs with strong linkage disequilibrium (LD) ($R^2 < 0.001$, window size = 10,000kb) were deleted to avoid possible bias. Significantly heterogeneous SNPs were excluded, and the remaining anthropometric-associated SNPs were selected as valid IVs using the heterogeneity test. In some of the exposure-outcome analyses, some SNPs were excluded due to the lack of available proxies, as they were not present in the outcome **(Supplementary Table 1)**. Finally, a total of 807 anthropometric-associated SNPs were available, including 18 HC-associated SNPs, 46 WC-associated SNPs, 32 WHR (BMI adjusted)-associated SNPs, 485 BMI-associated SNPs, and 226 BF-associated SNPs (**Supplementary Table 1**).

Outcome Data Sources

Concerning the outcome datasets, genetic associations for NAFLD agent-related SNPs were extracted from the UK biobank, which included all European subjects, comprising 8,434 cases and 770,180 controls[19]. The GWAS findings were harmonized with those of GWAS exposed to the same effect allele. Effects allele frequencies (minor allele frequency ≤ 0.5) were employed to harmonize the palindromic SNPs. For IVs that could not be found in the results, attempts were made to include suitable proxy SNPs ($r^2 \geq 0.8$) by searching for them on the website (http://www.mulinlab.org/vportal/index.html).

Ethical Approval

This study was conducted using public open data from the MRC-IEU database and the FinnGen research project; thus, no ethical review was necessary.

Statistical Analyses

UVMR Analysis

Inverse variance weighted (IVW), weighted median (WM), weighted model, and MR Egger methods were used to evaluate the causal effects of HC, WC, WHR (BMI adjusted), and BF (per standard deviation [SD]) on the risk of NAFLD. The IVW model, which was defined as the primary approach, can provide causal effect estimates with optimal precision when all IVs are valid and minimize the impact of heterogeneity. To assess weak IV bias, the variance (R^2) and approximated *F*-statistic for per-exposure explained by IVs were calculated. This evaluation, which is based on the inherent flaws in IV selection, is generally considered to be sufficiently instrumental for *F*-statistics greater than 10[20,21]. The formula used is $F = R^2 \times (N - 2) / (1 - R^2)$, where R^2 represents the variance of exposure explained by each IV[22]. In addition, FDR correction was implemented to control false positives that tend to occur due to multiple testing.

MVMR Analyses

Smoking and BMI were the primary confounders associated with exposure (anthropometric indicators) and outcome (NAFLD). Therefore, to determine the independent effects of anthropometric indicators on the risk of NAFLD development, MVMR analyses were employed. In MVMR analyses, MR-PRESSO models were employed to correct for horizontal pleiotropy in causal effects, estimate heterogeneity, and exclude potential outliers. A significance level of $P < 0.05$ in the MVMR analysis was considered indicative of

statistical significance.

Sensitivity Analysis

To ensure the stability of the study's results, several sensitivity analyses were conducted. First, the nonexistence of horizontal pleiotropy, a fundamental premise for satisfying causal inference, was analyzed using the MR-PRESSO approach. Findings with a p-value greater than 0.05 indicated the nonexistence of horizontal pleiotropy[23]. Second, Cochran's Q statistic was employed to eliminate heterogeneity of IVs, with p-values over 0.05 indicating the absence of heterogeneity in the IVs[24]. Furthermore, leave-one-out (LOO) analyses were performed to exclude single SNPs exerting a substantial influence on the findings[25]. An online tool "mRnd" was used to calculate the statistical efficacy (power). R programming software (version 4.0.3) with the "MR," "TwoSampleMR," "MVMR," and "MRPRESSO" packages was used to perform all statistical analyses and data visualization.

Replication in Another European-based Population

Anthropometric Indicators Replication

Published anthropometric indicators GWAS data were extracted from other consortiums ($P < 5 \times 10^{-08}$, $R^2 < 0.001$, window size = 10,000kb). The GWAS data for HC were obtained from Neale Lab's 2017 summary database, which included 336,639 individuals in this study (http://www.nealelab.is/uk-biobank). Summary data for WC and BF were derived from the MRC-IEU database and exported from the GWAS pipeline using the UK Biobank's pheasant-derived variables. Furthermore, the Within Family GWAS consortium was the source of data for WHR (BMI adjusted). In this genetic epidemiological study, sample analyses of related individuals (such as siblings or parent-child triples) were performed (https://www.withinfamilyconsortium.com). The BMI data were obtained from the GWAS study conducted

by Martina *et al.*, which included 236,781 individuals[26]. All data were extracted from published studies, and consequently, no additional ethical review was necessary. **Figure 1** shows the details of the exposure and outcome data.

