



Smoking and alcohol by HPV status in head and neck cancer: a Mendelian randomization study

Received: 12 February 2024

Accepted: 12 August 2024

Published online: 07 September 2024

 Check for updates

Abhinav Thakral^{1,32}, John J.W. Lee^{2,3,32} , Tianzhichao Hou¹, Katrina Hueniken¹, Tom Dudding^{4,5}, Mark Gormley^{4,5}, Shama Virani⁶, Andrew Olshan⁷, Brenda Diergaarde⁸, Andrew R. Ness⁹, Tim Waterboer¹⁰, Karl Smith-Byrne⁶, Paul Brennan⁶, D. Neil Hayes¹¹, Eleanor Sanderson⁴, M. Catherine Brown¹², Sophie Huang^{12,13}, Scott V. Bratman^{12,13}, Anna Spreafico^{12,14}, John De Almeida^{2,12}, Joel C. Davies^{2,3}, Laura Bierut¹⁵, Gary J. Macfarlane¹⁶, Pagona Lagioui¹⁷, Areti Lagioui¹⁸, Jerry Polesel¹⁹, Antonio Agudo²⁰, Laia Alemany²⁰, Wolfgang Ahrens²¹, Claire M. Healy²², David I. Conway²³, Mari Nygard²⁴, Cristina Canova²⁵, Ivana Holcatova²⁶, Lorenzo Richiardi²⁷, Ariana Znaor²⁸, David P. Goldstein^{2,12}, Rayjean J. Hung^{29,30}, Wei Xu^{1,30,33}, Geoffrey Liu^{10,12,30,33}  & Osvaldo Espin-Garcia^{1,30,31,33}

HPV-positive and HPV-negative head and neck squamous cell carcinoma (HNSCC) are recognized as distinct entities. There remains uncertainty surrounding the causal effects of smoking and alcohol on the development of these two cancer types. Here we perform multivariable Mendelian randomization (MR) to evaluate the causal effects of smoking and alcohol on the risk of HPV-positive and HPV-negative HNSCC in 3431 cases and 3469 controls. Lifetime smoking exposure, as measured by the Comprehensive Smoking Index (CSI), is associated with increased risk of both HPV-negative HNSCC (OR = 3.03, 95%CI:1.75-5.24, $P = 7.00 \times 10^{-5}$) and HPV-positive HNSCC (OR = 2.73, 95%CI:1.39-5.36, $P = 0.003$). Drinks Per Week is also linked with increased risk of both HPV-negative HNSCC (OR = 7.72, 95%CI:3.63-16.4, $P = 1.00 \times 10^{-7}$) and HPV-positive HNSCC (OR = 2.66, 95%CI:1.06-6.68, $P = 0.038$). Smoking and alcohol independently increase the risk of both HPV-positive and HPV-negative HNSCC. These findings have important implications for understanding the modifying risk factors between HNSCC subtypes.

Head and neck cancer, the seventh most common malignancy worldwide, accounts for more than 870,000 cases and 440,000 deaths annually^{1,2}. Head and neck squamous cell carcinomas (HNSCC) account for ~90% of all cases, with tobacco use and excessive alcohol consumption considered among the most significant modifiable risk factors³⁻⁶. Additionally, Human Papilloma Virus (HPV) plays a significant role in the pathogenesis of HNSCC, particularly for oropharyngeal cancer (OPC). HPV-associated HNSCC is now recognized as a

distinct entity from HPV-negative HNSCC with different epidemiology, risk factors, treatment regimens, and prognosis⁷⁻¹¹. The latest edition of American Joint Committee on Cancer (AJCC) staging of OPC further highlights this distinction by incorporating HPV status in its staging¹².

Although the association between smoking and HPV-negative HNSCC is well established, uncertainties persist regarding the causal effects of tobacco smoking and HPV-positive HNSCC. Some studies have indicated a positive correlation, while others have found no link

between tobacco smoking and HPV-positive HNSCC^{13,14}. These uncertainties can be attributed, in part, to the limitations of observational studies, alongside small cohorts for this relatively rare cancer, heterogeneity in study design between cohorts, and lack of distinction between HPV-positive and negative HNSCC within the studies. Similarly, uncertainty persists regarding the differences in the association of alcohol consumption with HPV-positive HNSCC and HPV-negative HNSCC, although the interactive effects of alcohol consumption and HPV status in increasing the risk of HPV-positive OPC has been previously reported¹⁵. Our study examines the association of tobacco smoking and alcohol consumption with the risk of each distinct cancer separately in one of the largest study populations to date.

Mendelian randomization (MR) is a genetic epidemiological approach that utilizes single nucleotide polymorphisms (SNPs) randomized during meiosis as instrumental variables to infer the effect of an exposure on an outcome. This approach attempts to mitigate the limitations of observational studies, such as confounding, reverse causation, and measurement error¹⁶. MR is based on three key assumptions: 1) the genetic variants used as instruments for the exposure must be valid and robustly associated with the exposure, 2) there should be no measured or unmeasured confounding of the association between the genetic instrument and the outcome, and 3) the variants should have no independent effect on the outcome other than through the exposure of interest¹⁷. MR is useful for exposures such as smoking and alcohol consumption, where obtaining unconfounded estimates by randomizing individuals to such exposures would be unfeasible and unethical. Furthermore, multivariable MR allows for the simultaneous estimation of the independent and joint effects of two or more exposures on an outcome¹⁸, which is particularly relevant given that the combined exposure to tobacco and alcohol has been demonstrated to exert a significant synergistic effect on the incidence of HNSCC¹⁹.

Large-scale genome-wide association studies (GWAS) reported SNPs reliably associated with smoking and drinking behaviors^{20,21}. Using these SNPs as genetic instruments for the exposures and outcome data obtained from a large HNSCC GWAS²², we performed MR to estimate the risk effect of tobacco smoking and alcohol consumption on HPV-positive and HPV-negative HNSCC subtypes. A previous study from our group used MR to assess the independent causal effects of smoking and alcohol on HNSCC using summary genetic data²³. The current study utilizes individual-level genetic data plus the available HPV status information to conduct an MR study evaluating HNSCC risk stratified by HPV status. We used univariable and multivariable MR methods to demonstrate independent causal effects of smoking as well as drinking behaviors on the risk of both HPV-negative and HPV-

positive HNSCC. We also investigated the interactive effects between smoking and drinking behaviors with the two cancer types via factorial MR. Our study highlights similarities and differences between HPV-positive and HPV-negative HNSCC risk factors.

Results

Baseline characteristics of the study population, including smoking and drinking behavior exposures, stratified by HPV status, are presented in Supplementary Table 1. The numbers of independent SNPs included as instrument variables for each smoking and alcohol use behavior are provided in Table 1. The results for univariable MR are summarized in Table 1 and Fig. 1 for the primary smoking and alcohol consumption exposures evaluated (SI, CSI, DPW). The genetic instruments for SI comprised of 57 SNPs were found to be associated with the risk of both HPV-positive HNSCC [IVW, OR (95% CI) = 2.37 (1.33, 4.24), $P = 0.0003$] and HPV-negative HNSCC [IVW OR (95% CI) = 1.81 (1.19, 2.76), $P = 0.0005$].

