A risk prediction model for head and neck cancers incorporating lifestyle factors, HPV serology and genetic markers

Sanjeev Budhathoki, Brenda Diergaarde, Geoffrey Liu, Andrew Olshan, Andrew Ness, Tim Waterboer, Shama Virani, Patricia Basta, Noemi Bender, Nicole Brenner, Tom Dudding, Neil Hayes, Andrew Hope, Shao Hui Huang, Katrina Hueniken, Beatriz Kanterewicz, James D McKay, Miranda Pring, Steve Thomas, Kathy Wisniewski, Sera Thomas, Yonathan Brhane, Antonio Agudo, Laia Alemany, Areti Lagiou, Luigi Barzan, Cristina Canova, David I. Conway, Claire M. Healy, Ivana Holcatova, Pagona Lagiou, Gary J. Macfarlane, Tatiana V. Macfarlane, Jerry Polesel, Lorenzo Richiardi, Max Robinson, Ariana Znaor, Paul Brennan and Rayjean J. Hung

Supplementary Information

Supplementary Methods

Estimation of absolute risk

The absolute risk of developing head and neck cancer for an adult of age *a* years within a duration of τ years (i.e. within an interval of [*a, a+* τ]) was determined by integrating the equation below:

$$AR(a, a+\tau) = \int_{a}^{a+\tau} \lambda_0(t)\exp(Z\beta)\, exp\left(-\int_{a}^{t}[\lambda_0(u)\exp(Z\beta) + m(u)]du\right)dt$$

where $\lambda_0(t)$ is the baseline hazard function, *Z* is a set of risk factors, β is a vector of log relative risk, $m(t)$ is age-specific competing hazards of mortality, and u is the time interval for the estimation of the integral. The derivation of the equation has been described in detail elsewhere[1, 2]. The underlying assumption of the risk model is that risk factors act in a multiplicative fashion on the baseline hazard function. Odds ratios, estimated from cases and controls in our study with adjustment of age and other risk factors, were used as a measure of relative risk. The age-specific cancer rates and competing hazards for mortality (Supplemental Table 3) were obtained from Surveillance, Epidemiology, and End Results (SEER) Program and Centers for Disease Control and Prevention, National Center for Health Statistics database respectively[3, 4].

Model calibration

We evaluated calibration of the risk models in the UK Biobank cohort, which is a population-based prospective cohort study of over 500 000 participants. The details of the study design have been described previously[5]. In brief, participants of age ranging from 38 years to 73 years were recruited between 2006 and 2010 at multiple assessment centres across the United Kingdom. At baseline, all participants underwent a self-completed questionnaire survey which inquired about lifestyle risk factors such as smoking and alcohol use, and medical history and family history of cancer. In addition, extensive physical measurement and biospecimens were also collected at baseline. The information on cancer diagnosis was obtained through record linkage with death and cancer registries. For this study, participants were followed to the date of death, cancer diagnosis, or censoring date of March 31, 2016 (in England and Wales) and Oct 31, 2015 (in Scotland). A total of 481,881 participants were available for analysis including 749 cases of head and neck cancer. Genotyping was performed using the UK BiLEVE Axiom array and the UK Biobank Axiom array[6]. Imputation was based on the Haplotype Reference Consortium reference panel. We computed PRS in the UK Biobank using the same weights as in the model development set. Three variants [rs201982221, HLA-B (156-Trp), HLA-DRB1 (71-Glu)] were not genotyped or imputed in the UK Biobank and were not included in the calculation of PRS. We imputed serostatus of the UK Biobank participants by random binomial draw with the overall probability of seropositivity (0.86%) estimated from controls who were assayed in VOYAGER study.

UK Biobank is known to be a healthier population with higher social economic status, lower smoking rate and lower cancer incidence[7]. To account for the population-level difference in the risk profile in UK Biobank, we applied the recalibration approach with the models reported, using a random sample of 50% of the UK Biobank, while keeping the remaining 50% for strict prospective assessment of calibration. Recalibration is a standard statistical approach when a developed risk model is being imported into a population that may have different risk profiles, while keeping the model structure unchanged[8, 9]. The method details of recalibration have been reported previously[9, 10]. For our study, we computed the log-odds of HNC cancers (Z) in UKB based on the same coefficients of models we developed using the VOYAGER data. Then we fit a logistic regression

model in the 50% training sample with HNC cancer status as the outcome and Z as the sole predictor. The beta coefficient for Z, $\hat{\beta}_Z$, is the re-calibrated slope (i.e. the adjustment factor). The adjustment factors are summarized in Supplementary Table 7. The reported calibration is based on the 50% hold-out testing set. All absolute risk estimation and calibration analyses were performed in R statistical software using *iCARE* package.

Reference

1. Gail MH. Estimation and interpretation of models of absolute risk from epidemiologic data, including family-based studies. *Lifetime Data Anal* 2008;**14**: 18-36.

2. Pal Choudhury P, Maas P, Wilcox A, Wheeler W, Brook M, Check D, Garcia-Closas M, Chatterjee N. iCARE: An R package to build, validate and apply absolute risk models. *PLoS One* 2020;**15**: e0228198.

3. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2020 Sub (1975-2018) - Linked To County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.

4. Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov/ucd-icd10.html on Jul 19, 2021 12:15:53 PM.

5. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, Liu B, Matthews P, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;**12**: e1001779.

6. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;**562**: 203-9.

7. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, Collins R, Allen NE. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol* 2017;**186**: 1026-34.

8. Field JK, Vulkan D, Davies MPA, Duffy SW, Gabe R. Liverpool Lung Project lung cancer risk stratification model: calibration and prospective validation. *Thorax* 2021;**76**: 161-8.

9. Puddu PE, Piras P, Kromhout D, Tolonen H, Kafatos A, Menotti A. Re-calibration of coronary risk prediction: an example of the Seven Countries Study. *Sci Rep* 2017;**7**: 17552.

10. Winter A, Aberle DR, Hsu W. External validation and recalibration of the Brock model to predict probability of cancer in pulmonary nodules using NLST data. *Thorax* 2019;**74**: 551-63.

Supplemental Table 1. Summary of the SNPs included in the calculation of polygenic risk score for head and neck cancer

	Variants	Region	Risk allele	Risk allele frequency	Odds ratio	Reference (PMID)
Oral cavity cancer						
	rs10462706	5p15.33	C	0.85	0.74	27749845
	rs1229984	4q23	G	0.94	0.57	27749845
	rs6547741	2p23.3	G	0.46	0.83	27749845
	rs8181047	9p21.3	A	0.24	1.24	27749845
	rs928674	9q34.12	G	0.12	1.33	27749845
Oropharyngeal cancer						
	rs1229984	4q23	G	0.94	0.55	27749845
	rs3828805	6p21.32	C	0.72	1.37	27749845
	rs4713462	6p21.3	A	0.326	0.71	34642315
	HLA-B*1501	6p21.3	P/A	0.059	0.79	34642315
	HLA-B (156-Trp)	6p21.3	P/A	0.061	0.80	34642315
	HLA-DRB1*1301	6p21.3	P/A	0.067	0.49	34642315
	HLA-DRB1 (71-Glu)	6p21.3	P/A	0.145	0.59	34642315
	HLA-DQA1*0103	6p21.3	P/A	0.078	0.53	34642315
	HLA-DQB1*0603	6p21.3	P/A	0.073	0.53	34642315
	rs35189640	12q23.3	T	0.02	1.66	34642315
Head and neck cancer†						
	rs1494961	4q21.23	C	0.49	1.12	21437268
	rs1789924	4q23	C	0.61	1.12	21437268
	rs4767364	12q24.13	A	0.30	1.13	21437268
	rs971074	4q23	G	0.88	0.75	21437268
	rs1229984	4q23	G	0.94	0.56	27749845
	rs1453414	11p15.4	C	0.2	1.19	27749845
	rs79767424	5p14.3	C	0.97	0.55	27749845
	rs2299187	7q21.11	A	0.02	3.26	27173062
	rs201982221	10q26	D/I	0.02	1.74	34642315
	rs35189640	12q23.3	T	0.02	1.79	34642315

D/I, deletion/insertion; P/A, presence/absence for amino acid polymorphisms in HLA alleles

†Including SNPs for oral cavity cancer and oropharyngeal cancer

Supplemental Table 2. Age-specific incidence rates of head and neck cancer and all-other-cause mortality rates per 100 000 person-years in non-Hispanic White population in the United States[a]

Age	Head and neck cancer		Oral cavity cancer		Oropharyngeal cancer	
	Incidence	All-other-cause Mortality	Incidence	All-other-cause Mortality	Incidence	All-other-cause Mortality
Men						
40-44	11.0	263.7	2.5	264.8	3.1	264.7
45-49	23.7	394.0	5.0	396.9	8.1	396.5
50-54	42.2	591.0	7.9	597.2	14.1	596.3
55-59	64.8	871.5	12.0	882.3	20.4	880.8
60-64	85.8	1278.4	15.5	1293.9	24.4	1292.2
65-69	101.5	1875.0	18.1	1894.6	24.8	1892.8
70-74	111.5	2906.1	20.0	2929.0	23.9	2927.4
Women						
40-44	4.2	154.6	1.1	155.0	0.8	155.0
45-49	8.2	235.0	2.3	235.8	1.9	235.8
50-54	13.6	352.4	3.7	354.0	3.4	353.9
55-59	20.6	524.8	5.7	527.5	5.2	527.3
60-64	26.7	792.8	7.5	796.6	6.6	796.4
65-69	33.6	1218.5	9.5	1223.7	7.5	1223.5
70-74	36.8	1968.4	11.4	1975.4	7.8	1975.4

[a]Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER Research Data, 9 Registries, Nov 2020 Sub (1975-2018) - Linked To County Attributes - Time Dependent (1990-2018) Income/Rurality, 1969-2019 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2021, based on the November 2020 submission.

