Published in final edited form as:
Oral Oncol. 2010 February ; 46(2): 100. doi:10.1016/j.oraloncology.2009.11.004.

Moderate predictive value of demographic and behavioral characteristics for a diagnosis of HPV16-positive and HPV16-negative head and neck cancer

Gypsyamber D'Souza[1,*], **Hao H. Zhang**[2], **Warren D. D'Souza**[2], **Robert R. Meyer**[3], and **Maura L. Gillison**[4]

[1] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

[2] Department of Radiation Oncology, University of Maryland School of Medicine

[3] Computer Sciences Department, University of Wisconsin-Madison

[4] The Ohio State University Comprehensive Cancer Center-James Cancer Hospital and Solove Research Institute

Abstract

Background—Patients with HPV-positive and HPV-negative head and neck squamous cell carcinoma (HNSCC) are significantly different with regard to sociodemographic and behavioral characteristics that clinicians may use to assume tumor HPV status.

Methods—Machine learning methods were used to evaluate the predictive value of patient characteristics and laboratory biomarkers of HPV exposure for a diagnosis of HPV16-positive HNSCC compared to in-situ hybridization, the current gold standard.

Results—Models that used a combination of demographic characteristics such as age, tobacco use, gender, and race had only moderate predictive value for tumor HPV status among all patients with HNSCC (positive predictive value [PPV]=75%, negative predictive value [NPV]=68%) or when limited to oropharynx cancer patients (PPV=55%, NPV=65%) and thus included a sizeable number of false positive and false negative predictions. Prediction was not improved by the addition of other demographic or behavioral factors (sexual behavior, income, education) or biomarkers of HPV16 exposure (L1, E6/7 antibodies or DNA in oral exfoliated cells).

Conclusions—Patient demographic and behavioral characteristics as well as HPV biomarkers are not an accurate substitute for clinical testing of tumor HPV status.

Keywords

human papillomavirus; HPV; oropharyngeal; HNSCC; cancer; prediction; screening; antibodies; oral HPV; profiling; tobacco; in-situ; tumor; machine learning

*Corresponding Author: Gypsyamber D'Souza, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St. E6132B, Baltimore, MD 21205, gdsouza@jhsph.edu, (p) 410-502-2583 (f) 410-614-2632.

Conflicts of Interest Statement
None Declared

Introduction

Recent research has demonstrated that HPV-positive and HPV-negative head and neck squamous cell carcinomas (HNSCC) are two distinct cancers with different etiologies.[1] Human papillomavirus (HPV) is found in ~20% of all head and neck squamous cell carcinomas (HNSCC), and 50–85% of oropharynx cancers.[2] The majority (~90%) of HPV-associated HNSCC are caused by a single HPV type: HPV16.[3]

The patient populations[4], and clinical outcomes[56] of HPV-positive and negative HNSCC patients are significantly different. Median age for HPV-positive HNSCC patients is five years less than that for HPV negative HNSCC patients. HPV-positive patients are also less likely to use tobacco or alcohol, more likely to be White, and to have a higher median number of lifetime sexual partners.[1] In addition, survival is markedly better for patients with HPV-positive when compared with HPV-negative HNSCC[6], despite the fact that HPV-positive HNSCC are more likely to be detected as late-stage cancers.[5]

Given the different profiles for HPV-positive and HPV-negative HNSCC patients, physicians may consciously or unconsciously assume prognosis [7] or tumor HPV status on the basis of simple patient demographics. For example, the young, White patient with little tobacco and/or alcohol use history may be assumed to have HPV-associated HNSCC while older patients with a significant tobacco and/or alcohol use history may be assumed to have HPV-negative HNSCC. These assumptions may, in imperceptible ways, influence treatment decision making. Thus, we chose to evaluate how well demographics characteristics and biomarkers for HPV exposure predict tumor HPV status in a cohort of newly diagnosed HNSCC patients.

Patients and Methods

Study population and characteristics

A dataset from a cohort study of 255 patients diagnosed with incident HNSCC at the Johns Hopkins Hospital between 2000 and 2006 was used for this analysis. The majority of cases (94%) were included in a previous case-control study from this cohort.[1] Patients included in the analysis were diagnosed with oropharyngeal (n=119), oral cavity (63), larynx (50), paranasal sinus (8), hypopharynx (8), nasopharynx (3) and unknown primary (4) cancers. This study was approved by the Johns Hopkins Institutional Review Board.

