Supplementary Online Content

Agalliu I, Gapstur S, Chen Z, et al. Associations of oral α-, β-, and γ-human papillomavirus types with risk of incident head and neck cancer. *JAMA Oncology*. Published online January 21, 2016. doi:10.1001/jamaoncol.2015.5504.

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods

Cohort Selection

Of the 70,004 CPS-II-NC participants who provided a mouthwash sample, we excluded 16,664 who had a previous cancer diagnosis, 158 whose oral rinse specimens were inadequate, and two whose gender data were missing. Among the 53,180 participants in the at-risk cohort after exclusions, 69 were diagnosed with a primary incident head and neck cancer (HNC): 28 cancers of the larynx, 24 of the oral cavity, 13 of the oropharynx, two of the hypopharynx, and two unspecified oral locations, between the time of oral rinse collection (2001-2002) and the end of follow up (6/30/2009). HNC cases were first identified from self-reported biannual questionnaires, and then verified through medical records and/or linkage to tumor registries. Approximately 87% (n=60) of these cancers were squamous cell carcinomas.

Of the 55,866 PLCO control arm participants who provided a mouthwash sample, we excluded 5,526 who had a previous cancer diagnosis, and 6,870 whose oral rinse specimens were exhausted or unavailable. Among the 43,470 participants in the at-risk cohort after exclusions, 80 were diagnosed with a primary incident HNC: 32 cancers of the larynx, 26 of the oral cavity, 14 of the oropharynx, three of the hypopharynx, and five unspecified oral locations, between the time of oral rinse collection (2000-2004) and the end of follow up (7/31/2011). In the PLCO trial, incident HNC cases were initially identified via annual follow-up questionnaires and/or telephone calls and then confirmed by review of the medical records at 10 U.S. screening centers. A total of 72 HNC (90%) were squamous cell carcinomas.

In each respective cohort, three controls were selected for each case from the at-risk cohort who were alive on the diagnosis date of the case and who had no prior history of cancer on this date. Controls were individually matched to cases on gender, race (white, black or other/unknown), date of birth (±6 months), and date of oral rinse collection (±30 days for CPS-II-NC cohort, and ±3 months for PLCO trial).

HPV DNA Detection

Next Generation Sequencing Assay. This method consisted of three separate PCR amplification assays that targeted primer-binding sites within the L1 (NG-S and NG-F assays) and E1 ORFs (NG-E1) [1]. Each DNA sample was amplified using sample-specific 8-bp barcoded primers. Successful amplification of predicted fragment sizes was verified by gel electrophoresis and PCR products were pooled and sequenced on an Illumina HiSeq 2000/2500 (Illumina Inc., San Diego, CA) at the Epigenomics Shared Facility at Einstein, using 150-bp paired-end reads. The reads were de-multiplexed, filtered for quality, and blasted against a PV reference database as described previously[1]. The following criteria were used to define HPV positivity: an HPV type was considered positive if the HPV read counts were ≥ 100, 500 and 250 for the NG-S, NG-F and NG-E1 assays, respectively, and the percent of HPV reads for that type were ≥ 2%, ≥ 5% and ≥ 5% of the total HPV reads, respectively.

Real-Time PCR Assay for HPV16. This assay consisted of two PCR reactions, one amplified a fragment within the HPV16 E6 and the other a fragment within the L1 region. A fragment of the CSF cellular gene was simultaneously amplified in both PCRs as an internal control to quantitate the host cellular DNA.

Statistical Analysis

Imputation: From both cohorts combined, there were 3 cases and 9 controls with missing information on pack-years of smoking, and 23 cases and 53 controls with missing data on alcohol drinks/week. We imputed these missing data by assigning the median values of pack-years of smoking or alcohol drinks/week in specific strata defined by study (CPS-II-NC or PLCO), case-control status, and smoking status (never, former or current smoker), which was available for all subjects in both cohorts. Mean (or median) imputation is a simple way to address the issue of missing data; however, this method may underestimate the variance of data, in particular when the missing rate is high[2]. Although the percent of missing data was small in our study, this problem could be overcome by multiple imputation (MI) [3].

To verify the results from median imputation, we performed MI using R package mi [4] to impute missing values of pack-years of smoking and alcohol drinks per week. The mi package uses a chained equation approach algorithm that iteratively draws imputed values from the conditional distribution for each variable given the observed and imputed values of the other variables in the data. There were no meaningful differences between MI and median imputation in the results of associations between HPV exposures and HNSCC cancer risk.