The genetic instrument for CSI comprised by 90 SNPs with independent and robust associations with the lifetime smoking exposure indicator was associated with the risk of both HPV-negative HNSCC [IVW OR (95% CI) = 2.59 (1.37, 4.92), $P = 0.0004$] and HPV-positive HNSCC [IVW OR (95% CI) = 2.6 (1.2, 5.65), $P = 0.02$]. The odds ratios correspond to a standard deviation change in CSI, which is equivalent to an individual smoking 20 cigarettes a day for 15 years and quitting 17 years ago, or an individual smoking 60 cigarettes a day for 13 years and quitting 22 years ago.

Using 25 independent SNPs associated with DPW, increased alcohol consumption was associated with the risk of both HPV-negative HNSCC [IVW OR (95%CI) = 6.79 (2.68, 17.16), $P = 5.21E-05$] and HPV-positive HNSCC [OR (95%CI) = 3.58 (1.27, 10.14), P value = 0.02].

Multivariable MR results are summarized in Table 2 and Fig. 2. After controlling for DPW, lifetime smoking exposure as measured by CSI was associated with an increased risk of both HPV-negative HNSCC [OR (95%CI) = 3.03 (1.75, 5.24), $P = 7.00E-05$] and HPV-positive HNSCC [OR (95%CI) = 2.73 (1.39, 5.36), $P = 0.003$]. After controlling for CSI, the number of DPW was linked with the risk of both HPV-negative HNSCC [OR (95%CI) = 7.72 (3.63, 16.4), $P = 1.00E-07$] and HPV-positive HNSCC [OR (95%CI) = 2.66 (1.06, 6.68), $P = 0.038$]. The estimates of associations from the ridge regression MVMR analyses using the optimal lambda penalty parameter and MVMR Egger regression were consistent with the estimates from the IVW MVMR method (Table 2). Owing to the weak instrument strengths (<10), we conducted Q-statistic minimization yielding a Q-statistic of 188.20 ($P = 0.36$), implying a lack of heterogeneity after correction for weak instrument bias.

Table 1 | Univariable Mendelian randomization of smoking and alcohol consumption exposures on HNSCC stratified by HPV status

HPV status	Method	Smoking initiation	Comprehensive smoking index	Drinks per week			
		OR (95% CI)* SNPs, $N = 57$	P value	OR (95% CI) SNPs, $N = 90$	P value	OR (95% CI) SNPs, $N = 25$	P value
HPV-negative HNSCC vs controls	MR Egger	29.5 (3.46, 251)	3.11E-03	2.46 (0.19, 31.5)	0.49	13.4 (3.36, 53.7)	1.26E-03
	IVW	1.82 (1.20, 2.76)	4.99E-03	2.59 (1.37, 4.92)	3.45E-03	6.79 (2.68, 17.2)	5.21E-05
	Weighted median	2.20 (1.20, 4.04)	0.01	3.87 (1.6, 9.33)	2.61E-03	9.78 (3.24, 29.5)	5.25E-05
	Weighted mode	3.28 (0.92, 11.6)	0.07	6.78 (0.89, 51.7)	0.07	10.4 (3.5, 31.1)	3.14E-04
HPV-positive HNSCC vs controls	MR Egger	6.88 (0.35, 135)	0.21	0.42 (0.02, 9.06)	0.58	4.53 (0.93, 22.0)	0.07
	IVW	2.37 (1.33, 4.24)	3.46E-03	2.60 (1.20, 5.65)	0.02	3.58 (1.27, 10.1)	0.02
	Weighted median	2.41 (1.04, 5.58)	0.04	1.45 (0.47, 4.41)	0.52	3.17 (0.79, 12.7)	0.1
	Weighted mode	2.63 (0.53, 13.19)	0.24	1.03 (0.08, 13.9)	0.98	2.82 (0.73, 10.8)	0.14

All tests are two-sided. Odds ratios for CSI and drinks per week are for a standard deviation change.

HPV human papillomavirus, MR Mendelian randomization, OR odds ratio, CI confidence interval, IVW inverse variance-weighted.

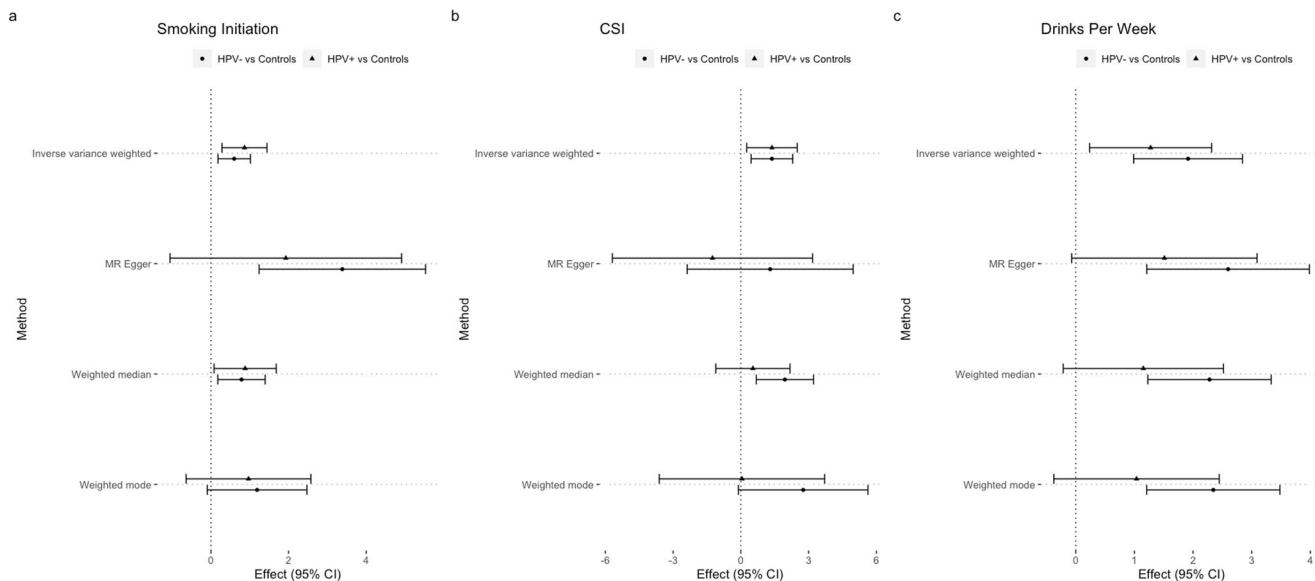


Fig. 1 | Forest plots of univariable Mendelian randomization effects of smoking and alcohol use exposures on HNSCC risk stratified by human papillomavirus (HPV) status. Univariable estimates were obtained using summary-level data from the GWAS of **a** smoking initiation ($n = 1232,091$), **b** comprehensive smoking index ($n = 462,690$), and **c** drinks per week ($n = 941,280$) on HPV-positive HNSCC risk ($n = 1105$ cases and 3469 controls) and HPV-negative HNSCC ($n = 2326$ cases and 3469 controls). Smoking initiation estimates are reported per log odds increase, while comprehensive smoking index and drinks per week are reported per SD increase in drinks per week. Error bars represent 95% confidence intervals. All statistical tests were two-sided. *CSI* comprehensive smoking index, *MR* Mendelian randomization.