Demographic variables included age (continuous), gender, race and ethnicity (categorized as White non-Hispanic, Black non-Hispanic, and Other which included those of Asian and Middle Eastern race and/or Hispanic ethnicity), a continuous measure of lifetime pack-years of tobacco use, and number of lifetime oral sex partners in ordinal categories (0,1, 2–5,6–10, 11–15,16–25, 26–50, 51–100, >100).

Laboratory analysis

As outlined below, four different biomarkers for HPV16 exposure were considered in this analysis.

HPV16 in-situ hybridization (gold-standard)—Patients were classified as having either HPV16-positive or HPV-16 negative HNSCC based upon HPV16 detection in formalin-fixed and paraffin-embedded tumors using the in situ hybridization/catalyzed signal amplification method for biotinylated probes (Dako GenPoint, Carpinteria, CA),[8] the current gold standard for determining case HPV status.[9] Specific staining of tumor cell nuclei for HPV16 defined a positive tumor.

Serologic analysis—Serum antibodies to the HPV16 major capsid protein, L1, were detected using a virus-like-particle based, enzyme-linked immunosorbent assay (ELISA).[10] Serum antibodies to the HPV 16 oncoproteins, E6 and E7, were detected in an ELISA that utilizes the gluthione S-transferase (GST) capture method with bacterially expressed full-length E6 and E7 fused to GST as the antigen.[11] The assay cutoff points for seropositivity to both L1 and E6/E7 were set using a low-risk reference population and dichotomized as seropositive or seronegative.

HPV16 L1 antibodies are a measure of ever having been infected with HPV at any site in the body and thus serve as a marker for HPV exposure. HPV16 E6 and E7 antibodies are measures of ever having had these HPV16 oncogenes expressed at any site in the body and thus are a marker for cancer.

HPV16 DNA in oral exfoliate cells: purification and detection—The presence of HPV16 DNA in purified DNA from oral exfoliated cells was measured using real-time PCR targeted to the E6 coding region.[12] Individuals with ≥ 1 copy of HPV16 were considered positive. Detection of HPV DNA in oral exfoliate cells is a measure of current oral infection and/or presence of an HPV-positive oral cancer.

Predictive Models

We used machine learning (ML) algorithms to optimize prediction of tumor HPV16 status among HNSCC cases as determined by the current gold standard assay, HPV16 in situ hybridization. ML algorithms attempt to learn complex patterns in datasets and then make knowledge-driven decisions to form predictions based on that data. Decision trees, an example of a ML construct, map observations about an item to conclusions about its target value via binary tests involving attribute ranges or properties.[13, 14] Another ML approach, Support Vector Machines [15, 16], was also evaluated and produced similar results (data not shown). Prediction models were created using a random sample of half of the data (*training subset)* and these models were then validated using the other half of the dataset (*testing subset*), a process referred to as cross-validation. This process was repeated 50 times and the results of these models were averaged (Table 2).

With decision tree algorithms value ranges that lead to the best prediction of the data are selected at each node until sub-groups with the same outcome value are achieved or no further distinguishing predictors can be determined. The large and complex decision tree created from the training data is pruned when applied to the test datasets to remove branches with higher error rates as well as divisions that provide little gain in statistical accuracy. Thus decision trees are designed to provide prediction mechanisms that can be applied to new datasets (i.e. to predict patient diagnoses) as opposed to descriptive statistical tests which summarize observations within a dataset but are less useful for predicting what will be observed in other populations.

Differences in patient characteristics between HPV16 positive and negative patients were compared using the Chi-square test. Prediction accuracy was evaluated using positive predictive value (PPV= percentage of truly positive among those predicted HPV positive via a given test), and negative predictive value (NPV=percentage of truly negative among those predicted negative). Sensitivity (percentage predicted positive among all truly positive) and specificity (percentage predicted negative among all truly negative) were also calculated.

