ORIGINAL ARTICLE

The number of lifetime sexual partners may not be associated with cervical cancer: A Mendelian randomization study

Kan Chen¹, Li Wan¹, Fang Wang¹, Yingxue Liu¹, Jinyan Tang², Lu Han^{1,*}

¹Department of Gynaecology, Xinjiang Uygur Autonomous Region Hospital of Traditional Chinese Medicine, Urumqi 830000, Xinjiang Uygur Autonomous Region, China

²Xinhua Hospital of Ili Kazak Autonomous Prefecture, Ili Kazak Autonomous Prefecture, Urumqi 835000, Xinjiang Uygur Autonomous Region, China

ABSTRACT

Objective: The objective of this study was to investigate the causal relationship between the number of lifetime sexual partners (NLSP) and cervical cancer (CC) using the Mendelian randomization (MR) method. Methods: Genome-wide association study (GWAS) data on NLSP and CC were obtained from the integrative epidemiology unit open genome-wide association studies (IEU OpenGWAS) project. To assess the correlation between NLSP and CC risk, we employed the inverse-variance weighted (IVW) method, applying a preset threshold to select single nucleotide polymorphisms (SNPs) closely related to NLSP. Instrumental variables (IVs) were constructed using isolated SNPs. Heterogeneity among the SNPs was evaluated using the Cochran Q test. The presence of abnormal SNPs was tested with MR pleiotropy residual sum and outlier test (MR-PRESSO). The Mendelian randomization-Egger (MR-Egger) intercept test was conducted to examine horizontal pleiotropy among the SNPs. Additionally, the "leave-one-out" sensitivity analysis was performed to assess whether the MR results were influenced by any single SNP. Results: A total of 63 SNPs correlated with NLSP were screened. IVW analysis revealed no causal relationship between NLSP and CC, with an odds ratio (OR) of 1.001, 95% confidence interval (CI): 0.996 - 1.005, P = 0.797. The Cochran Q test indicated no significant heterogeneity among the included SNPs (Q = 73.051, P = 0.07). The MR-Egger intercept value was 1.61×10^{-5} (P = 0.903), suggesting no genetic pleiotropy among the screened SNPs. MR-PRESSO did not identify any outlier SNPs. Furthermore, the "leave-one-out" sensitivity analysis indicated that the causal estimates were unlikely to be influenced by specific SNP effects. Conclusion: Our findings suggest that there may be no causal relationship between the NLSP predicted by genetics and the risk of CC.

Key words: Mendelian randomization, lifetime number of sexual partners, cervical cancer, causality

INTRODUCTION

Cervical cancer (CC) stands as the most prevalent malignant tumor-like disease in gynecology, commonly presenting with clinical symptoms such as vaginal bleeding, discharge, and secondary tumor-related manifestations. Existing research suggests that its occurrence is primarily linked to *Human Papillomavirus* (HPV) virus infection, with sexual behavior, the number of deliveries, and other biological factors potentially contributing to its pathogenesis.^[1,2] Since the mid-1970's, the incidence of CC has declined by more than half due to widespread screening practices. The introduction of the first HPV vaccine in 2006 has further reduced CC

*Corresponding Author:

Han Lu, Xinjiang Uygur Autonomous Region Hospital of Traditional Chinese Medicine, Urumqi 830000, Xinjiang Uygur Autonomous Region, China. Email: cckk960504@163.com; ORCID: https://orcid.org/0009-0009-6245-8432 Received: 28 October 2024: Revised: 18 December 2024; Accepted: 24 December 2024

https://doi.org/10.54844/cif.2024.0769

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which allows others to copy and redistribute the material in any medium or format non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