Table 2 | Multivariable Mendelian randomization for smoking and alcohol consumption with risk of HNSCC by HPV status

Outcomes	Exposures	F statistic	Q-statistic for Instrument Validity	P value	Method	OR (95% CI)	P value
HPV-negative HNSCC vs controls	CSI	30.4	207.2	0.09	IVW	3.03 (1.75, 5.24)	7.00E-05
					MVMR Egger	2.49 (1.40, 4.43)	0.002
					Ridge Regression	4.93 (2.24, 10.66)	
	Drinks per Week	6.13			IVW	7.72 (3.63, 16.4)	1.00E-07
					MVMR Egger	14.02 (5.42, 36.28)	5.00E-08
					Ridge Regression	7.70 (2.94, 16.23)	
HPV-positive HNSCC vs controls	CSI	30.4	174.8	0.63	IVW	2.73 (1.39, 5.36)	0.003
					MVMR Egger	2.05 (1, 4.12)	0.048
					Ridge Regression	3.90 (1.62, 9.67)	
	Drinks per Week	6.13			IVW	2.66 (1.06, 6.68)	0.038
					MVMR Egger	6.51 (2.01, 21.12)	0.002
					Ridge Regression	2.49 (0.89, 5.63)	

All tests are two-sided. Odds ratios for CSI and drinks per week are for a standard deviation change. HPV human papillomavirus, CSI comprehensive smoking index, IVW inverse variance-weighted, MVMR multivariable Mendelian randomization, OR odds ratio, CI confidence interval.

Results of the factorial MR analysis are provided in the supplementary materials (Supplementary Table S2). We did not observe interaction effect between any smoking phenotypes with DPW, though factorial MR may lack the power to detect interactions, hence these results should be interpreted with caution. Additional smoking exposure phenotypes (AI, CPD, SC) were evaluated in exploratory analyses (Supplementary Table S3). Briefly, we did not identify different effects of these exposures on the two cancers, except CPD, which was positively associated with risk of HPV-negative HNSCC [for a single cigarette per day increase in smoking intensity, IVW OR (95% CI) = 1.59 (1.17, 2.17), $P = 0.0003$].

Lastly, we evaluated the association of risk tolerance and high-risk sexual behaviors with each cancer subtype as genetic loci of these exposures have been shown to overlap with those of

smoking and alcohol exposures²⁴. There were no associations between risk tolerance and number of sexual partners with the risk of HPV-positive or HPV-negative HNSCC (Supplementary Table S4).

Discussion

Utilizing univariable and multivariable MR, our study evaluated the causal effects of multiple smoking and alcohol use behaviors on the risk of HPV-positive and HPV-negative HNSCC. We observed that smoking and alcohol consumption independently increased the risk of both HPV-positive and HPV-negative HNSCC. Specifically, we found that both smoking and alcohol consumption were associated with both cancer types. These findings align with previous observational studies on the role of smoking and alcohol use in HPV-negative

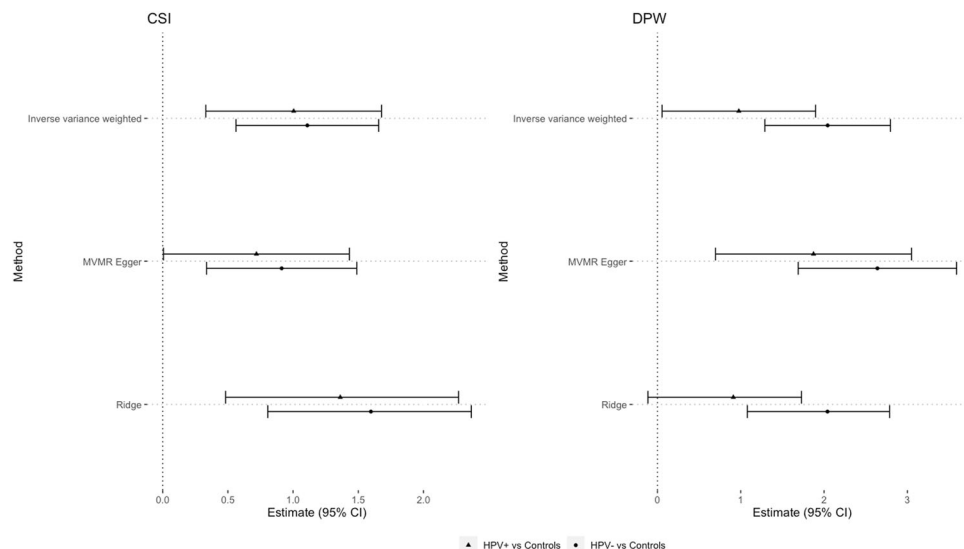


Fig. 2 | Forest plot of multivariable Mendelian randomization (MR) effects of lifetime smoking exposure and drinks per week on HNSCC risk stratified by human papilloma virus (HPV) status, using different MR approaches. Effect estimates were obtained using summary-level data for drinks per week ($n = 226,223$) and the comprehensive smoking index ($n = 226,223$) on HPV-positive HNSCC risk ($n = 1105$ cases and 3469 controls) and HPV-negative HNSCC

($n = 2326$ cases and 3469 controls). Comprehensive smoking index and drinks per week estimates are reported per SD change. Error bars represent 95% confidence intervals. All statistical tests were two-sided. CSI comprehensive smoking index; “drinks” refers to alcoholic drink equivalents, IVW inverse variance-weighted, HPV-, HPV-negative, HPV+, HPV-positive, CI confidence interval.

HNSCC, while providing evidence for the effects of these behaviors on HPV-positive HNSCC risks.

Large pooled observational studies have consistently supported tobacco smoking as an independent risk factor for HNSCC^{25–27}. In a meta-analysis of 15 case-control studies involving 10,244 HNSCC patients, Hashibe et al. reported a pooled OR of 2.13 for the association of cigarette smoking and HNSCC compared to never-smokers²⁶. More recently, Gormley et al. investigated the association between smoking and alcohol consumption on the risk of oral and OPC using a multivariable MR approach²³. After controlling for alcohol consumption, they report supporting evidence for a direct causal effect of lifetime smoking behavior on head and neck cancer risk (OR 2.6, 95% CI 1.7–3.9). When stratified by cancer subsite, the causal effect of cigarette smoking on OPC risk was even stronger, with risk estimates of 3.7 (95% CI 2.3–6.0) compared to 2.5 (95% CI 1.5–4.1) for oral cavity cancer. Utilizing MR on a subset of this study’s cohort, we report these associations separately for HPV-positive and HPV-negative HNSCCs.