Results

Characteristics of cases

When tumor HPV16 status was evaluated by in-situ hybridization for HPV16, 71% (85 of 119) oropharyngeal were HPV16-positive. In contrast, 5.1% (7 of 136) of non-oropharyngeal HNSCC were HPV16-positive including, two (50%) of four unknown primary HNSCC, three (6.0%) of 50 laryngeal, one (1.6%) of 63 oral cavity, one (12%) of eight paranasal sinus and none (0%) of eleven hypopharynx and nasopharnx cases.

Characteristics of the 92 HPV16-positive and 163 HPV-negative HNSCC in this study are compared in Table 1. HPV-positive cases were more likely than HPV-negative cases to be young (median age 52 vs. 60, $p<0.001$), White (93% vs. 82%, $p=0.02$), to make over $50,000 a year (70% vs. 48%, $p=0.001$), to report ever having been diagnosed with an sexually transmitted infection (34% vs. 17%, $p=0.003$), and to have a higher lifetime number of oral sexual partners ($p<0.001$). In contrast, traditional head and neck risk factors were less common in HPV-positive cases than in HPV-negative cases, including tobacco use ($p<0.001$), and poor oral hygiene ($p=0.001$) (Table 1).

Although tobacco use is associated with HPV-negative tumor status, there were HPV negative cases who were non-smokers as well as HPV-positive cases who were heavy smokers. Among the 54 oropharyngeal cancer cases who reported never smoking, most but not all (87%) were HPV16-positive. More surprising is that 51% of 41 oropharyngeal cases who reported heavy smoking (>20 pack-years) were HPV16-positive. Among non-OP HNSCC cases, HPV-positivity was rare among both the 42 non-smoking (4.8%) and the 80 (6.2%) heavy smoking cases.

Predicting HPV16 tumor status of HNSCC cases using demographic and risk behavior data

Despite the significant association of demographic and behavioral variables with HPV16 tumor status (Table 1) the ability to use these variables to *predict* tumor HPV16 status (using machine learning methods) was only moderate (Table 2). Specifically, when tobacco use, age, gender, and race were considered, there was only moderate predictive ability for tumor HPV status among oropharyngeal patients (positive predictive value: PPV=55%, negative predictive value: NPV=66%). Addition of sexual risk behavior, income and education did not further improve tumor HPV prediction (data not shown). When all HNSCC were considered, demographic variables had better predictive ability for tumor HPV16 status (PPV=75%), and similar NPV, but still included a sizable proportion of false positive and false negative predictions.

Predicting HPV16 tumor status of HNSCC cases using HPV biomarkers

As there was interest in whether non-tumor biomarkers could distinguish HPV-positive and HPV-negative cancers, we evaluated three other HPV16 measures. Of the 85 oropharyngeal cancers positive for HPV16-positive by in-situ hybridization, 63% were seropositive for HPV16 L1 antibodies, 78% were seropositive for E6 and/or E7 antibodies, and 19% had HPV16 DNA detected in oral exfoliated cells. Of the 34 oropharyngeal cancers that were HPV16-negative, 82% were seronegative for HPV16 L1 antibodies, 73% were seronegative for E6 and/or E7 antibodies, and 97% had no HPV16 DNA detected in oral exfoliated cells. Results were similar when all HNSCC were considered (Table One).

When tumor HPV16 status was *predicted* using both HPV biomarkers and demographic variables, negative and positive prediction accuracy was better than with demographic variables alone (Table 2), but still included a sizeable proportion of false positive and false negative predictions. PPV and NPV were higher when patient characteristics and biomarkers

were used to predict tumor HPV status among all HNSCC (PPV=79–83%; NPV=74–85%) than when only oropharyngeal cancers were considered (PPV=59–60%; NPV=71–77%; Table 2).

Discussion

While patient characteristics are statistically associated with tumor HPV status, that association does not necessarily translate into clinically useful diagnostic tools. Despite the use of sophisticated machine learning methods to develop prediction models from several likely risk factors, predictive ability of patient characteristics for tumor HPV status was only moderate. The presence of HPV-positive cases that have the profile of traditional (HPV-negative) HNSCC patients, as well as, more rarely, HPV-negative cases that have a profile more commonly observed among HPV-positive cases limits the ability of any model based on these characteristics to distinguish tumor etiology.