prevalence, making it one of the most preventable cancers. However, according to the latest global cancer statistics in 2023, CC remains the third most common female malignancy, with the second highest morbidity and mortality rates. Differences in CC screening and HPV vaccine coverage lead to varying incidence trends across age, race, and ethnicity, influenced further by economic ability and educational environment.^[3,4] These disparities pose a significant threat to women's health and increase the economic burden on society. Numerous observational studies have indicated that high-risk sexual behaviors are a primary cause of CC incidence, yet they have failed to establish a causal relationship between the two.^[5-7] Due to the private nature of sexual behavior and the traditional observational epidemiological design of these studies, sample sizes are often limited, and they are susceptible to confounding factors and reverse causality. These limitations hinder the ability to prove a causal link between exposure and outcome. Mendelian randomization (MR) offers a solution to these challenges. MR addresses the limitations of observational studies by leveraging whole genome sequencing data and the Mendelian second law to reveal causality.^[8] Its core principle involves using single nucleotide polymorphisms (SNPs) as instrumental variables (IVs) to demonstrate a causal relationship between exposure and outcome. Since parental alleles are passed to offspring following the "random distribution principle" during meiosis, MR is often referred to as a "natural randomized controlled trial," circumventing the interferences of reverse causation and confounding factors in traditional epidemiological studies.^[9] Sexual behavior encompasses various aspects, including the number of lifetime sexual partners (NLSP), sexual age, and frequency. Currently, only a genome-wide association studies (GWAS) database on NLSP is available. Therefore, this study aims to explore the causal relationship between NLSP and CC using the MR method.

DATA AND METHODS

Research principles

The rationale behind MR lies in leveraging genetic variants associated with both the exposure and the outcome as IVs to deduce the presence of a causal relationship between them. In this study, we employed the NLSP as the exposure factor. Genetic variants, specifically SNPs significantly correlated with the NLSP, served as the IVs, while CC constituted the outcome variable. Following the removal of outliers, we conducted a two-sample MR (TSMR) analysis, incorporating heterogeneity and pleiotropy tests, to assess causality. The robustness of our findings was subsequently verified. The fundamental steps in our methodology encompassed: acquiring GWAS summary data, screening and evaluating SNPs, performing statistical analyses, and conducting quality assessments. The accuracy of our MR analysis hinges on three core assumptions, as outlined in: (1) IVs need to be closely related to exposure; (2) IVs was independent of confounding factors that affected the "exposure outcome."; (3) IVs only affected the outcome through exposure but not through other means, as shown in (Figure 1).^[10]

Data sources

Relevant data on the NLSP ^[11] and CC ^[12] were retrieved from the integrative epidemiology unit open genomewide association studies (IEU open GWAS database) (accessible at https://gwas MRcieu.ac.uk/). The study dataset for the NLSP, identified as ukb-b-4256, comprised 378,882 individuals and encompassed a total of 9,851,867 SNPs. Meanwhile, the CC study, identified as ieu-b-4876, included a sample size of 199,086 participants with 850,626 SNPs. Notably, both datasets originated from European populations.

Screening of IVs

IVs were screened based on the fundamental assumptions underlying MR. SNPs significantly associated with the exposure were extracted from the pooled GWAS database using a threshold of $P < 5 \times 10^{-8}$ to satisfy the association assumption. The TSMR tool in R4.2.2 software was utilized for clump calculations. The parameters kb = 10,000 and $r^2 = 0.001$ were set to cluster SNPs, thereby excluding the influence of linkage disequilibrium and ensuring that the IVs did not interact with each other.^[13] Subsequently, the SNPs significantly correlated with the NLSP were retrieved from the GWAS data on CC. The exposure and outcome data were adjusted to ensure consistency in direction, and SNPs with palindromic structures were eliminated. The exposure and outcome datasets were then collated and merged. To address horizontal pleiotropy, outlier tests were conducted using Mendelian randomization pleiotropy residual sum and outlier test (MR-PRESSO).^[14] This step involved removing outliers and excluding SNPs directly associated with CC,^[15] thereby refining the set of IVs used in the MR analysis.