To address the correlation between smoking and alcohol consumption, as well as to simultaneously explore their independent effects, we performed multivariable MR analyses on HPV-positive and negative HNSCC groups separately. Multivariable MR extends the basic MR framework to accommodate the complexity of multiple correlated exposures, enabling the evaluation of the independent causal effects of smoking and alcohol use on HNSCC risk. In our separate assessments of HPV-positive and HPV-negative HNSCCs, after correcting for alcohol consumption, we observed an independent causal effect of lifetime smoking on the risk of HPV-associated HNSCC, providing evidence of a significant contribution of smoking to the risk of HPV-associated HNSCC. While the association between smoking and HPV-negative HNSCC is well established, there has been uncertainty regarding the influence of smoking on the risk of developing HPV-associated HNSCC. Previous studies have reported conflicting evidence, with some demonstrating a positive association while others reporting no interactions between tobacco smoking and HPV status¹⁴. In North America, the incidence of HPV-associated HNSCC has risen over the past few decades despite declining smoking rates, in direct opposition to the decreasing incidence of all other HNSCCs²⁸. A pooled study by Anantharaman et al.¹³ reported associations of smoking with

an increased risk of HNSCC in models stratified by HPV16 seropositivity. Smoking is thought to act synergistically with HPV infection to increase the risk of developing cancer²⁹. This is possibly due to smoking suppressing mediators of immune function, thus facilitating the persistence of HPV infection, which is a crucial step in the development of HPV-related cancers³⁰. Notably, in our study, while smoking initiation increased the risk of HPV-positive HNSCC (Supplementary Table S3), we found no association between CPD and HPV-positive HNSCC. In contrast, CSI, a comprehensive index of smoking initiation, smoking intensity, and duration of exposure, increased the risk of both cancers.

Our MR analyses also revealed independent associations between alcohol consumption and increased risks of both HPV-negative and HPV-positive HNSCC. The strong co-existence of smoking and alcohol use has made it difficult to determine the independent effects of each. In one study, the joint effect of tobacco and alcohol was found to be more than multiplicative, but no marginal effect of alcohol use among never tobacco users was observed²⁶. In contrast, Gormley et al. reported an independent causal effect of alcohol consumption in oral and OPCs when controlling for smoking using an MR approach, although HPV status was not accounted for²³. In a meta-analysis evaluating traditional OPC risk factors, the summary odds ratio for the risk of OPC was 3.76 for heavy alcohol drinking and HPV negativity, whereas it was 39.32 for HPV positivity and no alcohol drinking¹⁵. Interestingly, the risk of OPC among those who were heavy alcohol drinkers and HPV-positive was 27.10, suggesting the presence of an interactive effect between alcohol use and HPV status in increasing the risk of cancer development. The factorial MR analysis did not detect any interactive effects between smoking and alcohol use, however the absence of interaction should be interpreted with caution. Multiplicative joint effects of smoking and alcohol use on head and neck cancer have been previously described from large observational cohort studies²⁶. Factorial MR has been shown to be limited in statistical power compared to conventional epidemiological approaches, due to variance and bias represented in genetic instruments³¹. Past investigations employing this approach in the context of cardiovascular disease and diabetes have yielded inconclusive results^{32,33}.

Our study has several strengths. Firstly, large, pooled analyses with individual-level data were performed incorporating individual-level HPV status. MR is a powerful approach to evaluating causal relationships between exposures and outcomes by utilizing genetic variants as instrument variables and subsequently overcoming limitations of conventional epidemiological approaches, such as confounding and reverse causality³⁴. We also used summary statistics from large GWAS studies of smoking, alcohol use, and head and neck cancer, utilizing numerous SNPs to ensure robust associations of our genetic instruments. As for limitations, several of the genetic loci used in our study have been previously associated with other exposures, such as sexual behaviors, which is a purported risk factor for HPV infection^{24,35}. MR approaches to delineate independent causal effects of sexual activity, such as the number of sexual partners, on HNSCC risk have so far been limited due to correlated pleiotropy and non-specification of these sexual behavior instruments³⁶. Furthermore, the lack of sex-specific instrument exposure information prevented the assessment of smoking and alcohol use stratified by sex, which is particularly relevant given the differences in exposures seen across males and females. Factorial MR may be inefficient at detecting statistical interactions due to the variance explained by genetic instruments and the potential for weak instrument bias, compared to the robustness of a clinical trial or observational studies³¹. Finally, it is important to note that while MR approaches can suggest potential causal relationships, additional evidence is required to confirm causal mechanisms. HPV-positive OPCs are considered to have a distinct etiopathogenesis compared to their HPV-negative counterparts, often with less pronounced associations with smoking and alcohol use. In our study, the apparent lack of disparity in the impact of risk factors could suggest a more nuanced and complex interaction between HPV status and these carcinogens than previously understood. Mechanistic studies that explore the biological interactions between HPV oncogenes and carcinogen-induced DNA damage in epithelial cells could provide further clarity.

In conclusion, we demonstrate that smoking and alcohol consumption have independent causal effects on the risk of both HPV-positive HNSCC and HPV-negative HNSCC. Using a multivariable MR approach, we show that the influence of lifetime smoking is similarly associated with both cancer types. Furthermore, we observed statistically significant results linking increased alcohol consumption in both HPV-positive and negative HNSCC. These results shed new light on possible modifying risk factors for HPV-positive HNSCC.

Methods

The study protocol was approved by the Voyager Consortium, with research consent obtained by the institutional review boards or ethics committees of each participating institution within the consortium. All participants, including cases and controls, provided written informed consent. The complete list of collaborating studies and their respective institutions can be found at <https://voyager.iarc.who.int/co-investigators/>.

The study design was an MR analysis of smoking and alcohol exposures on the risk of HNSCC stratified by HPV status.

The study population consisted of individuals included in the VOYAGER (Human

Papillomavirus, Oral and Oropharyngeal Cancer Genomic Research) consortium³⁷. Within VOYAGER, OncoArray data were available from a total of 3431 cases and 3469 controls from Europe and North America²². In brief, all VOYAGER studies are hospital- or population-based case-control studies, except for the UK's Head and Neck 5000 (HN 5000) case series. Individual studies obtained informed consent from all participants and ethical approval from their respective Institutional Review Boards. All studies utilized standardized instruments to collect information on sociodemographic and clinical characteristics, including information on smoking and alcohol-related behaviors.