Recent research indicates that HPV-positive HNSCC has better response to therapy and improved survival compared to HPV-negative HNSCC.[6, 17, 18] Clinical trials are currently being designed to evaluate whether tailoring cancer therapy according to HPV status can improve patient outcomes. Currently, clinical guidelines for the treatment of HPV-positive and negative HNSCC do not differ, and tumor HPV testing is not yet the standard of care. However, because of its prognostic significance, some treatment centers are routinely evaluating the HPV status of oropharyngeal tumors. In centers where HPV testing is not performed, patient profiles may consciously or unconsciously be used as surrogates for HPV status and thus might influence treatments prescribed.

While HPV-associated HNSCC are often incorrectly viewed as a disease limited to non-smokers non-drinkers, this study highlights that many HPV-positive oropharynx cancers occur among individuals who smoke. As HPV accounts for the majority of oropharyngeal cancers in non-smokers and causes a significant proportion of oropharyngeal cancer in smokers, routine HPV testing of oropharynx cancers should be considered.

Prediction of tumor HPV status was better when all HNSCC were considered than when only oropharyngeal cancers were evaluated. Although sensitivity of patient characteristics for tumor HPV status was higher among oropharyngeal cancers, the specificity was poor (i.e. many false positives). Some of these oropharyngeal cancer patients whose tumors were negative for HPV16 but had characteristics similar to those of patients with HPV16-positive tumors may have cancer caused by other oncogenic HPV types not tested for; thus we may have underestimated specificity in these models.

The predictive ability of any test depends on sensitivity, specificity, which are inherent properties of a test, as well as the prevalence of the outcome (in this case HPV) in that population. Therefore, in populations with a low proportion of HPV-associated HNSCC, patient characteristics and biomarkers would be expected to have even lower PPV & NPV than observed in this study. Conversely, in populations where a larger proportion of HNSCC are due to HPV, higher predictive ability is expected.

Although both HPV antibodies and HPV DNA in exfoliated oral cells are associated with tumor HPV status they had moderate predictive ability for tumor HPV status in this study. This moderate predictive ability is not surprising given what is known about the limitations of each of these biomarker assays. Sensitivity of HPV16 L1 antibodies, a marker of lifetime HPV exposure, is 50–70%, as many individuals exposed to HPV never seroconvert.[19] These antibodies also lack specificity for the oral cavity [9, 19] as they represent lifetime history of HPV exposure at any bodysite (anogenital as well as oral). E6/E7 serology, a marker for HPV oncogene expression, suffers the same sensitivity and specificity limitations.[19] Detection of

HPV DNA in oral exfoliated cells is believed to represent current oral infection, and thus has imperfect sensitivity for HPV-associated HNSCC caused by an HPV infection likely acquired many years previously that may no longer be detectable in exfoliated cells. Specificity of HPV DNA in oral exfoliate cells is also limited as current transient infections not related to cancer may also be detected.

This research has several limitations. Tumors were only tested for one type of HPV (HPV16) and thus may have misclassified as HPV-negative some cancers caused by other oncogenic HPV types. However, over 90% of HPV-HNSCC are caused by HPV16. In addition when a subset of fresh frozen tumors from 60 cases in this study were tested for 37 types of HPV, other oncogenic HPV infections were only detected in 3.3% that had been thought to be HPV-negative[20] suggesting misclassification of HPV status was minimal. An additional limitation is that the data was from a single institution with a relatively high proportion of HPV-associated oropharyngeal cancers and thus will not represent the PPV and NPV of a lower prevalence setting. However, sensitivity and specificity of these measures will not change and given the inadequate prediction observed in a high prevalence setting the conclusions will apply to lower prevalence settings where even poorer prediction is expected.

This study demonstrates that assumptions about HNSCC HPV status should not be made based on either demographic, behavioral, or non-tumor HPV biomarkers as they do not have sufficient predictive ability to warrant use in HPV classification of HNSCC cases when tumor HPV detection is possible. This limited predictive ability underscores the adverse clinical impact that assuming HPV status based on patient characteristics could cause; for example unconsciously decreasing intensity of therapy for patients assumed incorrectly to be HPV-positive may negatively impact patient outcome.

Acknowledgments

Supported in part by grants from the Damon Runyon Cancer Research Foundation (to MG) and the National Institute of Dental and Craniofacial Research (DE016631 to MLG).