Statistical analysis

Estimation of the causal effect of the NLSP and CC This study will employ five distinct methods to estimate the causal relationship between the exposure and outcome variables: the the inverse-variance weighted (IVW) method, the Weighted Median (WM) method, the Mendelian randomization-Egger (MR-Egger) method, the Simple Mode method, and the Weighted Mode method.^[16] The IVW method serves as the primary tool for causality estimation and is widely regarded as the standard approach in MR analysis. When all screened



Figure 1. The Mendelian randomization model of the number of lifetime sexual partners and cervical cancer. IVs, instrumental variables

SNPs are valid IVs, the IVW method can accurately estimate the causal effect by combining the Wald ratios calculated for each IV, utilizing a random effects model in the presence of heterogeneity and a fixed effects model in its absence.^[14] The MR-Egger method is employed to detect potential pleiotropy,^[17] while the WM method necessitates that more than half of the IVs be valid SNPs to provide consistent causal effect estimates.^[18] In scenarios where up to half of the IVs may be invalid, the WM method can still calculate consistent estimates of causal effects.^[19] Lastly, the Simple Mode method allows for grouping of SNPs with similar effects, based on the similarity of the estimated causal effects,^[10] thereby facilitating a more nuanced analysis of the causal relationship between the exposure and outcome variables.

Quality control

In this study, we employed a series of rigorous methods to assess the stability and reliability of the MR results. Firstly, we calculated the F values for each individual SNP and conducted a weak IV bias test on the selected IVs. The *F* value was derived using the formula $F = \beta^2$ exposure/SE² exposure, where β represents the allelic effect size of the exposure, and SE denotes the standard error of the exposure.^[20] Secondly, we evaluated the heterogeneity among the SNPs using the Cochran Q test. A statistically significant result from the Cochran Q test indicated the presence of notable heterogeneity within the study. Thirdly, we utilized the MR-Egger method to detect any pleiotropic effects associated with the SNPs. A statistically significant MR-Egger test result suggested that the analysis contained significant pleiotropy. Fourthly, we analyzed MR residuals and outliers to identify any outlier SNPs in the results. If any outliers were detected, they were excluded from the analysis, and the process was repeated. Fifthly, to further assess the stability and reliability of our results, we recalculated the combined effect of the remaining SNPs after individually removing each one. For the purposes of this study, we conducted the MR analysis and quality control using the TSMR package in R version 4.2.2 software, with a significance level set at $\alpha = 0.05$.

RESULTS

Screened IVs

In this study, we identified 63 SNPs that exhibited a significant correlation with the NLSP. Subsequently, we retrieved and extracted the pooled data from CC, noting that 3 SNPs were absent from the CC GWAS dataset. Additionally, two palindromic sequences, namely rs2194027 and rs28457046, were excluded from our analysis. Furthermore, outliers were not considered based on the MR-PRESSO test. Ultimately, a total of 58 SNPs were selected for analysis, all of which had F values exceeding 10, indicating a reduced likelihood of the presence of weak variables (Table 1).

 Table 1: Summary of the number of lifetime sexual partners and cervical cancer.

Item	Sample size	No. of SNPs	Race	Gender	Year
Lifetime number of sexual partners	378,882	9,851,867	European	Male and female	2018
Cervical cancer	199,086	8,506,261	European	Female	2021

MR results for both samples

IVW results showed that odds ratio (OR) = 1.001, 95% confidence interval (CI): 0.996 - 1.005, P = 0.797, suggesting that the NLSP may not be related to the incidence of CC, the forest diagram analyzed by MR is shown in Figure 2, and the MR-Egger method as a complementary to the IVW result also shows that there is no obvious causal relationship between NLSP and CC, and shows the relevant results of the five MR Tests (Table 2). The scatter plot is shown in (Figure 3).