NAFLD Replication

Outcome data were obtained from FinnGen, a large cohort study comprising genomic and health data from 500,000 Finland BioBank individuals (https://www.finngen.fi). Similar allele frequency distributions to other European populations were observed in Finland; however, significant strengths, including uncommon variants in intricate phenotypes and unique group genetic history, were noted[27].

Results

Baseline Characteristics

A total of 807 anthropometric indicator-associated SNPs, including 18 HC-associated SNPs, 46 WC-associated SNPs, 32 WHR (BMI adjusted)-associated SNPs, 485 BMI-associated SNPs, and 226 BF-associated SNPs, were identified from three independent GWAS analyses **(Supplementary Table 1)**. Among the 807 anthropometric indicator-associated SNPs, all F-statistics were greater than 10, indicating a low likelihood of weak IVs among the included SNPs. In addition, the selected IVs explained approximately 0.84% (HC), 1.13% (WC), 1.56% (WHR [BMI adjusted]), 4.85% (BMI), and 3.60% (BF) of the phenotype variances. Low degrees of sample overlap between exposure and outcome were indicated by the sample overlap results, with overlap rates within datasets being less than 15%.

Stage 1 Causal Associations between Anthropometric Indicators and NAFLD Risk

All findings were based on the IVW random-effects model because evidence of horizontal pleiotropy was

not found ($P > 0.05$), but heterogeneity was present ($P < 0.05$). The findings of the IVW approach revealed causal relationships between five anthropometric indicators and NAFLD: HC ($OR_{IVW} = 1.70$, 95%CI: 1.27, 2.27, $P = 3.29 \times 10^{-04}$; $OR_{WM} = 1.76$, 95%CI: 1.25, 2.48, $P = 1.20 \times 10^{-03}$), WC ($OR_{IVW} = 1.80$, 95%CI: 1.42, 2.27, $P = 1.24 \times 10^{-06}$; $OR_{WM} = 1.88$, 95%CI: 1.40, 2.51, $P = 2.42 \times 10^{-05}$), WHR (BMI adjusted) ($OR_{IVW} = 1.42$, 95%CI: 1.20, 1.69, $P = 6.47 \times 10^{-05}$; $OR_{WM} = 1.35$, 95%CI: 1.06, 1.71, $P = 1.40 \times 10^{-03}$), BMI ($OR_{IVW} = 1.60$, 95%CI: 1.43, 1.79, $P = 1.49 \times 10^{-16}$; $OR_{WM} = 1.64$, 95%CI: 1.38, 1.94, $P = 1.16 \times 10^{-08}$), and BF ($OR_{IVW} = 1.67$, 95%CI: 1.37, 2.04, $P = 2.50 \times 10^{-07}$; $OR_{WM} = 1.91$, 95%CI: 1.50, 2.43, $P = 1.36 \times 10^{-07}$) (**Figure 2 & Supplementary Table 2**). Furthermore, sensitivity analyses using the LOO approach indicated that no single SNPs drove these findings after the stepwise elimination of individual SNPs (all error lines were on either side of zero) (**Supplementary Figure 1-5**). Considering the positive false due to multiple testing, significant causal relationships between HC, WC, WHR (BMI adjusted), BMI, BF, and NAFLD were still observed in the results after FDR correction ($P < 0.01$) (**Supplementary Figure 6**).

Causal Associations between NAFLD and Anthropometric Indicators

No significant genetic predictive correlation was observed when NAFLD was tested as an exposure factor for inverse correlation, as indicated by estimates derived from the IVW method ($P > 0.05$). This further supports the notion that reversed causal relationships with NAFLD do not confound anthropometric indicators (**Supplementary Table 2**). In addition, the result of no significant genetic prediction of associations between NAFLD and anthropometric indicators was supported by sensitivity analysis.

Replication of Results within Different GWAS of Exposures and Outcomes

With the above IVs selection standard, 1,052 SNPs associated with anthropometric indicators, including

390 HC-associated SNPs, 209 WC-associated SNPs, 15 WHR- associated SNPs, 79 BMI-associated SNPs, and 359 BF-associated SNPs, were selected for causal estimation in the replicated exposures **(Supplementary Table 1)**. For HC, WC, WHR (BMI adjusted), BMI, and BF, the *F*-statistics were all over 10, indicating that the chosen IVs were sufficiently robust and not susceptible to the influence of weak IVs. Furthermore, the statistical power of these replication exposures was similarly calculated, and the selected IVs were shown to explain approximately 5.50% (HC), 3.38% (WC), 0.99% (WHR [BMI adjusted]), 2.17% (BMI), and 4.60% (BF) of the phenotype variances.