References

1. Gillison ML, D'Souza G, Westra W, Sugar E, Xiao W, Begum S, et al. Distinct risk factor profiles for human papillomavirus type 16-positive and human papillomavirus type 16-negative head and neck cancers. J Natl Cancer Inst 2008 Mar 19;100(6):407–20. [PubMed: 18334711]
2. WHO. IARC Monographs on the Evaluation of carcinogenic risks to humans. Lyon, France: International Agency for Research on Cancer; 2007.
3. Kreimer AR, Clifford GM, Boyle P, Franceschi S. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. Cancer Epidemiol Biomarkers Prev 2005 Feb;14(2):467–75. [PubMed: 15734974]
4. Benard VB, Johnson CJ, Thompson TD, Roland KB, Lai SM, Cokkinides V, et al. Examining the association between socioeconomic status and potential human papillomavirus-associated cancers. Cancer 2008 Nov 15;113(10 Suppl):2910–8. [PubMed: 18980274]
5. Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. J Natl Cancer Inst 2000;92(9):709–20. [PubMed: 10793107]
6. Fakhry C, Westra WH, Li S, Cmelak A, Ridge JA, Pinto H, 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 2008 Feb 20;100(4):261–9. [PubMed: 18270337]
7. Chen LM, Li G, Reitzel LR, Pytynia KB, Zafereo ME, Wei Q, et al. Matched-pair analysis of race or ethnicity in outcomes of head and neck cancer patients receiving similar multidisciplinary care. Cancer Prev Res (Phila Pa) 2009 Sep;2(9):782–91. [PubMed: 19737985]

8. Huang CC, Qiu JT, Kashima ML, Kurman RJ, Wu TC. Generation of type-specific probes for the detection of single-copy human papillomavirus by a novel in situ hybridization method. Mod Pathol 1998 Oct;11(10):971–7. [PubMed: 9796725]
9. Adelstein, DJ.; Ridge, JA.; Gillison, ML.; Chaturvedi, AK.; D'Souza, G.; Gravitt, PE., et al. Squamous Cell Head and Neck Cancer and the Human Papillomavirus. Head Neck; Summary of a State of the Science Meeting; November 9–10, 2008; Washington, D.C. 2009. p. 1391-1422.
10. Viscidi RP, Ahdieh-Grant L, Clayman B, Fox K, Massad LS, Cu-Uvin S, et al. Serum immunoglobulin G response to human papillomavirus type 16 virus-like particles in human immunodeficiency virus (HIV)-positive and risk-matched HIV-negative women. J Infect Dis 2003;187(2):194–205. [PubMed: 12552444]
11. Gange SJ, Barron Y, Greenblatt RM, Anastos K, Minkoff H, Young M, et al. Effectiveness of highly active antiretroviral therapy among HIV-1 infected women. J Epidemiol Community Health 2002 Feb;56(2):153–9. [PubMed: 11812817]
12. Gravitt PE, Peyton C, Wheeler C, Apple R, Higuchi R, Shah KV. Reproducibility of HPV 16 and HPV 18 viral load quantitation using TaqMan real-time PCR assays. J Virol Methods 2003 Sep;112 (1–2):23–33. [PubMed: 12951209]
13. Kotsiantis SB. Supervised machine learning: a review of classification techniques. Informatica 2007;31:249–68.
14. Witten, I.; Frand, E. Data Mining: Practical machine learning tools and techniques. 2. San Francisco: Morgan Kaufmann; 2005.
15. Platt, J. Fast training of support vector machines using sequential minimal optimization. In: Burges, CJC., editor. Advances in kernel methods: support vector learning. MA, USA: MIT Press Cambridge; 1999. p. 185-208.
16. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KK. Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Netw 2000;11(5):1188–93. [PubMed: 18249845]
17. Kumar B, Cordell KG, Lee JS, Worden FP, Prince ME, Tran HH, et al. EGFR, p16, HPV Titer, Bcl-xL and p53, sex, and smoking as indicators of response to therapy and survival in oropharyngeal cancer. J Clin Oncol 2008 Jul 1;26(19):3128–37. [PubMed: 18474878]
18. Kumar B, Cordell KG, Lee JS, Prince ME, Tran HH, Wolf GT, et al. Response to therapy and outcomes in oropharyngeal cancer are associated with biomarkers including human papillomavirus, epidermal growth factor receptor, gender, and smoking. Int J Radiat Oncol Biol Phys 2007;69(2 Suppl):S109–11. [PubMed: 17848274]
19. Gravitt, P.; Viscidi, R. Chapter 5: Measurement of Exposure to Human Papillomavirus. In: Rohan, T.; Shah, K., editors. Cervical Cancer: From Etiology to Prevention. London: Kluwer Academic Publishers; 2004. p. 119-41.
20. D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM, et al. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med 2007 May 10;356(19):1944–56. [PubMed: 17494927]