Quality control

The Cochran Q test yielded results with Q = 73.051 and P = 0.07, suggesting an absence of heterogeneity among the selected SNPs. Furthermore, the MR-Egger analysis produced a value of 1.61×10^{-5} with P = 0.903, indicating that the causal effect analysis was unperturbed by pleiotropic effects. The funnel plot exhibited approximate symmetry on both sides (Figure 4),



Figure 2. Forest plot of the number of lifetime sexual partners and cervical cancer. The horizontal axis represents the exposure effect, while the vertical axis represents the names of single nucleotide polymorphisms. Each black dot signifies an individual single nucleotide polymorphism. The horizontal lines depict the effect sizes and confidence intervals of each individual single nucleotide polymorphism on the outcome.

 Table 2: The results of the five Mendelian randomization methods.

Method	nSNP	OR (95% CI)	P value
MR Egger	58	0.999 (0.980 - 1.019)	0.950
Weighted median	58	1.002 (0.999 - 1.008)	0.423
Inverse variance weighted	58	1.001 (0.996 - 1.005)	0.797
Simple mode	58	1.007 (0.992 - 1.022)	0.363
Weighted mode	58	1.007 (0.992 - 1.021)	0.378

suggesting a minimal likelihood of being influenced by potential biases. Additionally, the MR-PRESSO results indicated no detection of outlier SNPs (P = 0.072). The "leave-one-out" method revealed that, upon sequentially eliminating individual SNPs through a one-by-one elimination test, the results for the remaining 57 SNPs were consistent with those obtained from the IVW analysis that encompassed all SNPs. Notably, all these SNPs resided on the right side of the null line (Figure 5), affirming the reliability of the MR analysis.

DISCUSSION

This TSMR analysis has confirmed the association between NLSP and CC. Utilizing five distinct analytical methods, namely the IVW, MR-Egger, WM, Weighted Mode-based Estimate (WME), and Simple Mode, the study revealed the absence of a statistically significant causal link between NLSP and CC. It is welldocumented that chronic HPV infection constitutes a pivotal risk factor for the development of CC. Preventive measures aimed at controlling HPV infection have the potential to diminish the incidence of CC. Notably, contemporary research has established HPV infection as the most prevalent sexually transmitted disease, with high-risk sexual behaviors significantly enhancing women's susceptibility to HPV. These behaviors include having multiple sexual partners, engaging in frequent sexual activity, and initiating sexual activity at an early age.[21,22]

A study conducted in China, with a sample size of 98, 036 participants, revealed that the primary risk factors



Figure 3. Scatter plot of the number of lifetime sexual partners and cervical cancer. Different colors represent different Mendelian randomization methods. Each black dot signifies an individual single nucleotide polymorphism. The horizontal lines depict the 95% confidence interval for the corresponding exposure factors, while the vertical lines represent the confidence intervals for the outcome indicators.



Figure 4. Funnel plot of the number of lifetime sexual partners and cervical cancer. Each black dot represents an individual single nucleotide polymorphism, and the funnel plot shows a roughly symmetrical distribution on both sides, indicating a low likelihood of being affected by potential bias.



Figure 5. "Leave-one-out" sensitivity analysis. Each black dot represents an individual single nucleotide polymorphism. The impact of the remaining single nucleotide polymorphisms on the outcome, after removing individual single nucleotide polymorphisms, is shown. The horizontal lines represent the 95% confidence interval. This figure demonstrates the robustness of the Mendelian randomization analysis results.

