

Research Article

Mendelian Randomization Explores the Causal Relationship between Household Income and Breastfeeding

Han Kong^{1#}, Huifang He^{2#}, Chaoqun Wang^{3#}, Jinchai Long², Shuaining Tang², Qingjie Hu², Siqi Yin, Hai Li², Naizhen Wang², Qiang Gao², Meizi Song², Yu Liu^{2,4*} and Yonggang Zhang^{5*}

¹Department of Economic Management, Shandong Water Conservancy Vocational College, China

²Department of Breast Surgery, The First Affiliated Hospital of Hainan Medical University, China

³Department of Nuclear Medicine, Hainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, China

⁴Department of general surgery, First Clinical Medical College, Nanjing University of Chinese Medicine, China

⁵Department of General Surgery People's Hospital of Rizhao, China

[#]These authors have contributed equally to this work

***Corresponding author**

Yonggang Zhang, Department of General Surgery
People's Hospital of Rizhao, China

Submitted: 12 May, 2025

Accepted: 20 May, 2025

Published: 23 May, 2025

Copyright

© 2025 Kong H, et al.

OPEN ACCESS

Keywords

- Household Income
- Breastfeeding;
- Mendelian Randomization
- Causal Relationship

Abstract

Objective: To explore the causal relationship between household income and breastfeeding by applying Mendelian Randomization (MR) analysis.

Methods: We obtained data on household income and breastfeeding practices from the IEU Open GWAS database derived from European populations. The causal relationship between household income and breastfeeding was analyzed by five methods: Inverse Variance Weighted (IVW), MR Egger regression, weighted median, weighted, and simple models. Cochran's Q test and MR Egger regression were performed to analyze heterogeneity and horizontal pleiotropy, respectively.

Results: IVW associated higher household income with an increased rate of breastfeeding (OR = 1.05, $P = 4.05 \times 10^{-5}$). This finding was corroborated by weighted median (OR = 1.06, $P = 5.5 \times 10^{-4}$) and simple (OR = 1.09, $P = 0.02$) models. Reverse analysis showed that breastfeeding did not increase family income ($P > 0.05$). Quality control analyses indicated the absence of heterogeneity and pleiotropy.

Conclusion: Our study disclosed a positive causal relationship between household income and an increased rate of breastfeeding in Europeans. However, a causal relationship between breastfeeding and increased household income was not evident.

INTRODUCTION

The intricate relationship between socioeconomic factors and infant feeding practices, particularly breastfeeding, has garnered significant attention in both public health and academic research. The World Health Organization recommends exclusive breastfeeding during the first six months of life, as it is widely recognized as a crucial determinant of infant health and development, providing essential nutrients and immunologic benefits that can shape health outcomes across an individual's lifespan. However, disparities in breastfeeding rates often reflect underlying socioeconomic inequalities, raising concerns regarding the long-term health implications

for both infants and communities. For example, lower socioeconomic status is frequently associated with reduced breastfeeding initiation and duration, leading to adverse health outcomes for children from economically disadvantaged backgrounds [1,2].

Current strategies to promote breastfeeding often encounter substantial barriers that are linked primarily to socioeconomic conditions and cultural perceptions regarding infant feeding. Initiatives to support breastfeeding are frequently undermined by factors such as limited access to resources, lack of education regarding the benefits of breastfeeding, and negative societal attitudes towards breastfeeding in public [3]. Despite the

growing body of literature on the impact of socioeconomic status on breastfeeding, the nuances of this relationship remain insufficiently explored, particularly regarding the causal pathways that connect family income to infant feeding practices. Although some studies have reported robust associations between socioeconomic factors and breastfeeding rates [4], others have yielded mixed results, highlighting the complexity of this relationship. This discordance underscores a critical gap in the understanding of how family income affects breastfeeding behavior, and indicates a pressing need for further investigation [5]. A more nuanced exploration of this relationship could yield valuable insights to inform public health policies and interventions to promote breastfeeding, especially among low-income populations.