Subsequently, the two sources of exposure (discovery and replication) were analyzed in a bidirectional MR analysis with outcomes from various GWAS (UK biobank and FinnGen). The causal relationship between an increased risk of NAFLD and higher WC, WHR (BMI adjusted), BMI, and BF was supported by our cross-validation comparison of four sets of MR findings; sensitivity analysis findings remained consistent. Although the existence of heterogeneity between SNPs was indicated by the evidence, the overall level of pleiotropy based on the MR-Egger intercept was not significant. In our inverse MR analysis of anthropometric indicators (discovery database) with NAFLD (FinnGen research project), the presence of NAFLD resulted in a decrease in BMI, indicating that the previously observed causal relationship between BMI and NAFLD was influenced by reverse causality bias. In addition, FDR correction was equally performed on the validated results, resulting in the loss of some weak causal relationships. The findings of the four MR analyses supported the fact that an increase in WC (OR_{IVW} = 1.80, 95%CI: 1.42, 2.27, $P = 1.24 \times 10^{-06}$; OR_{WM} = 1.88, 95%CI: 1.40, 2.51, $P = 2.42 \times 10^{-05}$) and BF (OR_{IVW} = 1.67, 95%CI: 1.43, 1.79, $P = 2.50 \times 10^{-07}$; OR_{WM} = 1.91, 95%CI: 1.50, 2.43, $P = 1.36 \times 10^{-07}$) would result in a higher risk of NAFLD after the exclusion of reverse causality (**Figure 3-5**). **Supplementary Figure 6** shows the IVW

results of the FDR correction.

MVMR Analysis of Potential Effects between WC and NAFLD, BF and NAFLD

After adjusting for BMI and smoking, the findings of the MVMR demonstrated that the relationship between the genetically determined increase in WC and an increase in the risk of NAFLD remained significant (OR = 1.58, 95% CI: 1.06, 2.35, P = 0.025), indicating that WC may be a factor for increased risk of NAFLD. Nevertheless, a causal relationship between BF and the risk of NAFLD was not supported by the MVMR findings (OR_{IVW} = 1.02, 95% CI: 0.60, 1.73, P = 0.941) (**Figure 6)**. Furthermore, causal relationships between HC, WHR (BMI adjusted), BMI, and the risk of NAFLD were not supported by the cross-validation findings.

Discussion

The primary causes of the increased global burden of NAFLD are changes in lifestyle and dietary habits, and identifying risk factors is particularly crucial for disease prevention and control[28]. Previous studies have demonstrated that anthropometric indicators may primarily be risk factors for developing NAFLD[29]. Nevertheless, conflicting findings exist from various studies, and the causal relationships still need to be clarified[4,30]. Risky relationships between WC and BF for NAFLD were observed from repeated-validation TSMR. After excluding the confounding effects of BMI and smoking, the direct causal effect of BF on NAFLD could not be pursued in subsequent MVMR studies.

As a critical indicator of abdominal obesity (AO) patterns, a strong relationship exists between WC and the early risk of all-cause mortality. An increase of 10 cm in WC for an individual results in an increase of 11% points in the risk of all-cause mortality[31]. Furthermore, positive relationships with the risk of morbidity and mortality of major chronic diseases were exhibited by WC[32]. The importance of WC as an indicator of

population health and long-term adverse outcomes was emphasized by these results. In addition, a relationship between WC and NAFLD was observed. A study in an adolescent population indicated that AO was a predictor of pulmonary fibrosis in children and adolescents with NAFLD, with WC as the primary measure[33]. Furthermore, a Korean study discovered a direct relationship between a larger WC and an increased risk of developing NAFLD[33]. This result was supported by a study of the Iranian population[34]. Although numerous studies have demonstrated that AO is a risk factor for NAFLD, with WC being the primary measure, most of them have been cross-sectional studies or single-measurement findings and the long-term dynamic changes in an individual's WC in response to changes in growth, development, and lifestyle, among other things, have not been sufficiently evaluated. From a genetic viewpoint, our results indicated a causal relationship between greater WC and higher NAFLD. The biological relationship can be explained by several possible mechanisms. On the one hand, AO reflects the excessive accumulation of visceral adiposity (VAT), which increases WC; on the other hand, an increase in WC results in a substantial collection of VAT, enabling the continuous secretion of pro-inflammatory cytokines, such as TNF-α. This activates inflammatory signaling pathways, thereby mediating metabolic homeostasis and insulin sensitivity of the organism[35–37].