HNSCC cases comprised the following International Classification of Disease Volume 10 (ICD-10): oral cavity (C02.0-C02.9, C03.0-C03.9, C04.0-C04.9, C05.0-C06.) oropharynx (C01.9, C02.4, C09.0-C10.9), hypopharynx (C13.0-C13.9) and overlapping (C14 and combination of other sites). Further stratification based on HPV status was performed to evaluate differences between HPV-positive and HPV-negative cancers. HPV-positive cancers were defined as OPC patients with positive HPV16 antibody status as the primary classifier, given that up to 90% of HPV-positive OPCs are attributed to HPV type 16 (HPV16)³⁸. OPC cases were classified as HPV-positive or HPV-negative based on a previously validated and described HPV16 seropattern algorithm³⁹, acknowledging that this method may lead to an underrepresentation of seropositivity for other high-risk HPV subtypes such as HPV 18, 31, and 33. For cases where the HPV16 antibody status was indeterminable, we utilized the expression of p16 as a surrogate marker. p16 is a cellular protein whose overexpression is an indirect measure of HPV-associated oncogenic activity, rather than a direct viral marker. OPC patients with unknown HPV status were excluded from analyses ($n = 102$). HPV-negative OPCs were pooled together with oral cavity cancer (OCC) cases as HPV-negative cancers. With an estimated prevalence of 5% or less in OCC, HPV is considered to have a limited role in the development of carcinomas of the oral cavity^{40–42}. Consequently, all OCCs in our study population were assumed to be HPV-negative. Since OCC shares similar risk factors of excessive smoking and alcohol consumption with HPV-negative OPC, we hypothesized that these tumor types have a similar etiology. After excluding people with primary tumor sites other than the oral cavity or oropharynx ($n = 299$) and unknown HPV status (among the OPC subgroup; $n = 102$), the final cohort consisted of 1105 patients classified as HPV-positive HNSCC and 2326 patients as HPV-negative HNSCC (Supplementary Table S1).

Genotyping, genetic data acquisition, quality control, and imputation

Individual-level genetic data were obtained from the VOYAGER consortium, with genotyping performed using the Illumina OncoArray⁴³ as described previously²². Genotyping data were accessed through the database of genotypes and phenotypes (dbGaP) project number phs001202.v1.p1⁴⁴. Standard quality control for the genotyping array included strand correction following standard pipelines⁴⁵, sex checking, missing rates, duplicates or relatedness, outlying heterozygosity rates, and population stratification. 486,987 SNPs were included after applying standard quality control procedures. Analyses were performed using PLINK v1.90b4.4⁴⁶ and EIGENSTRAT v6.1.4^{47,48}. Imputation was performed via the TOPMed Imputation Server⁴⁹ using software Eagle v2.4⁵⁰ and Minimac4⁵¹ with the TOPMed r^2 as a reference panel. Post-imputation quality control removed variants with r^2 values less than 0.3 and minor allele frequency less than 0.01.

For the exposures, we used the summary statistics and definitions from the GSCAN meta-Genome-Wide Association Study (meta-GWAS) of smoking and alcohol use behaviors conducted using data from 1.2 million individuals²⁰. Specifically, we defined: Smoking Initiation (SI) as a dichotomous variable of never smoker *versus* ever smoker, the latter defined as having smoked more than 100 cigarettes during lifetime; Alcohol Use Intensity/Drinks per Week (DPW) as a continuous variable of the average number of standardized alcoholic DPW; Comprehensive Smoking Index (CSI) as an independent and comprehensive indicator of smoking^{52,53}. We also investigated additional smoking behaviors from the GSCAN GWAS including: Age of Smoking Initiation (ASI) as a continuous variable of age at which participant started smoking cigarettes regularly, with regularly defined as >5 cigarettes/week; Smoking Intensity/Cigarettes per Day (CPD) as a continuous variable of the average number of cigarettes smoked per day; and Smoking Cessation (SC) as a dichotomous variable of former smoker *versus* current smoker. These GSCAN phenotypes/behaviors have been shown to be heritable and having a sufficient variation in population samples by the

Tobacco and Genetics (TAG) consortium³⁵. Further, these phenotypes have been shown to be reliable and valid measures of tobacco and alcohol use in terms of morbidity and mortality²⁴.

For multivariable MR, we used DPW and CSI³⁴. CSI was used in lieu of the four GSCAN smoking behaviors, which are correlated and interdependent and would thus be unsuitable for a multivariable model. A GWAS for the CSI variable conducted on the UK biobank³⁶ was used to obtain the instruments for this variable. We further investigated potential interaction effects between smoking and drinking on the risk of both HPV-positive and negative HNSCC using factorial MR.

Outcome

We conducted GWAS of risk of HNSCC stratified by HPV status, with the comparison groups consisting of HPV-positive HNSCC *versus* controls, and HPV-negative HNSCC *versus* controls³⁴.

GWAS of HNSCC risk stratified by HPV status were conducted using additive models. The log-odds of the outcome were regressed on the genetic variable with age, sex, and the first 7 genetic principal components (PC) as covariates. Population stratification was evaluated to determine ancestry using principal component analysis (PCA) with PC plots provided in Supplementary Fig. 1. The ethnicity of the population is described in Supplementary Table 1 based on self-reported information. The genetic association tests were performed using PLINK v1.90b4.4⁴⁶.

Mendelian randomization

MR is a statistical method that combines quantified estimates from associations between genetic instruments and the risk factor (smoking and alcohol use) with parallel associations between these genetic variants and the outcome, to determine an estimate of the risk factor's impact on the outcome (HNSCC). All steps of the **univariable MR** were performed using the *TwoSampleMR* package v0.5.6 in the R statistical language⁵⁵. The genetic instruments associated with smoking and alcohol use were selected based on the *P* values of association. Using a reference population of 1000 genomes' European superpopulation, all SNPs with *p* values $< 5 \times 10^{-8}$ were selected as potential index SNPs and then pruned using a clumping window of 10,000 base pairs (bp), with an r^2 cutoff of 0.001 to ensure independence. Secondary SNPs in LD were removed at a threshold level set at *P* values < 1 . The proxies for SNPs missing from the outcome GWAS were generated using the LD proxy tool (European superpopulation reference) and r^2 cutoff of 0.8; SNPs with the highest r^2 were selected. The "harmonise_data" function from the *TwoSampleMR* package was used to harmonize the SNPs between exposures and outcomes. The default action to infer the positive strand alleles using allele frequencies for palindromes was used. MR analyses with the Inverse-Variance-Weighted (IVW) method, MR Egger⁵⁶, MR weighted median⁵⁷(3), MR weighted mode, and MR-PRESSO⁵⁸ were performed for each of the behavior outcomes with each of the three outcomes (HPV-positive HNSCC patients *versus* controls, HPV-negative *versus* controls, and HPV-positive *versus* HPV-negative). The application of multiple MR methods allows the assessment of causal effects across various statistical assumptions, thereby adjusting for potential pleiotropy and invalid instruments. For the first four methods, the "mr()" function from the *TwoSampleMR* package with default parameters (z test distribution, alpha of 0.05, q threshold of 0.05, phi parameter of 1, Huber loss function, Cov parameter of 0, penk parameter of 20, over-dispersion, no shrinkage, and 1000 bootstraps) was used. For the MR-PRESSO method, R package *MR-PRESSO* v1.0⁵⁸ was used. The No Measurement Error (NOME) assumption was assessed using the I^2 statistic. These results are provided in the Supplementary Materials (Supplementary Table S5, 6).

Pleiotropy arises when a genetic variant is linked to the outcome of interest through multiple pathways, which may not necessarily involve the exposure under investigation. The presence of pleiotropy can alter both the magnitude and direction of the association between the exposure and outcome. To evaluate whether the assumptions of MR hold true, multiple MR methodologies are employed to assess the consistency of findings across different approaches. Pleiotropy was assessed using *MR-PRESSO*, and directional pleiotropy was assessed using the intercept for MR Egger regression (Supplementary Table 7).