The primary objective of this study was to investigate causal relationships between family income and breastfeeding behaviors from both forward and reverse perspectives by utilizing a bidirectional two-sample Mendelian Randomization (MR) approach, utilizing publicly available data from Genome-Wide Association Studies (GWAS) of European populations. This methodological framework permitted a rigorous examination of causal relationships between family income and breastfeeding practices, leveraging Single Nucleotide Polymorphisms (SNPs) as instrumental variables. The advantage of employing MR lies in its ability to infer causal relationships while controlling for confounding factors and minimizing the impact of reverse causation, a common challenge in observational studies [6].

By elucidating these relationships, this study aimed to enhance the understanding of how socioeconomic factors shape maternal health decisions, and to ultimately inform public health initiatives and interventions to improve breastfeeding rates among economically disadvantaged families. This exploration not only holds the potential to advance academic discourse but may also translate into practical strategies for addressing health disparities related to infant nutrition [7,8].

In summary, this study is positioned to fill a critical research void by elucidating the interplay between family income and breastfeeding practices. By employing advanced statistical approaches and leveraging genetic data, we sought to provide insights that may inform public health strategies to promote optimal infant feeding practices in the context of socioeconomic inequality.

MATERIALS AND METHODS

Source MR Must Follow Three Core Assumptions

1. Genetic variation and exposure factors are highly

correlated; 2. Genetic variation is not affected by confounding factors such as environment; 3. Genetic variation can only affect outcomes through exposure factors, not through other means [9]. To mitigate the impact of population stratification, all included samples were from the European population. All instrumental variables involved in MR analysis were downloaded from the IEU Open GWAS database (<https://gwas.mrcieu.ac.uk/datasets/>). Household incomes were obtained from GWAS data released in 2018 (ID: ukb-b-7408) acquired from 39,7751 European participants. Breastfeeding data were obtained from GWAS data released in 2021 (ID: ukb-1-33) involving 25,5881 European participants, including 181,621 patients and 74,260 controls. Positive research used household income as the exposure factor and breastfeeding as the outcome factor; reverse analysis designated breastfeeding as an exposure factor and household income as an outcome factor. Because all data used in this investigation were published previously, an additional ethical review of our study was not required.

Selection of Instrumental Variables and Quality Control

We first extracted SNPs significantly associated with whole genome exposure ($P < 5 \times 10^{-8}$) [10]. Second, the parameter threshold for linkage disequilibrium was set to $r^2 < 0.001$, with a regional range of kb=10000, to remove SNPs with linkage disequilibrium. Third, SNPs with minor allele frequencies below 0.01 were removed. Finally, we removed palindrome and incompatible alleles. We used F-score to determine whether the selected SNP was affected by weak instrumental variables. F-score was calculated by employing the formula $F = \beta^2 / s^2$ [11], where β is the effect value of SNP exposure and s is the standard error of β . An F-value greater than 10 indicates an absence of bias; consequently, SNPs with F-scores < 10 were excluded [12].

Statistical Analysis

The Inverse Variance Weighted (IVW) method was used as the principle approach for calculating causal effects. The MR Egger regression, weighted median, simple, and weighted modes were used for the comprehensive evaluation of causal effects [13]. IVW is the primary analytical tool used to estimate potential causal relationships because it provides the most accurate results; selected SNPs are effective instrumental variables [14]. The MR Egger regression and weighted median methods are used to improve IVW estimation as they provide more reliable estimates in a wider range of cases [15,16]. The weighted mode clusters SNPs into a subset based on the similarity of causal effects, thereby estimating causal

effects in the subset with the highest number of SNPs [17]. The strength of the causal relationship between household income and breastfeeding was expressed using Odds Ratio (OR) and 95% confidence interval (95% CI). Heterogeneity testing evaluated the differences between individual instrumental variables, using Cochran's Q test with a $P > 0.05$ indicating the absence of heterogeneity [18]. The level of pleiotropy between multiple instrumental variables was identified by the intercept term of the MR Egger method, with $P > 0.05$ indicating an absence of pleiotropy [19]. We constructed funnel plots to evaluate potential horizontal pleiotropy, similar to the method used in meta-analyses to assess bias. The 'leave one out' sensitivity analysis was used to evaluate whether the causal relationship between exposure and outcomes was influenced by any single SNP [20], and the results were presented in forest plots. All the above MR and quality control analyses were conducted using R software (version 4.4.1) and R package Two Sample MR (version 0.6.8).