for women's susceptibility to HPV infection included the age of first sexual intercourse (OR = 1.62, 95% CI: 1.16 – 2.26), NLSP (OR = 1.49, 95% CI: 1.27–1.76), and the frequency of sexual intercourse (OR = 1.95, 95% CI: 1.45-2.62).^[23] Additionally, a Canadian study also demonstrated a strong association between HPV infection and NLSP (OR = 3.04, 95% CI: 1.99 - 4.65).^[24] Relevant research has concluded that sexual behavior is linked to the incidence of CC. The present study aimed to investigate whether a causal correlation exists between NLSP and CC from a genetic perspective through MR. However, the results indicated that NLSP may not elevate the risk of CC, which somewhat contrasts with the existing observational studies. Possible reasons for this discrepancy may include: (1) the NLSP may not increase the risk of CC in the case of fulfilling the safe sex conditions may have little effect on HPV infection; (2) The risk of infecting HPV has decreased with the use of the HPV vaccine; (3) screening for CC is becoming more common, and CC is being managed before it occurs; (4) HPV infection pathways are complex and interact with each other, while the human immune system clears HPV infection to a certain extent, which does not necessarily progress to CC. (5) The relevant GWAS data in this study were derived from the European population, and the GWAS data were small and the population was limited. Compared with the annual incidence of CC in the world and China, the small sample size may limit the statistical power of the study, increase the margin of error, and lead to less accurate estimates of associations, taking into account differences in ethnic, economic conditions and cultural norms. Different populations may have different genetic backgrounds, healthcare, and patterns of sexual behavior, which raises questions about the potential impact on the reliability of the results, possibly affecting the causal relationship between sexual partners and CC risk. This study is statistical and cannot further explore the mechanism. Combined with the current research, it can only be concluded that there is a correlation between the number of lifetime sexual partners and cervical cancer, but there may not be causality. Follow-up studies should be carried out among different ethnic groups and include larger and more representative samples to ensure the reliability of the conclusions. Further stratified studies should be conducted on whether there is safe sex in multiple sexual partners and the frequency of sexual life, so as to further explore the impact of the NLSP on CC. Further studies are expected to explore the causal effects of relevant exposure factors on disease. Although MR has significant advantages in establishing causality, limitations related to population diversity and sample size should be carefully addressed. These factors are essential to accurately interpret the results and ensure their applicability in different contexts. In summary, this study adopted TSMR Method to explore the causal relationship between the NLSP and CC and the results showed that there was no causal relationship between the NLSP and CC from the perspective of genetics.

The GWAS data were small and the population was limited. Compared with the annual incidence of CC in the world and China, the small sample size may limit the statistical power of the study, increase the margin of error, and lead to less accurate estimates of associations, taking into account differences in ethnic, economic conditions and cultural norms. Different populations may have different genetic backgrounds, healthcare, and patterns of sexual behavior, which raises questions about the potential impact on the reliability of the results, possibly affecting the causal relationship between sexual partners and CC risk. This study is statistical and cannot further explore the mechanism. Combined with the current research, it can only be concluded that there is a correlation between the NLSP and CC, but there may not be causality. Follow-up studies should be carried out among different ethnic groups and include larger and more representative samples to ensure the reliability of the conclusions.

CONCLUSION

In summary, this study adopted the TSMR Method to discuss the causal relationship between the NLSP and CC, and the results indicated that there was no causal relationship between the NLSP and CC from the perspective of genetics.

DECLARATION

Author contributions

Chen K conceived and wrote the paper, Wan L, Wang F searched the related data, Liu YX, Tang JY carried out statistical analysis, Han L reviewed the manuscript. All authors reviewed the manuscript.

Ethics approval

Not applicable.

Source of funding

This study was supported by the Open Project of the State Key Laboratory of Co-construction of the Causes and Prevention of High Morbidity in Central Asia (SKL-HIDCA-2020-ZY6).

Conflict of interest

The authors have declared that they have no competing interests.

Data availability statement

The original contributions or analyzed during the current study are available from the article. Further inquiries can be made to the corresponding author.