Permutation for Multiple Comparisons: A permutation procedure was used to account for multiple comparisons of several HPV types and species [5,6]. In brief, for each of 10,000 replicates, the matched pairs were permuted by shuffling the case-control status within a pair. For each permuted dataset, conditional logistic regression models were fit for HPV types or species and the minimum of these p-values were kept. This provided an empirical distribution of p-values under the null hypothesis of no association. The permutation p-value for an HPV type was obtained by comparing their observed p-values to this empirical distribution. Permutation p-values can be interpreted as the probability of observing a p-value less than or equal to what was observed under the null hypothesis of no association with HNSCC for any of the HPV types or species analyzed. After this procedure, an HPV type or species was considered to be statistically significantly associated with risk of HNSCC if the permuted p-value was ≤ 0.05 (two-sided).

eReferences:

1. Fonseca AJ, Taeko D, Chaves TA, et al. HPV infection and cervical screening in socially isolated Indigenous women inhabitants of the Amazonian rainforest. *PLoS One.* 2015;10(7):e0133635
2. Barzi F, Woodward M. Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. Am J Epidemiology 2004; 160(1):34-45.
3. Rubin, D. B. Multiple imputation after 18+ years. Journal of American Statistical Association, 1996: 91(434):473–489.
4. Su YS, Gelman A, Hill J, Yajima M. Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. Journal of Statistical Software 2011, 45(2), 1-31.
5. Westfall, P. H. and Young, SS. (1993). Resampling-Based Multiple Testing: Examples and Methods for p-Value adjustment. Wiley, New York.
6. Dudoit S, Popper-Shaffer J, Boldrick JC. Multiple hypothesis testing in microarray experiments. Statistical. Sci. 2003; 18 (1), 71-103.

eTable 1. Selected lifestyle characteristics of incident cases of head and neck squamous cell carcinoma (HNSCC) and matched controls

Characteristics	CPS-II-NC Cohort			PLCO Cohort		
	HNSCC N=60	**Controls N=180**	**p[1]**	**HNSCC N=72**	**Controls N=216**	**p[2]**
BMI Group (kg/m^2); n (%)			0.57			0.64
< 25	21 (35.0)	60 (33.3)		28 (38.9)	68 (31.5)	
25-29.9	22 (36.7)	79 (43.9)		29 (40.3)	99 (45.8)	
≥ 30	10 (16.7)	29 (16.1)		13 (18.1)	45 (20.8)	
Missing	7 (11.7)	12 (6.7)		2 (2.8)	4 (1.9)	
Education; n (%)			0.05			0.32
<12 grade	7 (11.7)	12 (6.7)		9 (12.5)	17 (7.9)	
HS/Vocational	22 (36.7)	52 (28.9)		26 (36.1)	75 (34.7)	
Some College	15 (25.0)	31 (17.2)		17 (23.6)	38 (17.6)	
College Graduate	10 (16.7)	36 (20.0)		10 (13.9)	37 (17.1)	
Graduate Degree	6 (10.0)	48 (26.7)		10 (13.9)	49 (22.7)	
Marital Status; n (%)			0.83			0.51
Married	43 (71.7)	132 (73.3)		53 (73.6)	179 (82.9)	
Separated/Divorced	2 (3.3)	4 (2.2)		8 (11.1)	17 (7.9)	
Widowed	3 (5.0)	10 (5.6)		9 (12.5)	15 (6.9)	
Never Married	-	3 (1.7)		2 (2.8)	5 (2.3)	
Missing	12 (20.0)	31 (17.2)		-	-	

[1] P-values comparing CPS-II-NC cases and controls; [2] P-values comparing PLCO cases and controls

eTable 2. Associations of Beta and Gamma HPVs species and types with risk of incident HNSCC subtypes

2A. Oropharynx Cancer					
	Cases N = 25	Controls N = 75	Adjusted Model *		
Beta HPV Species†	n (%)	n (%)	OR*	95% CI	p
Any Beta HPV	13 (52.0)	41 (54.7)	1.77	0.42 - 7.40	0.43
Any Beta1 HPV	10 (40.0)	29 (38.7)	3.12	0.61 - 16.09	0.17
Any Beta2 HPV	9 (36.0)	26 (34.7)	2.05	0.50 - 8.34	0.32
Specific Beta HPV Types†					
Beta1 HPV5	3 (12.0)	3 (4.0)	7.42	0.98 - 56.82	0.054
Beta1 HPV36	2 (8.0)	5 (6.7)	1.93	0.23 - 16.35	0.55
Clade Beta 1 HPVs 5, 36, 47, & 143	5 (20.0)	7 (9.3)	5.01	0.76 - 32.89	0.09
Beta2 HPV17	2 (8.0)	4 (5.3)	2.60	0.32 – 20.87	0.37
Beta2 HPV38	6 (24.0)	4 (5.3)	**7.28**	**1.33 - 39.72**	**0.02**
Gamma HPV Species†					
Any Gamma HPV	10 (40.0)	23 (30.7)	**4.64**	**1.03 – 20.91**	**0.045**
Any Gamma7 HPV	5 (20.0)	8 (10.7)	**4.42**	**1.00 – 19.50**	**0.05**
Any Gamma10 HPV	2 (8.0)	4 (5.3)	2.36	0.24 - 22.85	0.46
Any Gamma12 HPV	2 (8.0)	4 (5.3)	2.52	0.32 – 20.08	0.38

*OR and 95% CI are estimated from conditional logistic regression models adjusted for smoking, alcohol consumption, HPV16 DNA detection
and study cohort (CPS-II-NC vs. PLCO)

† For beta and gamma HPV species and types data were presented if the prevalence of HPV exposure was ≥5% in controls (with the exception of beta1 HPV5).