RESULTS

Causal Relationship between Household Income and Breastfeeding (Positive Study)

MR Analysis Results: Significant ($P < 5 \times 10^{-8}$) and independent ($r^2 < 0.001$, kb = 10000) SNPs were included. After removing the palindrome sequence, 43 instrumental variables were included. For these instrumental variables, all F-values were >10 . These variables conform to the strong correlation hypothesis of MR and are less affected by the bias of weak instrumental variables. IVW associated increased household income with a higher rate of breastfeeding (OR = 1.05, 95% CI: 1.03-1.08, $P = 4.06 \times 10^{-5}$). The simple (OR = 1.09, 95% CI: 1.02-1.18, $P = 0.02$) and weighted median modes (OR = 1.06, 95% CI: 1.02-1.09, $P = 5.5 \times 10^{-4}$) showed that higher household income increased the rate of breastfeeding. Furthermore, the causal effects identified by the five analysis methods were consistent (OR values were all greater than 1), supporting a positive causal relationship between household income and breastfeeding (Table 1).

Sensitivity Analyses: Cochran Q test showed $Q = 45.95$, $P = 0.27$, indicating no heterogeneity of SNPs ($P > 0.05$). The MR Egger intercept method (intercept = 1.2×10^{-4} , $P = 0.9$) demonstrated no horizontal pleiotropy ($P > 0.05$) in the selected instrumental variable of household income. The "leave one" method analysis disclosed that the lack of a single SNP did not affect the causal correlation between household income and breastfeeding. Funnel plot symmetry indicated robust and reliable results (Table 2, Figure 1).

Table 1: Bidirectional MR analysis between household income and breastfeeding.

Exposure	Outcome	Method	nSNP	beta	P	OR	95%CI
Household income	Breastfed	MR Egger	43	0.04	0.4	1.04	0.95-1.15
		Simple mode	43	0.09	0.02	1.09	1.02-1.18
		Weighted mode	43	-4.20E-04	1	1	0.94-1.07
		Weighted median	43	0.06	5.50E-04	1.06	1.02-1.09
		IVW	43	0.05	4.05E-05	1.05	1.03-1.08
Breastfed	Household income	MR Egger	11	-0.46	0.46	0.63	0.19-2.0
		Simple mode	11	0.29	0.33	1.35	0.76-2.38
		Weighted mode	11	0.28	0.36	1.33	0.75-2.36
		Weighted median	11	0.09	0.59	1.09	0.78-1.54
		IVW	11	0.04	0.76	1.04	0.79-1.37

MR: Mendelian Randomization; SNP: Single Nucleotide Polymorphisms; IVW: Inverse Variance Weighted.

Table 2: Heterogeneity and pleiotropy tests for instrumental variables.

Outcome	Heterogeneity		Pleiotropy	
	Q	P	intercept	P
Breastfed	45.95	0.27	1.20E-04	0.9
Household income	11.8	0.22	4.40E-03	0.4

Causal Relationship between Breastfeeding and Household Income (Reverse Study)

MR Analysis: Breastfeeding was used as the exposure factor and household income as the outcome factor, with the same SNPs included as in the forward study. Finally, 11 instrumental variables were selected, and all F-values were >10 , indicating no bias. All 5 MR analyses showed no causal relationship between breastfeeding and increased household income, and the differences were not statistically significant ($P > 0.05$) (Table 1).

Sensitivity Analyses: Cochran Q test showed $Q = 11.8$, $P = 0.22$, indicating no heterogeneity of SNPs ($P > 0.05$). The MR Egger intercept method (intercept = 4.4×10^{-3} , $P = 0.4$) disclosed no significant pleiotropy ($P > 0.05$) in the selected BC instrumental variables. The "leave one" method showed that the lack of a single SNP did not affect the causal correlation between breastfeeding and household income. Funnel plots showed symmetrical data distribution, indicating robust and reliable results (Table 2, Figure 1).