For multivariable (MV) MR, Inverse-Variance-Weighted (IVR) Egger MVMR regression^{59,60}, and Q-statistic minimization approach were performed using the *MVMR* R package v0.4. MVMR Ridge regression, a method that uses Ridge regression to shrink the regression estimates was also used⁶¹. As previously mentioned, we used the GSCAN drinking behavior, DPW, with a comprehensive measure of lifetime smoking exposure, CSI. The summary statistics for this index have previously been derived from the UK Biobank⁵³. Specifically, the covariances for pairwise associations between SNP-exposure effects were assumed to be zero as the summary statistics for each exposure were derived from independent, non-overlapping samples. Weak instruments were tested using the conditional F statistic with a threshold of 10. Horizontal pleiotropy was tested using a modified form of the Q-statistic with respect to differences in MVMR estimates across the set of instruments⁶². Causal effect estimation was performed using the IVW method wherever the assumptions of MVMR were met (strong instruments and no significant pleiotropy). Whenever assumptions were violated, we obtained more robust estimates through Q-statistic minimization, which is particularly effective when instruments are weak or exhibit pleiotropy. For the MVMR ridge regression, a sequence of penalties (lambdas) was used and the results for the best lambdas were compared with the MVMR results. MVMR Egger regression was performed using the *MendelianRandomization* R package v0.70.

For factorial MR, we first constructed the polygenic risk scores (PRS) for the exposures as instrument variables for each study participant³¹. The PRS aggregates the effects of multiple genetic variants to estimate an individual's genetic predisposition to the specific exposure in question. The SNPs used to construct the PRS, along with their effect sizes, were the same as those used in the univariable MR. Then, we performed a one-sample MR of each phenotype on both HPV-positive and HPV-negative HNSCC. Subsequently, we performed the instrumental variables regression of PRS of each of the four smoking phenotypes with DPW used as interaction term (smoking phenotype multiplied with DPW) on both HPV-positive and HPV-negative HNSCC based on two-stage least-squares model (*AER* R package v1.2-10)⁶³. Since we were not able to build the PRS for CSI, we did not investigate the interactive effect between CSI and drinking in their effect on the two cancers.

We evaluated the association of risk tolerance and high-risk sexual behaviors with each cancer type as several of the genetic loci used in our study have been previously associated with these two exposures²⁴. Further, sexual behaviors are a purported risk factor for HPV infection^{24,35}, and risk tolerance can predispose an individual to initiate high-risk behaviors like smoking, drinking, and unsafe sexual activity. The association of risk tolerance and high-risk sexual behaviors with the two cancers was assessed using the univariable MR approach with the same methodology as for the smoking and drinking behaviors. Summary statistics for these behaviors were obtained from a large meta-GWAS published recently²⁴.

Statistics and reproducibility. AT and TH performed all MR analyses independently, with replication of the same results and conclusions. GWAS data used in this study had been previously replicated in the

respective studies^{20,22,24}. Additional information on research design is available in the Nature Research Reporting Summary.

Data availability

VOYAGER GWAS data access was approved under the application, #24972: “Genetic Influence of Smoking in Head and Neck Cancer”. Summary-level analyses were performed using publicly available GWAS data. Full summary statistics for VOYAGER GWAS can be accessed via dbGaP (OncoArray: Oral and Pharynx Cancer; study accession number: phs001202.v1.p1), and published data from this study can be accessed at Lesseur, C. et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat. Genet.* **48**, 1544–1550 (2016)²². Summary statistics for smoking initiation and alcohol use (GSCAN GWAS) were downloaded online (<https://genome.psych.umn.edu/index.php/GSCAN>) and published in Liu, M.Z. et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237 (2019)²⁰. Risk tolerance/Sexual behaviors meta-GWAS (<http://www.thessgac.org/data>) is published in Karlsson L.R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **2019**;51(2):245–257²⁴. Data for CSI is published in Wootton, R.E. et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomization study. *Psychol. Med.* **1–9** (2019)⁵³. For access to VOYAGER individual-level data, an approved project proposal is required. The VOYAGER consortium welcomes applications from researchers to maximize the utility of the VOYAGER data for research on head and neck cancer and other cancer types. To request data, please go to: <https://voyager.iarc.who.int/data-access/>. Potential applicants are strongly encouraged to contact VOYAGER investigators before submitting their proposal, for assistance with completing the document and to answer any questions that might arise about research that uses this resource. Please contact: headspace@iarc.who.int. Project proposals will be discussed every three months for approval. Source data are provided with this paper.

Code availability

All code used in the present study is available at: <https://github.com/mrhnscc/MRHNSCC>⁶⁴.