2B. Oral Cavity Cancer					
	Cases N = 43	Controls N = 128	Adjusted Model *		
Beta HPV Species†	n (%)	n (%)	OR*	95% CI	p
Any Beta HPV	27 (62.8)	77 (60.2)	1.11	0.43 - 2.86	0.83
Any Beta1 HPV	22 (51.2)	57 (44.5)	1.37	0.54 - 3.45	0.51
Any Beta2 HPV	22 (51.2)	57 (44.5)	2.05	0.81 - 5.19	0.13
Any Beta3 HPV	10 (23.3)	19 (14.8)	1.66	0.49 - 5.59	0.41
Specific Beta HPV Types†					
Beta1 HPV5	12 (27.9)	17 (13.3)	**5.34**	**1.51 - 18.80**	**0.01**
Beta1 HPV19	5 (11.6)	7 (5.5)	**7.68**	**1.20 - 49.27**	**0.03**
Beta1 HPV36	10 (23.3)	11 (8.6)	**4.46**	**1.26 - 15.78**	**0.02**
Clade Beta1 HPVs 5, 36, 47, 143	14 (32.6)	21 (16.4)	**3.75**	**1.23 - 11.45**	**0.02**
Beta2 HPV17	5 (11.6)	6 (4.7)	**8.09**	**1.46 - 44.85**	**0.017**
Beta2 HPV37	5 (11.6)	8 (6.3)	2.68	0.51 - 13.94	0.24
Beta2 HPV38	7 (16.3)	17 (13.3)	2.26	0.54 - 9.54	0.26
Gamma HPV Species†					
Any Gamma HPV	20 (46.5)	50 (39.1)	1.71	0.63 – 4.66	0.29
Any Gamma10 HPV	8 (18.6)	12 (9.4)	2.09	0.58 – 7.55	0.26
Any Gamma11 HPV	4 (9.3)	4 (3.1)	**7.47**	**1.21 – 46.17**	**0.03**
Any Gamma12 HPV	6 (14.0)	7 (5.5)	**6.71**	**1.47 – 30.75**	**0.01**

*OR and 95% CI are estimated from conditional logistic regression models adjusted for smoking, alcohol consumption, HPV16 DNA detection
and study cohort (CPS-II-NC vs. PLCO)

† For beta and gamma HPV species and types data were presented if the prevalence of HPV exposure was ≥5% in controls (with the exception of any gamma 11 HPV species)

2C. Larynx Cancer[ξ]					
	Cases N = 64	Controls N = 193	Adjusted Model *		
Beta HPV Species[†]	n (%)	n (%)	OR*	95% CI	p
Any Beta HPV	43 (67.2)	114 (59.1)	1.92	0.77 - 4.80	0.16
Any Beta1 HPV	32 (50.0)	80 (41.5)	1.97	0.83 - 4.67	0.13
Any Beta2 HPV	34 (53.1)	92 (47.7)	1.78	0.76 - 4.09	0.17
Any Beta3 HPV	14 (21.9)	30 (15.5)	2.69	0.99 - 7.17	0.06
Specific Beta HPV Types[†]					
Beta1 HPV5	13 (20.3)	23 (11.9)	**2.71**	**1.00 - 7.43**	**0.05**
Beta1 HPV36	7 (10.9)	25 (13.0)	1.20	0.35 - 4.05	0.77
Clad Beta1 HPVs 5, 36, 47, 143	14 (21.9)	37 (19.2)	1.84	0.71 - 4.77	0.21
Beta2 HPV37	6 (9.4)	12 (6.2)	1.97	0.51 - 7.56	0.32
Beta2 HPV38	11 (17.2)	26 (13.5)	2.50	0.80 - 7.85	0.12
Gamma HPV Species[†]					
Any Gamma HPV	30 (46.9)	67 (34.7)	1.84	0.82 – 4.12	0.14
Any Gamma10 HPV	11 (17.2)	12 (6.2)	2.45	0.71 – 8.37	0.16
Any Gamma11 HPV	5 (7.8)	5 (2.6)	**7.49**	**1.10 – 51.04**	**0.04**
Any Gamma12 HPV	8 (12.5)	10 (5.2)	**5.31**	**1.13 – 24.95**	**0.03**

[ξ] Larynx cancer includes five cases of hypopharynx cancer

*OR and 95% CI are estimated from conditional logistic regression models adjusted for smoking, alcohol consumption, HPV16 DNA detection
and study cohort (CPS-II-NC vs. PLCO)

[†] For beta and gamma HPV species and types data were presented if the prevalence of HPV exposure was ≥5% in controls (with the exception of any gamma 11 HPV species)