Supplemental Table 3. Distribution of the selected characteristics in cancer cases

Variables	Categories	Hypopharynx cancer	Larynx cancer	Other cancer*
Total (n)		518	2379	1077
Sex, n (%)				
	Men	438 (84.7)	2041 (85.8)	728 (67.7)
	Women	79 (15.3)	337 (14.2)	348 (32.3)
	Missing	1	1	1
Age (years), mean (SD)		61.4 (9.9)	63.4 (10.6)	58.7 (12.7)
Tobacco Smoking status, n (%)				
	Never	27 (6.2)	133 (6.4)	252 (26.4)
	Former	175 (39.9)	943 (45.6)	301 (31.5)
	Current	237 (54.0)	991 (47.9)	402 (42.1)
	Missing	79	312	122
Tobacco Pack-years, median (IQR)		40 (31.5)	42 (35.6)	36 (32.6)
Alcohol drinking status, n (%)				
	Never	53 (12.4)	396 (19.3)	201 (21.0)
	Former	217 (50.8)	958 (46.7)	371 (38.8)
	Current	157 (36.8)	699 (34.0)	385 (40.2)
	Missing	91	326	120
Drink/week, median (IQR)		28 (38.3)	21 (28.8)	14.7 (28.6)
Education, n (%)				
	Postsecondary	76 (20.9)	424 (23.6)	311 (34.0)
	High school diploma	88 (24.2)	436 (24.3)	275 (30.1)
	None/elementary	199 (54.8)	937 (52.1)	329 (36.0)
	Missing	155	582	162

*includes cancers of the salivary gland (C07.9-C08.9), nasopharynx (C11.0-C11.9) and oral cavity-oropharynx-hypopharynx not otherwise specified (C02.8, C02.9, C05.8, C05.9, C14.0, C14.2, C14.8).

Supplemental Table 4. Cut-off of smoking, drinking and polygenic risk score for head and neck cancers

	Variables	Categories	Head and neck cancer	Oral cavity cancer	Oropharyngeal cancer
Men					
	Smoking status[a]	Moderate	<24 pack-years		
		Heavy	≥24 pack-years		
	Drinking status[b]	Never/low	<5.5 drinks/week		
		Moderate	5.5 – <14.7 drinks/week		
		Heavy	≥14.7 drinks/week		
	Polygenic risk score[c]	1st tertile	≤ -0.08	≤ -0.27	≤ -0.43
		2nd tertile	> -0.08, ≤ 0.48	> -0.27, ≤ 0.0004	> -0.43, ≤ 0.21
		3rd tertile	> 0.48	> 0.0004	> 0.21
		Median (Q1, Q3)	0.23 (-0.28, 0.65)	-0.16 (-0.37, 0.03)	0.03 (-0.92, 0.31)
Women					
	Smoking status[a]	Moderate	<14 pack-years		
		Heavy	≥14 pack-years		
	Drinking status[b]	Never/low	<2.2 drinks/week		
		Moderate	2.2 – <6.9 drinks/week		
		Heavy	≥6.9 drinks/week		
	Polygenic risk score[c]	1st tertile	≤ -0.05	≤ -0.27	≤ -0.38
		2nd tertile	> -0.05, ≤ 0.54	> -0.27, ≤ 0.007	> -0.38, ≤ 0.27
		3rd tertile	> 0.54	> 0.007	> 0.27
		Median (Q1, Q3)	0.26 (-0.23, 0.68)	-0.16 (-0.37, 0.03)	0.06 (-0.78, 0.31)

[a]The cut-off is based on sex-specific medians among ever smokers in the control group

[b]The cut-off is based on sex-specific tertiles in the control group

[c]The polygenic risk scores are computed for oral cavity and oropharyngeal cancer separately based on the loci reported for these tumor types. Loci reported for head and neck cancer or their anatomical subsites are included in the PRS for head and neck cancer overall. The cut-off is based on sex-specific tertiles in the control group

Supplemental Table 5a. Beta coefficients of risk factors in different models of head and neck cancer overall and oral cavity cancer