DISCUSSION

This study aimed to explore the causal relationship between household income and breastfeeding by using MR analysis. We obtained data on household income

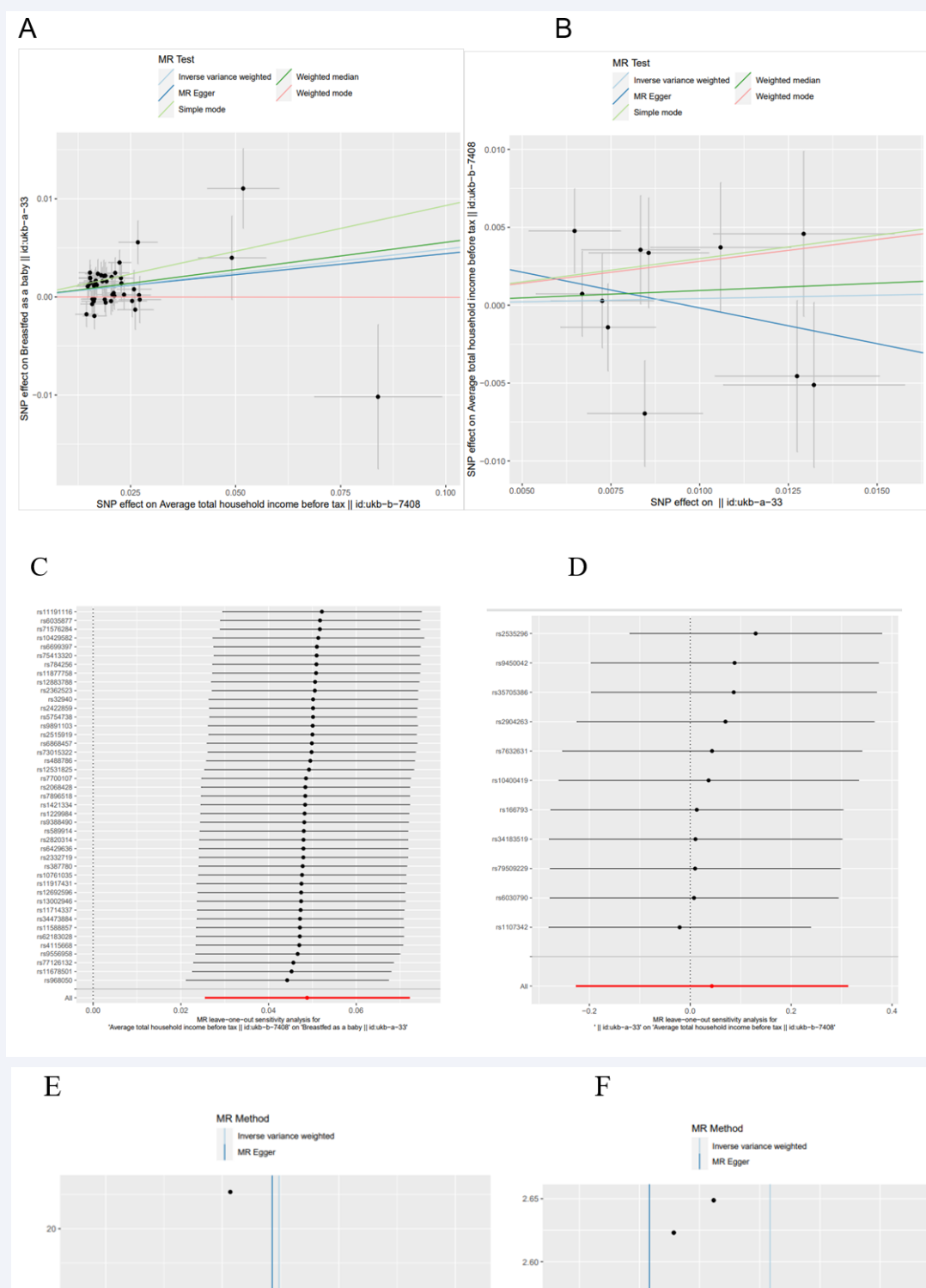


Figure 1 Scatter plots of causal associations, leave-one-out plots, and funnel plots generated by sensitivity analyses.

Scatter plot for estimating the effect of household income on the rate of breastfeeding.

Scatter plot for estimating the effect of breastfeeding on household income.

Leave-one-out analysis of the effect of household income on breastfeeding.

Leave-one-out analysis of the effect of breastfeeding on household income.

Heterogeneity of MR estimates of the impact of household income on breastfeeding.