BF measures the amount of fat in a person's body and has been identified as a crucial risk factor for cardiometabolic[38]. BF is strongly related to the risk of NAFLD, as indicated by the Rotterdam study, and this relationship is more visible in the female population. Furthermore, in a longitudinal relationship investigating the effect of BF changes on NAFLD incidence and remission, Kim *et al*. demonstrated that increased BF was longitudinally associated with increased risk of NAFLD and negatively associated with NAFLD regression[39]. From the two replication datasets, our combined TSMR findings indicated that

increased BF substantially increased the risk of NAFLD. After excluding the effects of BMI and smoking, MVMR rejected a direct causal impact of BF on NAFLD.

Furthermore, the relationships of HC, WHR ((BMI adjusted)), and BMI with NAFLD have been analyzed in numerous studies. A cross-sectional study of the prevalence of NAFLD in pathologically overweight women in South India revealed that BMI, BF, and BWP are three indicators that can be used to a large degree as indicators of the development of NAFLD[40]. Improved cardiovascular metabolism is associated with HC. The development of NAFLD may be triggered by excessive HC, as it has been closely associated with muscle mass, and improving muscle mass will reduce insulin resistance and decrease the likelihood of NAFLD[41,42]. Indicators commonly employed to evaluate nutritional conditions and obesity, such as BMI and WHR (BMI adjusted), primarily reflect the whole-body fat problem rather than VAT. WC is more reflective of the volume of VAT than BMI[43]. Consequently, some studies support visceral obesity with WC and BF as primary measures of a crucial risk factor for NAFLD[44]. Based on European GWAS pooled data, a previous study reported partially different findings compared with our findings, indicating that WHR may be a potential risk factor for NAFLD[45]. Several factors could explain this difference. First, the WHR (BMI adjusted) data used in the GWAS were corrected for BMI, indicating that changes in WHR independent of BMI might not be fully captured by the selection of WHR (BMI adjusted) SNPs. Second, the causal inferences included fewer SNPs, resulting in limited phenotypic differences and reduced statistical power to detect a true causal association between WHR and NAFLD risk. Furthermore, differences in study populations and sample sizes could have contributed to the inconsistent findings. To address these challenges, cross-repeated validation was performed using two independent GWAS datasets to eliminate bias resulting from data selection and enhance the credibility of our findings.

Several limitations require attention. First, the exposure and outcome datasets were obtained from European ancestry; therefore, these findings may not be generalizable to other populations with different genetic backgrounds. Further studies are required to validate these results in other ethnic populations. Second, although the F-statistic can be used to evaluate the first hypothesis, verifying the second and third hypotheses is generally challenging and may lead to potential bias. Third, the causal relationship between anthropometric indicators and NAFLD in different gender/age groups cannot be investigated due to the lack of personal demographic information on anthropometric indicators.

Although previous studies have extensively investigated the relationship between AO and NAFLD, there are limited studies on the association between other anthropometric measures[45,46]. This study expands and validates these results by incorporating a wider range of anthropometric measures, offering additional supporting evidence for a causal association in early-stage NAFLD. Compared with previous studies, this study has numerous notable advantages. Although the association between WC, BMI, and BMI-adjusted WHR and NAFLD has been partially examined in previous studies, evidence for a causal association between other anthropometric indicators (such as HC and BF) and NAFLD remains limited. To the best of our knowledge, the causal association between the five major anthropometric indicators and NAFLD was comprehensively assessed for the first time in this study using UVMR and MVMR approaches. This study provides valuable causal evidence and directionality for the early prediction and diagnosis of NAFLD. The results of this study were replicated using GWAS data from a European population to enhance the reliability of the findings and ensure robust conclusions. Data from two distinct sources were examined in this replication, thereby increasing the potential for identifying new opportunities. In our preliminary analysis, several unidirectional causal associations, such as HC, WC, WHR (BMI adjusted), BMI, and BF,

were identified. However, nonreproducible causal relationships between four indicators (HC, WHR, BMI, BF) and NAFLD were excluded from our conclusion to enhance the reproducibility of our approach and reinforce the strength of our conclusions.

Conclusions

This study demonstrates that genetically determined increased WC maintains a positive and causal association with NAFLD, even in the presence of confounders, including BMI and smoking. This underscores the potential of WC as a reliable indicator for the early identification and diagnosis of NAFLD.

Article Highlights

This study offers the first comprehensive assessment of causal associations between five anthropometric measures and Nonalcoholic fatty liver disease (NAFLD) by using both Univariate Mendelian randomization (UVMR) and Multivariable MR (MVMR) methods. Considering the possibility of potential chance in the results, we additionally selected another exposure and outcome Genome-wide association study (GWAS) data from European population for replication, including a cross-analysis of two different sources of data.