	Men				Women			
	Epi model		Epi & PRS		Epi model		Epi & PRS	
	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
Head and neck cancer								
Age, < 50 years	0.44	<0.01	0.45	<0.01	0.59	0.03	0.62	0.02
50 - < 55 years	0.09	0.52	0.08	0.54	0.04	0.85	0.04	0.83
55 - < 60 years	0.22	0.11	0.19	0.16	-0.11	0.59	-0.11	0.57
60 - < 65 years	0.18	0.18	0.18	0.18	-0.08	0.70	-0.07	0.71
65 - < 70 years	-0.06	0.67	-0.07	0.63	-0.23	0.27	-0.24	0.24
70 - < 75 years	-0.06	0.72	-0.07	0.66	-0.19	0.38	-0.18	0.42
≥ 75 years	0.01	0.95	0.02	0.92	0.11	0.61	0.11	0.58
Smoking [a], Moderate	0.19	0.03	0.20	0.02	0.28	0.04	0.27	0.05
Heavy	0.91	<0.01	0.93	<0.01	1.24	<0.01	1.24	<0.01
Drinking [b], Moderate	-0.09	0.31	-0.14	0.10	-0.35	0.01	-0.36	0.01
Heavy	0.53	<0.01	0.49	<0.01	0.42	<0.01	0.41	<0.01
Education, High school	0.83	<0.01	0.81	<0.01	1.17	<0.01	1.15	<0.01
None/elementary	0.68	<0.01	0.67	<0.01	1.60	<0.01	1.58	<0.01
PRS category, 2nd tertile			0.42	<0.01			0.55	<0.01
3nd tertile			0.86	<0.01			0.66	<0.01
Oral cavity cancer								
Age, < 50 years	0.31	0.18	0.34	0.15	0.03	0.94	0.03	0.93
50 - < 55 years	-0.10	0.64	-0.10	0.63	-0.22	0.43	-0.23	0.41
55 - < 60 years	0.02	0.91	0.04	0.85	-0.39	0.15	-0.41	0.13
60 - < 65 years	0.08	0.70	0.10	0.63	-0.15	0.57	-0.19	0.47
65 - < 70 years	0.00	0.98	0.03	0.88	-0.07	0.80	-0.12	0.64
70 - < 75 years	0.29	0.18	0.33	0.14	-0.04	0.88	-0.02	0.94
≥ 75 years	0.64	<0.01	0.64	<0.01	0.43	0.09	0.44	0.09
Smoking [a], Moderate	0.42	0.01	0.43	0.01	0.35	0.05	0.31	0.08
Heavy	1.18	<0.01	1.18	<0.01	1.22	<0.01	1.21	<0.01
Drinking [b], Moderate	-0.08	0.55	-0.12	0.35	-0.31	0.10	-0.31	0.11
Heavy	0.65	<0.01	0.62	<0.01	0.45	<0.01	0.43	0.01
Education, High school	0.99	<0.01	1.00	<0.01	1.34	<0.01	1.33	<0.01
None/elementary	0.97	<0.01	0.99	<0.01	1.87	<0.01	1.86	<0.01
PRS category[b], 2nd tertile			0.29	0.02			0.55	<0.01
3nd tertile			0.77	<0.01			0.76	<0.01

[a]The cut-off is based on sex-specific medians among ever smokers in the control group

[b]The cut-off is based on sex-specific tertiles in the control group

Supplemental Table 5b. Beta coefficients of risk factors in different models of oropharyngeal cancer

	Men						Women					
	Epi model		Epi & HPV		Epi, HPV & PRS		Epi model		Epi & HPV		Epi, HPV & PRS	
	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value	Estimate	P value
Age, < 50 years	0.74	<0.01	0.57	0.08	0.57	0.08	1.31	<0.01	1.57	0.01	1.60	0.01
50 - < 55 years	0.50	0.01	0.24	0.42	0.24	0.41	0.80	0.03	0.48	0.36	0.51	0.34
55 - < 60 years	0.67	<0.01	0.53	0.07	0.55	0.06	0.81	0.02	0.93	0.05	0.96	0.05
60 - < 65 years	0.53	<0.01	0.27	0.36	0.31	0.30	0.50	0.15	0.79	0.10	0.83	0.09
65 - < 70 years	0.34	0.08	0.49	0.10	0.49	0.10	0.18	0.63	0.40	0.44	0.44	0.40
70 - < 75 years	0.09	0.67	0.20	0.54	0.22	0.50	0.31	0.43	0.88	0.10	0.89	0.10
≥ 75 years	0.02	0.93	-0.03	0.95	-0.02	0.95	0.24	0.55	0.64	0.25	0.67	0.23
Smoking[a], Moderate	0.12	0.31	0.47	0.02	0.49	0.02	0.72	<0.01	1.02	0.01	1.04	<0.01
Heavy	0.75	<0.01	1.74	0.00	1.78	<0.01	1.26	<0.01	1.89	<0.01	1.89	<0.01
Drinking [b], Moderate	-0.10	0.38	-0.21	0.29	-0.24	0.22	-0.37	0.13	0.16	0.60	0.12	0.69
Heavy	0.49	<0.01	0.84	<0.01	0.80	<0.01	0.79	<0.01	1.06	0.00	1.05	<0.01
Education, High school	0.68	<0.01	0.57	<0.01	0.56	<0.01	0.69	<0.01	0.64	0.02	0.65	0.02
None/elementary	0.22	0.05	0.47	<0.01	0.50	<0.01	0.59	0.01	0.64	0.03	0.66	0.03
HPV seropositive			5.99	<0.01	5.96	<0.01			5.36	<0.01	5.32	<0.01
PRS category[b], 2nd tertile					0.11	0.51					-0.14	0.64
3nd tertile					0.50	<0.01					0.30	0.27

[a]The cut-off is based on sex-specific medians among ever smokers in the control group

[b]The cut-off is based on sex-specific tertiles in the control group

Supplemental Table 6. Odds ratios (ORs) and 95% confidence intervals (CIs) of head and neck cancer including and excluding HN5000 study