Table 1

Demographic and risk behavior information among 255 incident HNSCC cases diagnosed at Johns Hopkins hospital, by tumor HPV16 status.

Explanatory Variables	HPV16+ Cases N=92	HPV16− Cases N=163	p-value
Male	85%	76%	0.11
Caucasian, non-Hispanic	93%	82%	0.023
Income < $50,000	30%	52%	0.001
Currently married or lived as married	86%	62%	<0.001
Ever diagnosed with a STD	34%	17%	0.003
Age (in years)			
<45	12%	15%	<0.001
45–54	42%	22%	
55–64	34%	30%	
≥65	12%	33%	
Education			
Did not graduate high school	7.6%	21%	0.005
High school graduate	41%	44%	
College graduate	51%	35%	
Primary family member with history of HNSCC	9%	8%	0.92
Tooth loss			
None	74%	46%	<0.001
Some	12%	21%	
All	14%	33%	
Tobacco Use			
Current Smoker (within 1 year of diagnosis)	18%	37%	<0.001
Former Smoker	29%	33%	
Never smoker	52%	30%	
Alcohol Use			
Current drinker - drinks daily	8.7%	15%	0.19
Current drinker - less than daily	52%	44%	
Former drinker	18%	25%	
Never drinker	21%	15%	
Number of lifetime oral sex partners			
0	8%	32%	<0.001
1	14%	19%	
2 – 5	33%	31%	
6 – 15	24%	11%	
> 15	22%	7%	
ECOG Performance status			
0 (fully active)	83%	69%	0.009
≥1 (some physical restrictions)	10%	25%	

Explanatory Variables	HPV16+ Cases N=92	HPV16− Cases N=163	p-value
Unknown	8%	5%	
Tumor site			
Lip or oral cavity	2.2%	35%	<0.001
Oropharyngeal	89%	21%	
Nasopharynx, Hypopharynx, or Paranasal Sinus	1.1%	11%	
Larynx	3.3%	30%	
Unknown Primary	4.4%	2.6%	
HPV16 L1 Seropositive	59%	8%	<0.001
HPV16 E6 and/or E7 Seropositive	73%	15%	<0.001
HPV16 DNA detected in oral exfoliated cells	17%	2.0%	<0.001

Table 2

Prediction of head and neck squamous cell carcinoma (HNSCC) and oropharyngeal tumor HPV16 status using demographic and behavioral risk factors and non-tumor biomarkers. Positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity reported.

Variables used for prediction*	Oropharyngeal cancers				HNSCC			
	PPV	NPV	Sen	Spec	PPV	NPV	Sen	Spec
Tobacco use, age, gender	53%	61%	84%	25%	59%	66%	75%	53%
Tobacco use, age, gender, race^	55%	66%	84%	31%	75%	68%	76%	60%
HPV16 L1, tobacco use, age, gender, race^	59%	71%	84%	40%	79%	83%	63%	90%
HPV16 E6/7, tobacco use, age, gender, race^	60%	75%	85%	43%	73%	85%	73%	85%
Oral HPV16 DNA, tobacco, age, gender, race^	60%	77%	87%	41%	83%	74%	42%	97%

^Addition of lifetime number of oral sexual partners, family income, and education to multivariate model did not improve prediction

*In these models age was included as a continuous variable, gender was binary, race and ethnicity categories included White non-Hispanic, Black non-Hispanic, and Other which included those of Asian and Middle Eastern race and/or Hispanic ethnicity, pack-years of tobacco were included as a continuous measure, and number of lifetime oral sex partners was in ordinal categories (0,1, 2–5,6–10, 11–15,16–25, 26–50, 51–100, >100)