Heterogeneity of MR estimates of the impact of breastfed on household income.

and breastfeeding behavior from a GWAS database to investigate this relationship. Using multiple MR methods, including IVW and MR Egger regression, we evaluated the causal effects of household income on breastfeeding rates. Our findings disclosed a positive causal relationship between increased household income and the rate of breastfeeding in a European population, reinforcing the argument that socioeconomic factors play a pivotal role in maternal and child health decisions. This was consistent with previous research findings, which demonstrated that mothers with higher household income were more likely to breastfeed [21].

In contrast, the reverse analysis indicated that breastfeeding did not significantly increase family income, which challenges the assumption that breastfeeding might contribute to better financial outcomes for families. This implies that while breastfeeding provides numerous health benefits for infants and mothers, its impact on economic factors may not be as pronounced as previously thought. The absence of evidence of a reverse causal relationship aligns with findings that highlight the importance of socioeconomic factors in maternal health decisions [22].

Quality control analyses revealed no significant heterogeneity or pleiotropy, validating the robustness of our findings. The methodological rigor of this study including the use of multiple MR approaches strengthens the confidence in asserting a causal link between household income and breastfeeding. This is particularly crucial given the concerns regarding the validity of MR studies due to potential biases introduced by genetic variants that could have pleiotropic effects [23]. Moreover, these results underline the necessity for policymakers to consider economic support systems that promote breastfeeding as a public health strategy. Targeted interventions aimed at increasing household income could yield positive outcomes in breastfeeding rates, thereby improving infant health metrics and potentially reducing healthcare costs associated with formula feeding and its complications [24].

The implications of this study extend beyond the immediate findings, as they suggest a need for further exploration of the potential influence of socioeconomic factors on maternal behaviors and child health outcomes. Future research should focus on the pathways through which income affects breastfeeding practices, considering variables such as access to healthcare, maternal education, and community support systems [25]; and should address the development of comprehensive strategies to enhance breastfeeding rates through economic empowerment [26].

Limitations of this study must be acknowledged. First,

the reliance on GWAS data may have introduced bias, as the identified genetic variants may not have fully captured the complexities of socio-economic factors influencing both household income and breastfeeding practices. Additionally, the restriction of the analysis to a European population limits the generalizability of the findings to other ethnic groups or geographic regions. Furthermore, although multiple MR methods were employed to ensure robustness, unidentified and therefore unmeasured confounding factors may have influenced the observed associations. Lastly, the absence of heterogeneity and pleiotropy detected by our quality control analyses should be interpreted with caution, as genetic instruments may have pleiotropic effects that are not readily detectable.

In conclusion, this study establishes a positive causal relationship between increased household income and the likelihood of breastfeeding among Europeans, suggesting that socioeconomic factors play a significant role in breastfeeding practices. However, a reverse association, in which breastfeeding would influence household income, was not evident. Future studies should aim to expand the demographic scope and investigate the impact of additional confounding variables to provide a more comprehensive understanding of this relationship.

AUTHOR CONTRIBUTIONS

HK conceptualized the study. HH, JL, CW, ST, MS and QH, SY, HL, NW, YZ performed Mendelian randomization. YL, YZ designed the experiments. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

Joint Project on Health Science and Technology Innovation in Hainan Province (SQ2023WSJK0356), Key R&D Project of Hainan Provincial Department of Science and Technology (ZDYF2020139, ZDYF2018158), and College Student Innovation and Entrepreneurship Training Program Project (X202311810097).

REFERENCES

1. Engle PL. Influences of mothers' and fathers' income on children's nutritional status in Guatemala. *Soc Sci Med.* 1993; 37: 1303-1312.
2. Harkness S, Machin S, Waldfogel J. Evaluating the pin money hypothesis: the relationship between women's labour market activity, family income and poverty in Britain. *J Popul Econ.* 1997; 10: 137-158.
3. Ku I, Lee W, Lee S, Han K. The Role of Family Behaviors in Determining Income Distribution: The Case of South Korea. *Demography.* 2018; 55: 877-899.
4. Celi AC, Rich-Edwards JW, Richardson MK, Ken PK, Matthew WG. Immigration, race/ethnicity, and social and economic factors as