Variable	With HN5000		Without HN5000	
	OR (95% CI)	P value	OR (95% CI)	P value
Men				
Smoking status[a], Never	1 (Ref.)		1 (Ref.)	
Moderate	1.20 (1.01-1.43)	0.04	1.20 (1.00-1.45)	0.05
Heavy	2.58 (2.16-3.07)	<0.01	3.01 (2.51-3.62)	<0.01
Drinking status[b], Never/low	1 (Ref.)		1 (Ref.)	
Moderate	0.85 (0.72-1.01)	0.06	0.91 (0.76-1.09)	0.31
Heavy	1.57 (1.34-1.84)	<0.01	1.54 (1.30-1.84)	<0.01
Education, Postsecondary	1 (Ref.)		1 (Ref.)	
High school diploma	2.14 (1.83-2.52)	<0.01	2.55 (2.15-3.03)	<0.01
None/elementary	1.48 (1.20-1.82)	<0.01	3.24 (2.53-4.14)	<0.01
Polygenic risk score[b], 1st tertile	1 (Ref.)		1 (Ref.)	
2nd tertile	1.52 (1.29-1.79)	<0.01	1.40 (1.17-1.68)	<0.01
3rd tertile	2.34 (2.00-2.75)	<0.01	2.18 (1.83-2.58)	<0.01
Women				
Smoking status[a], Never	1 (Ref.)		1 (Ref.)	
Moderate	1.31 (1.00-1.70)	0.05	1.21 (0.91-1.60)	0.19
Heavy	3.67 (2.91-4.64)	<0.01	3.65 (2.86-4.65)	<0.01
Drinking status[b], Never/low	1 (Ref.)		1 (Ref.)	
Moderate	0.62 (0.47-0.82)	<0.01	0.81 (0.61-1.08)	0.15
Heavy	1.34 (1.05-1.70)	0.02	1.12 (0.86-1.46)	0.40
Education, Postsecondary	1 (Ref.)		1 (Ref.)	
High school diploma	2.75 (2.17-3.48)	<0.01	3.18 (2.49-4.05)	<0.01
None/elementary	2.31 (1.68-3.17)	<0.01	4.91 (3.34-7.22)	<0.01
Polygenic risk score[b], 1st tertile	1 (Ref.)		1 (Ref.)	
2nd tertile	1.60 (1.24-2.05)	<0.01	1.67 (1.28-2.18)	<0.01
3rd tertile	1.85 (1.45-2.37)	<0.01	1.78 (1.37-2.31)	<0.01

OR, odds ratio; CI, confidence interval

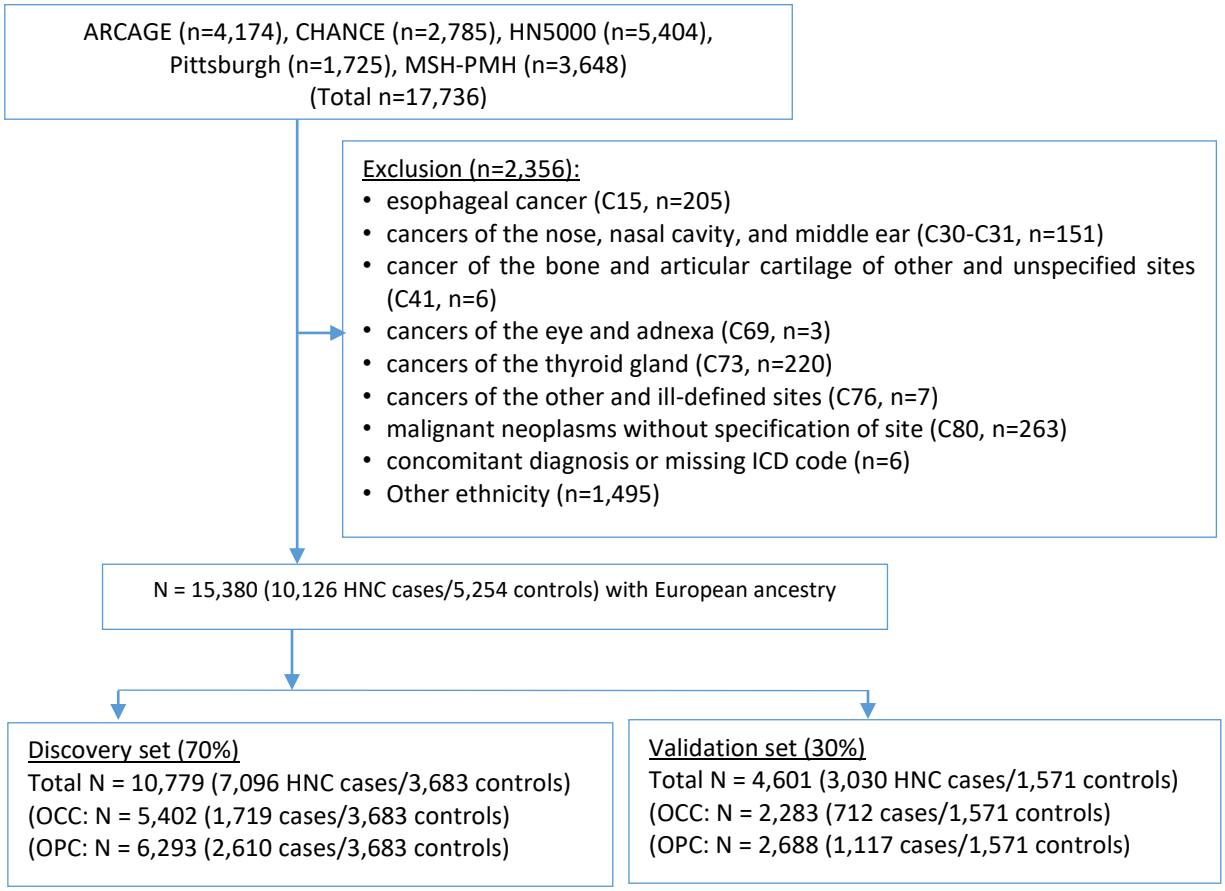
[a]The cut-off is based on sex-specific medians among ever smokers in the control group