References

- Bray, F. et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
- Jethwa, A. R. & Khariwala, S. S. Tobacco-related carcinogenesis in head and neck cancer. *Cancer Metastasis Rev.* **36**, 411–423 (2017).
- Boffetta, P., Hecht, S., Gray, N., Gupta, P. & Straif, K. Smokeless tobacco and cancer. *lancet Oncol.* **9**, 667–675 (2008).
- Gandini, S. et al. Tobacco smoking and cancer: a meta-analysis. *Int. J. Cancer* **122**, 155–164 (2008).
- Hashibe, M. et al. Evidence for an important role of alcohol-and aldehyde-metabolizing genes in cancers of the upper aerodigestive tract. *Cancer Epidemiol. Biomark. Prev.* **15**, 696–703 (2006).
- Hashibe, M. et al. Alcohol drinking in never users of tobacco, cigarette smoking in never drinkers, and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *J. Natl. Cancer Inst.* **99**, 777–789 (2007).
- Ang, K. K. et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N. Engl. J. Med.* **363**, 24–35 (2010).
- Elrefaey, S., Massaro, M., Chiocca, S., Chiesa, F. & Ansarin, M. HPV in oropharyngeal cancer: the basics to know in clinical practice. *Acta Otorhinolaryngol. Italica* **34**, 299 (2014).
- Fakhry, C. et al. Improved survival of patients with human papillomavirus-positive head and neck squamous cell carcinoma in a prospective clinical trial. *J. Natl. Cancer Inst.* **100**, 261–269 (2008).
- Kreimer, A. R. et al. Evaluation of human papillomavirus antibodies and risk of subsequent head and neck cancer. *J. Clin. Oncol.* **31**, 2708 (2013).
- Lang Kuhs, K. A. et al. Human papillomavirus 16 E6 antibodies in individuals without diagnosed cancer: a pooled analysis. *Cancer Epidemiol. Biomark. Prev.* **24**, 683–689 (2015).
- Lydiatt, W. M. et al. Head and Neck cancers-major changes in the American Joint Committee on cancer eighth edition cancer staging manual. *CA Cancer J. Clin.* **67**, 122–137 (2017).
- Anantharaman, D. et al. Combined effects of smoking and HPV16 in oropharyngeal cancer. *Int. J. Epidemiol.* **45**, 752–761 (2016).
- Skoulakis, A. et al. Do smoking and human papilloma virus have a synergistic role in the development of head and neck cancer? A systematic review and meta-analysis. *J. BUON* **25**, 1107–1115 (2020).
- Arif, R. T., Mogaddam, M. A., Merdad, L. A. & Farsi, N. J. Does human papillomavirus modify the risk of oropharyngeal cancer related to smoking and alcohol drinking? A systematic review and meta-analysis. *Laryngoscope Investig. Otolaryngol.* **7**, 1391–1401 (2022).
- Sanderson, E. et al. Mendelian randomization. *Nat. Rev. Methods Prim.* **2**, 6 (2022).
- de Leeuw, C., Savage, J., Bucur, I. G., Heskes, T. & Posthuma, D. Understanding the assumptions underlying Mendelian randomization. *Eur. J. Hum. Genet.* **30**, 653–660 (2022).
- Sanderson, E. Multivariable Mendelian randomization and mediation. *Cold Spring Harb. Perspect. Med.* **11**, a038984 (2021).
- Auguste, A. et al. Joint effect of tobacco, alcohol, and oral HPV infection on head and neck cancer risk in the French West Indies. *Cancer Med.* **9**, 6854–6863 (2020).
- Liu, M. et al. Data related to association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat. Genet.* **51**, 237–244 (2019).
- Furberg, H. et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
- Lesseur, C. et al. Genome-wide association analyses identify new susceptibility loci for oral cavity and pharyngeal cancer. *Nat. Genet.* **48**, 1544–1550 (2016).
- Gormley, M. et al. A multivariable Mendelian randomization analysis investigating smoking and alcohol consumption in oral and oropharyngeal cancer. *Nat. Commun.* **11**, 6071 (2020).
- Karlsson Linnér, R. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat. Genet.* **51**, 245–257 (2019).
- Cogliano, V. J. et al. Preventable exposures associated with human cancers. *J. Natl. Cancer Inst.* **103**, 1827–1839 (2011).
- Hashibe, M. et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol. Biomark. Prev.* **18**, 541–550 (2009).
- Vineis, P. et al. Tobacco and cancer: recent epidemiological evidence. *J. Natl. Cancer Inst.* **96**, 99–106 (2004).
- Sturgis, E. M. & Ang, K. K. The epidemic of HPV-associated oropharyngeal cancer is here: is it time to change our treatment paradigms? *J. Natl. Compr. Canc. Netw.* **9**, 665–673 (2011).
- Eldridge, R. C. et al. Smoking and subsequent human papilloma-virus infection: a mediation analysis. *Ann. Epidemiol.* **27**, 724–730.e721 (2017).
- Sinha, P., Logan, H. L. & Mendenhall, W. M. Human papillomavirus, smoking, and head and neck cancer. *Am. J. Otolaryngol.* **33**, 130–136 (2012).
- Rees, J. M., Foley, C. N. & Burgess, S. Factorial Mendelian randomization: using genetic variants to assess interactions. *Int. J. Epidemiol.* **49**, 1147–1158 (2020).
- Ference, B. A., Majeed, F., Penumetcha, R., Flack, J. M. & Brook, R. D. Effect of naturally random allocation to lower low-density lipoprotein

- cholesterol on the risk of coronary heart disease mediated by polymorphisms in NPC1L1, HMGCR, or both: a 2×2 factorial Mendelian randomization study. *J. Am. Coll. Cardiol.* **65**, 1552–1561 (2015).
33. Ference, B. A. et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N. Engl. J. Med.* **375**, 2144–2153 (2016).
 34. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
 35. Heck, J. E. et al. Sexual behaviours and the risk of head and neck cancers: a pooled analysis in the International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Int. J. Epidemiol.* **39**, 166–181 (2010).
 36. Gormley, M. et al. Investigating the effect of sexual behaviour on oropharyngeal cancer risk: a methodological assessment of Mendelian randomization. *BMC Med.* **20**, 40 (2022).
 37. Budhathoki, S. et al. A risk prediction model for head and neck cancers incorporating lifestyle factors, HPV serology and genetic markers. *Int. J. Cancer* **152**, 2069–2080 (2023).
 38. Castellsagué, X. et al. HPV involvement in head and neck cancers: comprehensive assessment of biomarkers in 3680 patients. *J. Natl Cancer Inst.* **108**, djv403 (2016).
 39. Ferreiro-Iglesias, A. et al. Germline determinants of humoral immune response to HPV-16 protect against oropharyngeal cancer. *Nat. Commun.* **12**, 5945 (2021).
 40. Chi, A. C., Day, T. A. & Neville, B. W. Oral cavity and oropharyngeal squamous cell carcinoma—an update. *CA Cancer J. Clin.* **65**, 401–421 (2015).
 41. Combes, J.-D. & Franceschi, S. Role of human papillomavirus in non-oropharyngeal head and neck cancers. *Oral. Oncol.* **50**, 370–379 (2014).
 42. Mirghani, H., Amen, F., Moreau, F. & Lacau St Guily, J. Do high-risk human papillomaviruses cause oral cavity squamous cell carcinoma? *Oral. Oncol.* **51**, 229–236 (2015).
 43. Consortium, O. Consortium launches genotyping effort. *Cancer Discov.* **3**, 1321–1322 (2013).
 44. Mailman, M. D. et al. The NCBI dbGaP database of genotypes and phenotypes. *Nat. Genet.* **39**, 1181–1186 (2007).
 45. Wrayner, N. (Accessed 09/20/2021). Strand. Wellcome Centre for Human Genetics. <https://www.well.ox.ac.uk/~wrayner/strand/>.
 46. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
 47. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLOS Genet.* **2**, e190 (2006).
 48. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
 49. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
 50. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
 51. Fuchsberger, C., Abecasis, G. R. & Hinds, D. A. minimac2: faster genotype imputation. *Bioinformatics* **31**, 782–784 (2015).
 52. Leffondré, K., Abrahamowicz, M., Xiao, Y. & Siemiatycki, J. Modeling smoking history using a comprehensive smoking index: application to lung cancer. *Stat. Med.* **25**, 4132–4146 (2006).
 53. Wootton, R. E. et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *Psychol. Med.* **50**, 2435 (2020).
 54. Prentice, R. L., Vollmer, W. M. & Kalbfleisch, J. D. On the use of case series to identify disease risk factors. *Biometrics* **4**, 445–458 (1984).
 55. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenotype. *Elife* **7**, e34408 (2018).
 56. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
 57. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
 58. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).
 59. Rees, J. M. B., Wood, A. M. & Burgess, S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. *Stat. Med.* **36**, 4705–4718 (2017).
 60. Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**, 251–260 (2015).
 61. Mariosa, D. et al. Body size at different ages and risk of 6 cancers: a Mendelian randomization and prospective cohort study. *J. Natl Cancer Inst.* **114**, 1296–1300 (2022).
 62. Sanderson, E., Spiller, W. & Bowden, J. Testing and correcting for weak and pleiotropic instruments in two-sample multivariable Mendelian randomization. *Stat. Med.* **40**, 5434–5452 (2021).
 63. Kleiber, C. & Zeileis, A. *Applied econometrics with R*. (Springer Science & Business Media, 2008).
 64. Thrakal, A. et al. mrhnscc/MRHNSCC: v1 (Version v1). Zenodo. <https://doi.org/10.5281/zenodo.12558118> (2024).