- predictors of breastfeeding initiation. *Arch Pediatr Adolesc Med*. 2005; 159: 255-260.
5. Zhang Q, Zhang X, Cui Q, Weining C, Ling H, Yexin Z, et al. The Unequal Effect of the COVID-19 Pandemic on the Labour Market and Income Inequality in China: A Multisectoral CGE Model Analysis Coupled with a Micro-Simulation Approach. *Int J Environ Res Public Health*. 2022; 19: 1320.
 6. Vasilyeva N, Blanchard T, Lombrozo T. Stable Causal Relationships Are Better Causal Relationships. *Cogn Sci*. 2018; 42: 1265-1296.
 7. Carneiro AV. Causal relationships in medicine. *Rev Port Cardiol*. 2005; 24: 443-449.
 8. Ancin-Murguzur FJ, Hausner VH. Causalize R: a text mining algorithm to identify causal relationships in scientific literature. *Peer J*. 2021; 9: e11850.
 9. Birney E. Mendelian randomization. *Cold Spring Harb Perspect Med*. 2022; 12: a041302.
 10. Fang A, Zhao Y, Yang P, Xuehong Z, Edward LG. Vitamin D and human health: evidence from Mendelian Randomization studies. *Eur J Epidemiol*. 2024; 39: 467-490.
 11. Davies NM, Holmes MV, Davey SG. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ*. 2018; 362: k601.
 12. Shu MJ, Li JR, Zhu YC, Hang S. Migraine and ischemic stroke: a mendelian randomization study. *Neurol Ther*. 2022; 11: 237-246.
 13. Chen LL, Yang HQ, Li HT, Chang H, Liu Y, Guoyue L. Insights into modifiable risk factors of cholelithiasis: A Mendelian randomization study. *Hepatology*. 2022; 75: 785-796.
 14. Pan SY, Zhang ZQ, Pang WY. The causal relationship between bacterial pneumonia and diabetes: a two-sample mendelian randomization study. *Islets*. 2024; 16: 2291885.
 15. Cheng ZJ, Gao L, Huang P, Dingwen L, Chunsheng B, Xiaoyan C. Genetic causal association between rheumatoid arthritis and periodontitis: a bidirectional two-sample Mendelian randomization analysis. *Clin Oral Investig*. 2024; 28: 107.
 16. Bowden J, Holmes MV. Meta-analysis and Mendelian randomization: a review. *Res Synth Methods*. 2019; 10: 486-496.
 17. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. 2017; 32: 377-389.
 18. Bowden J, Davey SG, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016; 40: 304-314.
 19. Yin KJ, Huang JX, Wang P, Xiao-Ke Y, Sha-Sha T, Hong-Miao L, et al. No genetic causal association between periodontitis and arthritis: a bidirectional twosample Mendelian randomization analysis. *Front Immunol*. 2022; 13: 808832.
 20. Hemani G, Zheng J, Elsworth B, Kaitlin HW, Valeriia H, Denis B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018; 7: e34408.
 21. Tanaka R, Tsuji M, Shimono M, Kaitlin HW, Valeriia H, Denis B, et al. Japan Environment and Children's Study Group. Association between maternal socioeconomic status and breastfeeding: Results from the Japan environment and children's study. *J Child Health Care*. 2024; 28: 715-728.
 22. Buchsbaum D, Seiver E, Bridgers S, Gopnik A. Learning about causes from people and about people as causes: probabilistic models and social causal reasoning. *Adv Child Dev Behav*. 2012; 43: 125-160.
 23. Statham G. Mechanisms, the interventionist theory, and the ability to use causal relationships. *Stud Hist Philos Sci*. 2018; 72: 22-31.
 24. Cummins DD, Lubart T, Alksnis O. Conditional reasoning and causation. *Mem Cognit*. 1991; 19: 274-282.
 25. Paris A, Amirthalingam G, Karania T, Foote IF, Dobson R, Noyce AJ, et al. Depression and dementia: interrogating the causality of the relationship. *J Neurol Neurosurg Psychiatr*. 2025; 96: 573-581.
 26. Nijs J, Van de Velde F, Cuyckens H. Is Word Order Responsive to Morphology? Disentangling Cause and Effect in Morphosyntactic Change in Five Western European Languages. *Entropy (Basel)*. 2025; 27: 53.