[b]The cut-off is based on sex-specific tertiles in the control group

Supplemental Table 7. Adjustment factors ($\hat{\beta}_Z$) for UKB

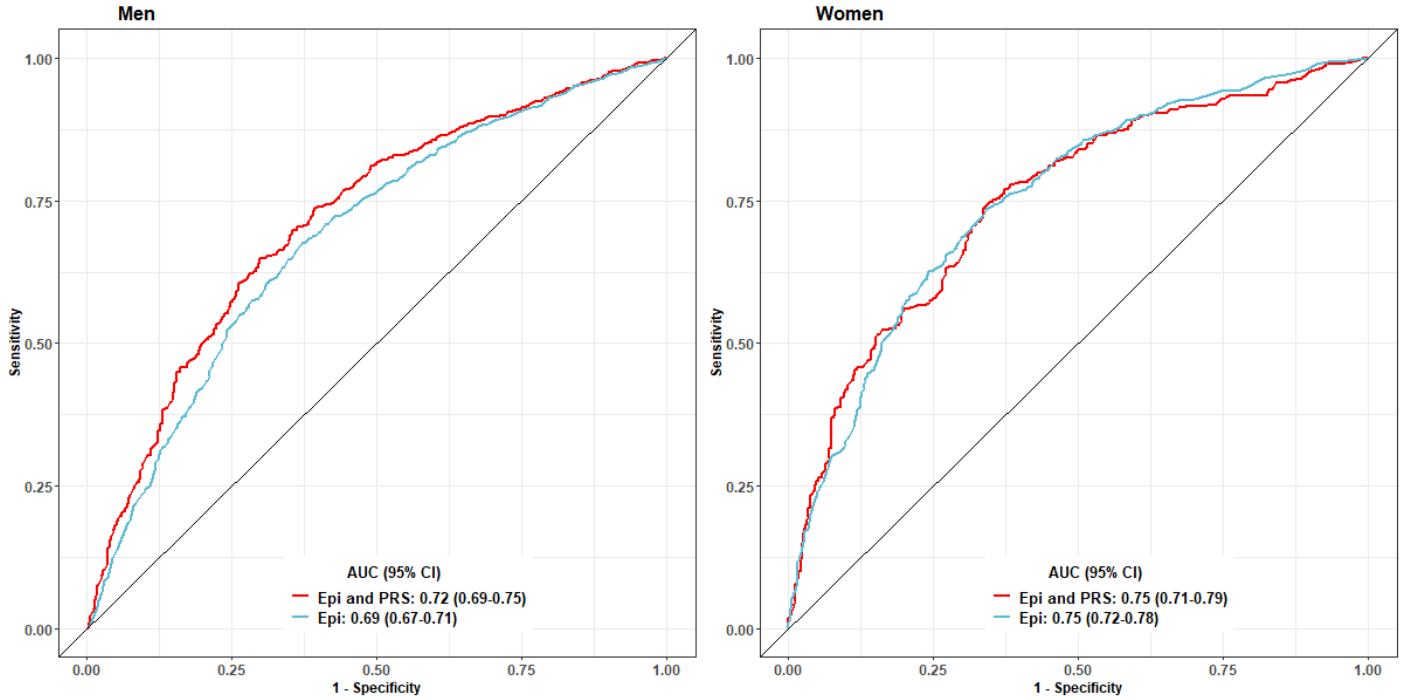
	Men	Women
Head and neck cancer	0.737	0.407
Oral cavity cancer	0.630	0.306
Oropharyngeal cancer	0.830	0.890