Acknowledgements

This study was funded in part by NIH/NIDCR R01 DE025712 (to P.B., B.D., and D.N.H.). Genotyping of the HNSCC cases and controls was performed at the Center for Inherited Disease Research (CIDR) and funded by NIH/NIDCR 1X01HG007780-O. The Alcohol-Related Cancers and Genetic Susceptibility Study in Europe (ARCAGE) was funded by the European Commission's fifth framework program (QLK1-2001-00182), the Italian Association for Cancer Research, Compagnia di San Paolo/FIRMS, Region Piemonte and Padova University (CPDA057222). We thank Dr. Wolfgang Ahrens, PhD (Universität Bremen, Germany) for his support in ARCAGE study. The Carolina Head and Neck Cancer Epidemiology (CHANCE) study was supported in part by the National Cancer Institute (R01-CA90731). The Head and Neck 5000 study was a component of independent research funded by the National Institute for Health Research (NIHR) under its Program Grants for Applied Research scheme (RP-PG-0707-10034). The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. Core funding was also provided through awards from Above and Beyond, University Hospitals Bristol and Weston Research Capability Funding, and the NIHR Senior Investigator award to Professor Andy Ness. Human papillomavirus (HPV) serology was supported by a Cancer Research UK Program Grant, the Integrative Cancer Epidemiology Program (grant number: C18281/A19169). The University of Pittsburgh head and neck cancer case-control study is supported by US National Institutes of Health grants P50CA097190 and P30CA047904. The MSH-PMH study is supported by the Canadian Cancer Society Research Institute, and CIHR Canada Research Chair to R.J.H. J.P. was partially supported by the Italian Ministry of Health "Ricerca Corrente". G.L. is funded by the Alan B. Brown Chair in Molecular Genomics and the Lusi Wong Foundation Fund. This study was supported by the Princess Margaret Cancer Center Head & Neck Translational Program, with philanthropic funds from the Wharton Family, Joe's Team, and Gordon Tozer. Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation,

the Government of Ontario, the Ontario Research Fund—Research Excellence, and the University of Toronto.

Author contributions

A.T., J.L., and O.E.G. conceived the study, and all three carried out data curation, analysis, interpretation of results, and manuscript writing. Clinical and genetic epidemiological data and methodology expertise were provided by T.H., K.H., T.D., M.G., M.B., S.H., S.B., A.S., J.D.A., J.D., L.B., D.G., and R.H. Guidance, expertise, and specific interpretation of multivariable MR methodology and analyses were provided by T.D., M.G., E.S., and K.S.B. Head and neck cancer genetic data were obtained through multiple collaborations from studies led by A.O., S.V., B.D., A.N., T.W., P.B., D.H., G.M., P.L., A.L., J.P., A.A., L.A., W.A., C.H., D.C., M.N., C.C., I.H., L.R., and A.Z. All authors contributed to the interpretation of the results and critical revision of the manuscript. This work was jointly supervised by W.X., G.L., and O.E.G.

Competing interests

The authors declare no competing interests. The funders had no role in the design of the study, the collection, analysis, and interpretation of data, the writing of the manuscript, or the decision to submit the manuscript for publication. The research was conducted independently by the authors, who are responsible for all the findings expressed in this article. The authors affirm that they have no financial or personal relationships that could have influenced the work reported in this paper. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policies, or views of the International Agency for Research on Cancer/World Health Organization.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51679-x>.

Correspondence and requests for materials should be addressed to John J.W. Lee or Geoffrey Liu.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

¹Department of Biostatistics, Princess Margaret Cancer Centre-University Health Network, University of Toronto, Toronto, Ontario, Canada. ²Department of Otolaryngology-Head and Neck Surgery, University of Toronto, Toronto, Ontario, Canada. ³Department of Otolaryngology - Head and Neck Surgery, Sinai Health System, Toronto, Ontario, Canada. ⁴MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 1TL, UK. ⁵University of Bristol Dental School, 1 Trinity Walk, Avon Street, Bristol BS2 0PT, UK. ⁶Genomic Epidemiology Branch, International Agency for Research on Cancer, World Health Organization, Lyon, France. ⁷Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, USA. ⁸Department of Human Genetics, School of Public Health, University of Pittsburgh, and UPMC Hillman Cancer Center, Pittsburgh, PA 15260, USA. ⁹University Hospitals Bristol and Weston NHS Foundation Trust National Institute for Health Research Bristol Biomedical Research Centre, University of Bristol, Bristol BS1 3NU, UK. ¹⁰Infections and Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹¹Division of Medical Oncology and Center for Cancer Research, University of Tennessee Health Science Center, Memphis, TN, USA. ¹²Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. ¹³Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada. ¹⁴Division of Medical Oncology and Hematology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada. ¹⁵Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. ¹⁶Epidemiology Group, School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Aberdeen, UK. ¹⁷Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian, University of Athens, Goudi, Greece. ¹⁸Department of Public and Community Health, School of Public Health, University of West Attica, Athens, Greece. ¹⁹Unit of Cancer Epidemiology, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, Aviano, Italy. ²⁰Cancer Epidemiology Research Programme, Catalan Institute of Oncology, L'Hospitalet de Llobregat, Barcelona, Spain; Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública, Madrid, Spain. ²¹Department of Epidemiological Methods and Etiological Research, Leibniz Institute for Prevention Research and Epidemiology-BIPS, Bremen, Germany; Department of Mathematics and Computer Science, University of Bremen, Bremen, Germany. ²²Department of Oral and Maxillofacial Surgery, Oral Medicine and Oral Pathology, Dublin Dental University Hospital, Trinity College Dublin, Dublin, Ireland. ²³School of Medicine, Dentistry, and Nursing, University of Glasgow, Glasgow, UK. ²⁴Department of Research, Cancer Registry of Norway, Norwegian Institute of Public Health, Oslo, Norway. ²⁵University of Padua, Padua, Italy. ²⁶Institute of Hygiene & Epidemiology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic. ²⁷Department of Medical Sciences, University of Turin and CPO-Piemonte, Turin, Italy. ²⁸Cancer Surveillance Branch, International Agency for Research on Cancer, World Health Organization, Lyon, France. ²⁹Prosserman Centre for Population Health Research, Lunenfeld-Tanenbaum Research Institute, Sinai Health, Toronto, Canada. ³⁰Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. ³¹Department of Epidemiology and Biostatistics, University of Western Ontario, London, Ontario, Canada. ³²These authors contributed equally: Abhinav Thakral, John J.W. Lee. ³³These authors jointly supervised this work: Wei Xu, Geoffrey Liu, Osvaldo Espin-Garcia.

✉ e-mail: Jongwook.lee@utoronto.ca; Geoffrey.Liu@uhn.ca