REFERENCES

- Cibula D, Rosaria Raspollini M, Planchamp F, et al. ESGO/ESTRO/ ESP Guidelines for the management of patients with cervical cancer–Update 2023. *Radiother Oncol.* 2023;184:109682.
- Williamson A-L. Recent developments in human papillomavirus (HPV) vaccinology. *Viruses*, 2023;15(7):1440.
- Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA A Cancer J Clinicians. 2023;73(1):17–48.
- Tadesse F, Megerso A, Mohammed E, Nigatu D, Bayana E. Cervical cancer screening practice among women: A community based crosssectional study design. *INQ*. 2023;60:00469580231159743.
- Prue G, Lawler M, Baker P, Warnakulasuriya S. Human papillomavirus (HPV): making the case for 'immunisation for all'. Oral Dis. 2017;23(6):726-730.
- Tesfaw K, Kindie W, Mulatu K, Bogale EK. Utilisation of cervical cancer screening and factors associated with screening utilisation among women aged 30–49 years in Mertule Mariam Town, East Gojjam Zone, Ethiopia, in 2021: a cross-sectional survey. *BMJ Open*. 2022;12(11):e067229.
- Gavrankapetanovic F, Sljivo A, Dadic I, Mehmedbasic N. Epidemiological aspects of age and genotypical occurrence of HPV infection among females of Canton Sarajevo over a 10-year period. *Mater Sociomed.* 2022;34(4):260–3.
- Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *Jama*. 2017;318(19):1925–1926.
- Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing mendelian randomization investigations: update for summer 2023. Wellcome Open Res. 2023;4:186.
- 10. Birney E. Mendelian randomization. *Cold Spring Harb Perspect Med.* 2022:17;12(4):a041302.
- Lu Z, Sun Y, Liao Y, et al. Identifying causal associations between early sexual intercourse or number of sexual partners and major depressive disorders: A bidirectional two-sample Mendelian randomization analysis. J Affect Disord. 2023;333:121–129.
- 12. Burrows K, Haycock P. Genome-wide Association Study of Cancer Risk in UK Biobank. University of Bristol. 2021.
- Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife*. 2018;7:e34408.
- Chen X, Kong J, Diao X, et al. Depression and prostate cancer risk: A mendelian randomization study. *Cancer Med.* 2020;9(23):9160–7.
- Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol.* 2016;45(6):1717–26.
- Hwang LD, Lawlor DA, Freathy RM, Evans DM, Warrington NM. Using a two-sample Mendelian randomization design to investigate a possible causal effect of maternal lipid concentrations on offspring birth weight. *Int J Epidemiol.* 2019;48(5):1457–1467.

- Slob EAW, Groenen PJF, Roy Thurik A, Rietveld CA. A note on the use of Egger regression in Mendelian randomization studies. *Int J Epidemiol.* 2017;46(6):2094–7.
- Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization *via* the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985–98.
- Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted Median estimator. *Genet Epidemiol.* 2016;40(4):304–14.
- Papadimitriou N, Dimou N, Tsilidis KK, et al. Physical activity and risks of breast and colorectal cancer: A Mendelian randomisation analysis. *Nat Commun.* 2020;11(1):597.
- 21. Chelimo C, Wouldes TA, Cameron LD, Mark Elwood J. Risk factors

for and prevention of human papillomaviruses (HPV), genital warts and cervical cancer. J Infect. 2013;66(3):207–17.

- 22. Torres-Poveda K, Ruiz-Fraga I, Madrid-Marina V, Chavez M, Richardson V. High risk HPV infection prevalence and associated cofactors: a population-based study in female ISSSTE beneficiaries attending the HPV screening and early detection of cervical cancer program. *BMC Cancer.* 2019;19(1):1205.
- Tana Z, Yinru Z, Rui W, Zhigang Z, Jie C, Kexin J. Meta-analysis of risk factors for HPV infection in Chinese women. *Chin J AIDS STD*. 2022;28(11):1334–8.
- Sellors JW, Mahony JB, Kaczorowski J, et al. Prevalence and predictors of human papillomavirus infection in women in Ontario, Canada. Survey of HPV in Ontario Women (SHOW) Group. CMAJ. 2000;163(5):503–8.