Supplemental Figure 1. Flowchart of the study subjects

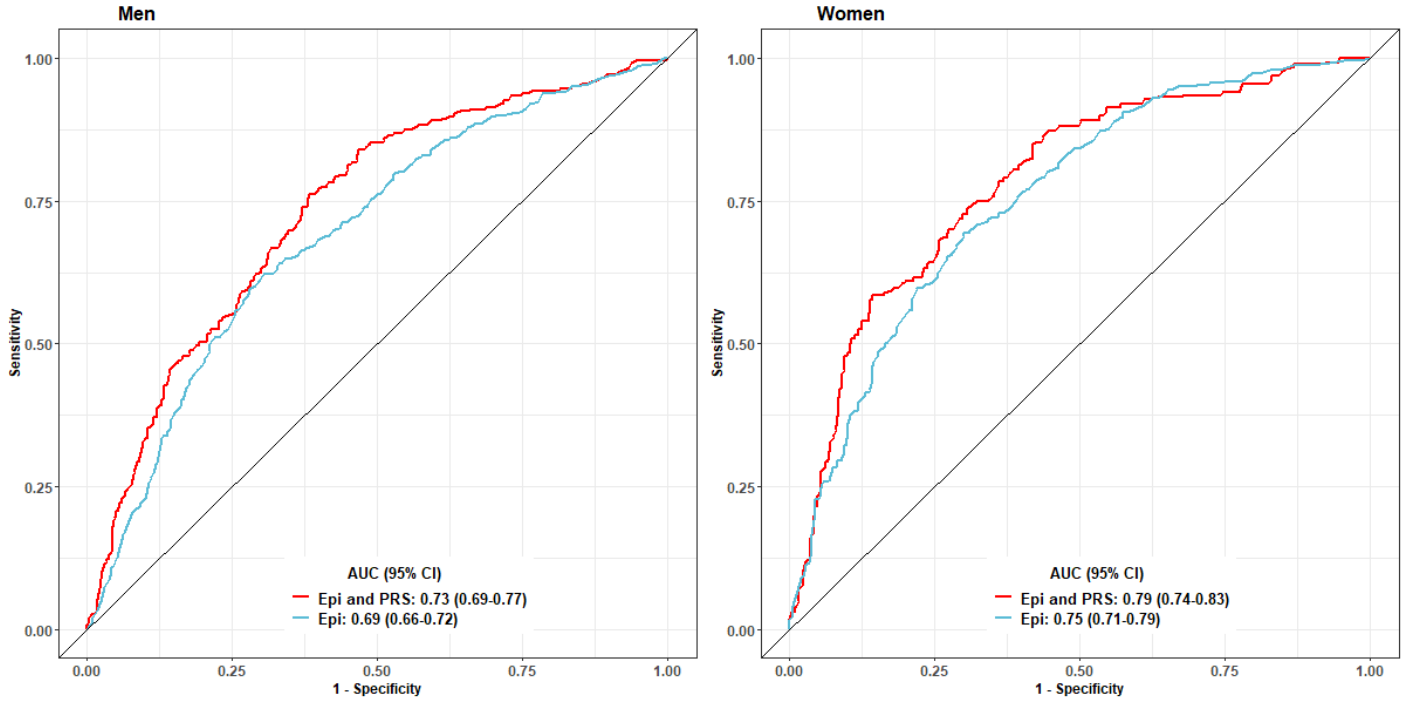


Supplemental Figure 2. Receiver Operating Characteristic Curves (ROCs) of risk models for head and neck cancer in hold-out testing set

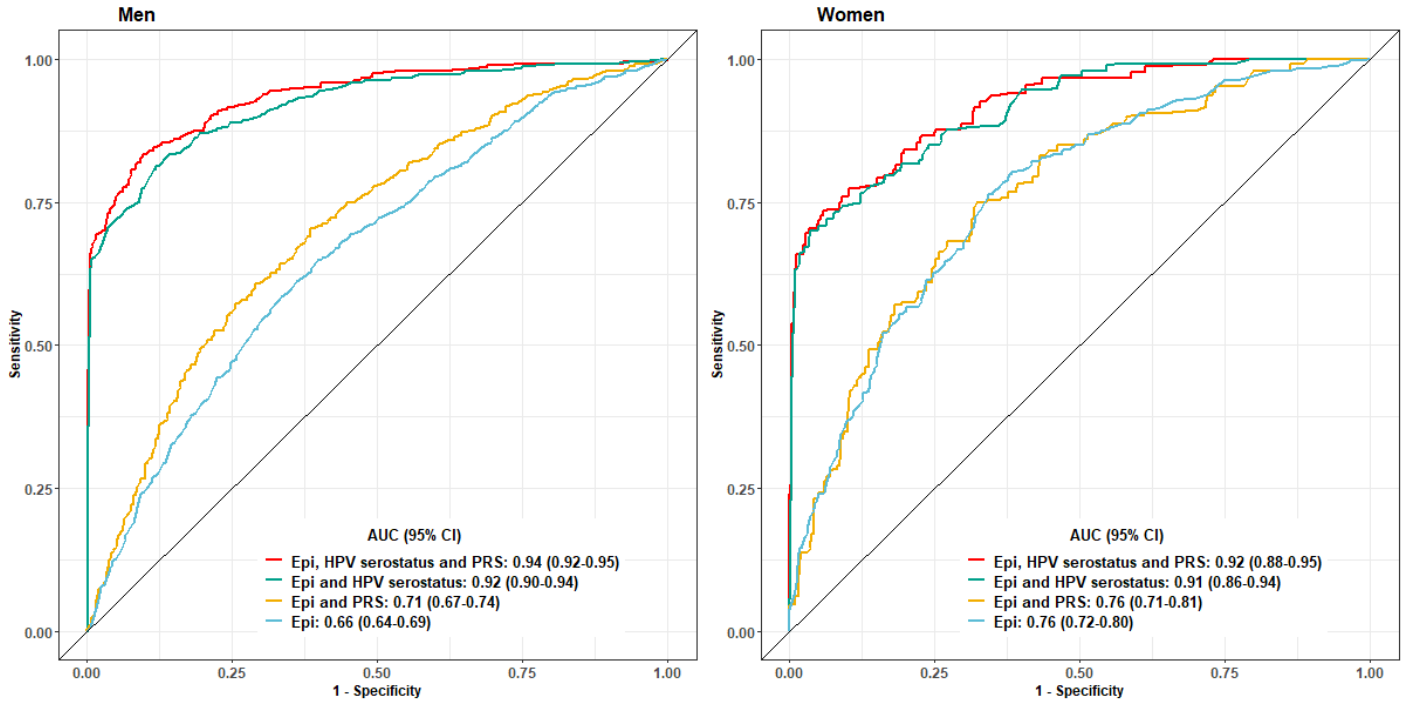
A. Head and neck cancer



B. Oral cavity cancer



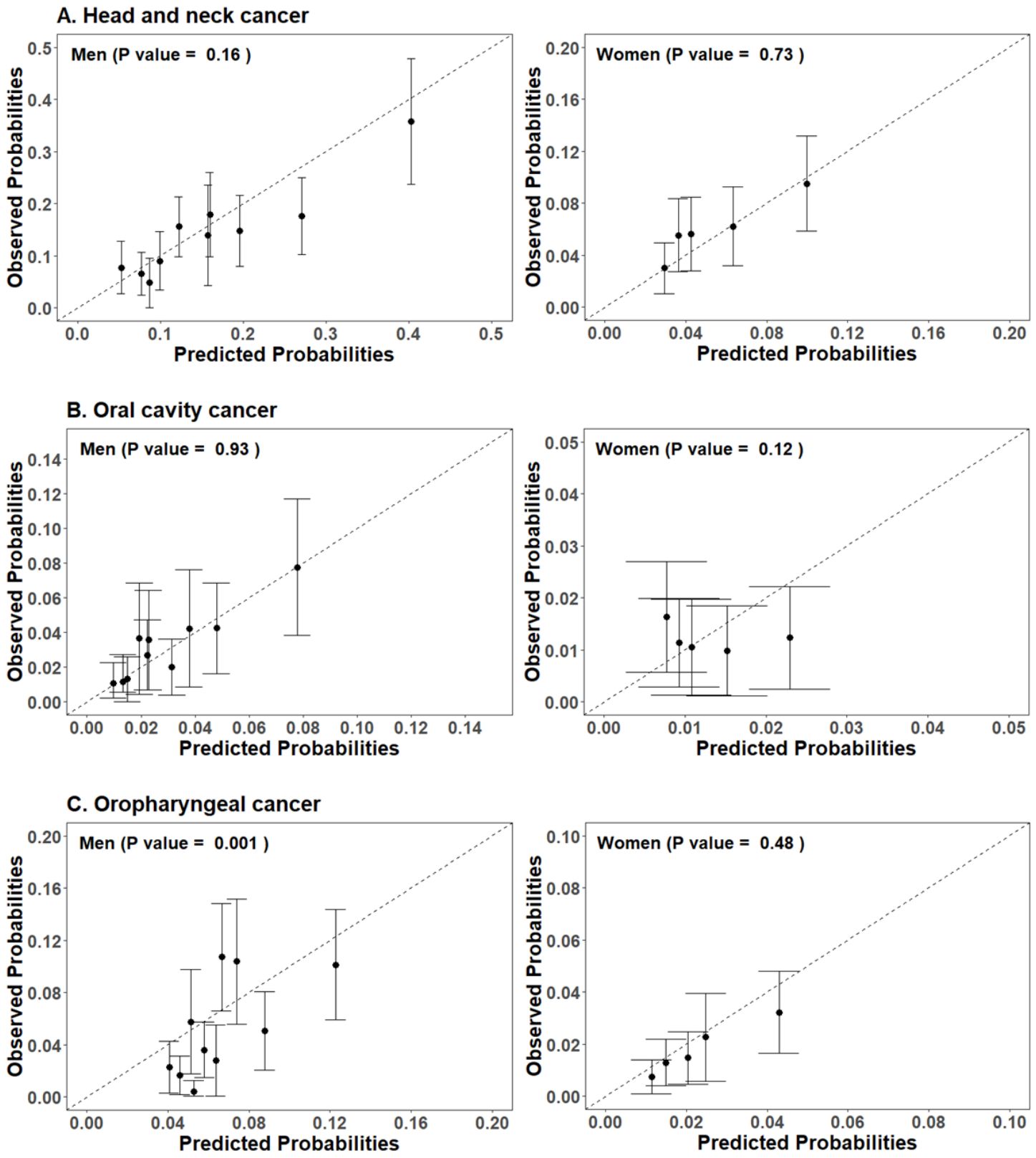
C. Oropharyngeal cancer



HPV, human papillomavirus; PRS, polygenic risk scores.

Epidemiological (epi) risk factor model includes age, smoking packyears, alcohol drinking intensity and education. The model for head and neck cancer overall (**A**) and oral cavity cancer (**B**) include epidemiological risk factors and polygenic risk score. The model of oropharyngeal cancer (**C**) includes epidemiological risk factor, HPV serostatus and polygenic risk score. The left and right panel shows the ROC curves of risk models for head and neck cancer in men and women, respectively.

Supplemental Figure 3. Calibration plot comparing predicted probability with observed probability.



The model for head and neck cancer overall (**A**) and oral cavity cancer (**B**) include epidemiological risk factors and polygenic risk score. The model of oropharyngeal cancer (**C**) includes epidemiological risk factor, HPV serostatus and polygenic risk score. The calibration lines for men (Left panel) are plotted in deciles of predicted probability and for women (Right panel) are plotted in quintile due to smaller sample size. P-values are based on Hosmer-Lemeshow test.