Modern Pathology (2017) 30, 1194–1203

p16 immunohistochemistry in oropharyngeal squamous cell carcinoma: a comparison of antibody clones using patient outcomes and high-risk human papillomavirus RNA status

Jeremy Shelton[1], Bibianna M Purgina[2], Nicole A Cipriani[3], William D Dupont[4], Dale Plummer[4] and James S Lewis Jr.[1,5]

[1]Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA; [2]Department of Pathology and Laboratory Medicine, University of Ottawa, Ottawa, ON, Canada; [3]Department of Pathology, University of Chicago Medicine, Chicago, IL, USA; [4]Department of Biostatistics and Preventive Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA and [5]Department of Otolaryngology, Vanderbilt University Medical Center, Nashville, TN, USA

High-risk human papillomavirus (HPV)-related oropharyngeal squamous cell carcinomas have a more favorable prognosis than HPV-negative ones. p16 immunohistochemistry has been recommended as a prognostic test in clinical practice. Several p16 antibodies are available, and their performance has not been directly compared. We evaluated three commercially available p16 antibody clones (E6H4, JC8 and G175-405) utilizing 199 cases of oropharyngeal squamous cell carcinoma from a tissue microarray, read by three pathologists with three different cutoffs for positivity: any staining, >50% and >75%. Positive predictive values for high-risk HPV status by RNA *in situ* hybridization for the E6H4, JC8 and G175-405 clones were 98%, 100% and 99% at the 75% cutoff, but negative predictive values were much more variable at 86%, 69% and 56%, respectively. These improved using the 50% cutoff, becoming similar for all three antibodies. Intensity varied substantially, with 85% of E6H4, 72% of JC8 and 67% of G175-405 showing strong (3+) intensity. With Kaplan–Meier survival plots at the 75% cutoff, the E6H4 clone showed the largest differential in disease specific and overall survival between p16-positive and -negative results. Decreasing the cutoff to 50% increased correlation with HPV *in situ* hybridization and improved the survival differential for the JC8 and G175-405 clones without worsening of performance for the E6H4 clone. Interobserver agreement was also assessed by kappa scores and was highest for the E6H4 clone. Overall, these study results show modest but important performance differences between the three different p16 antibody clones, suggesting that the E6H4 clone performs best because of strongest staining intensity, greatest differential in outcomes between positive and negative results, lowest interobserver variability, and lowest background, nonspecific staining. The results also suggest that a 75% cutoff is very functional but that, in this patient population with high HPV incidence, 50% and any staining cutoffs may be more effective, particularly for the non-E6H4 clones.
Modern Pathology (2017) **30,** 1194–1203; doi:10.1038/modpathol.2017.31; published online 16 June 2017

High-risk human papillomavirus (HPV)-related oropharyngeal squamous cell carcinoma is now well established as a unique entity with distinct epidemiology, biology and prognosis.[1–4] HPV-related oropharyngeal squamous cell carcinoma shows a predilection for younger, Caucasian males with higher numbers of sexual partners and lower or no tobacco use, in contrast to conventional head and neck squamous cell carcinoma patients. HPV-16, a known oncogenic virus in other body sites, tends to be the most commonly identified HPV type, with other high-risk types (18, 31, 33 and rarely others) causative in only 5–10% of patients.[1,5] Tumors almost always originate from the base of tongue and palatine tonsils, most patients have nodal metastasis at presentation,[1,6–8] and most have a

Correspondence: Dr JS Lewis Jr, MD, Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Room 3020D Vanderbilt University Hospital, Nashville, TN 37232-7415, USA.
E-mail: James.lewis@vanderbilt.edu
Received 17 December 2016; revised 2 March 2017; accepted 12 March 2017; published online 16 June 2017

distinctive, non-keratinizing morphology. Despite these seemingly aggressive characteristics, HPV-related oropharyngeal squamous cell carcinoma patients have consistently been shown to have a more favorable prognosis than those who are HPV negative.[3,4,8,9] Tumorigenesis is primarily thought to be driven by the HPV E6 and E7 proteins, which decrease levels of p53 and of functional Rb tumor-suppressor protein leading to aberrant overexpression of the cell cycle protein p16INK4a.[5–7] The resulting marked increase in p16 makes it a rational target for immunohistochemistry as a prognostic marker and surrogate marker for transcriptionally active high-risk HPV.[3]

p16 immunohistochemistry has become the recommended standalone prognostic test for patients with oropharyngeal squamous cell carcinoma as it is more cost-effective and less technically cumbersome than HPV-specific testing (ie, *in situ* hybridization and reverse-transcriptase PCR), is widely available and has high interobserver agreement in its assessment,[10] and has been consistently shown to have independent prognostic significance.[8] Under the new 8th edition American Joint Commission on Cancer (AJCC) staging guidelines, p16 immunohistochemistry as a standalone test is now required in order to stage oropharyngeal squamous cell carcinoma and cervical metastatic squamous cell carcinoma of unknown primary patients,[11] as p16-positive ones (as a surrogate marker for high-risk HPV status) have their own separate staging systems. Further, a College of American Pathologists evidence-based guidelines committee on this subject now recommends p16 IHC in all patients with newly diagnosed oropharyngeal squamous cell carcinoma in routine clinical practice.

Combined nuclear and cytoplasmic immunoreactivity in $>70\%$ of tumor cells is a commonly utilized cutoff for immunohistochemistry positivity based on large studies using this cutoff, and helping to exclude the significant fraction of oropharyngeal squamous cell carcinoma and non-oropharyngeal squamous cell carcinoma that may have partial p16 expression that does not signify the favorable biology of an HPV-related oropharyngeal squamous cell carcinoma.[1,6] This partial staining is a realistic problem occasionally encountered in clinical practice.[12–14] As p16 testing is becoming a recommended standard for oropharyngeal squamous cell carcinoma, more technical and method-based studies are needed to identify the best and most reproducible test for use in clinical practice, and as groundwork for anticipated laboratory proficiency testing. There are a number of commercially available immunohistochemical assays (antibody clones) for p16, and, although they largely have performed well at prognostication and patient risk stratification in large studies, there are some suggestions of variable performance characteristics.[12,14] As there is, to our knowledge, no study directly comparing performance of different p16 antibodies in oropharyngeal squamous cell carcinoma, and because patient outcomes are the standard by which they should be evaluated, we performed a comparative study of three different commercially available p16 antibodies in a large tissue microarray cohort of patients with robust clinical follow-up information.

Materials and methods

A tissue microarray of retrospectively identified oropharyngeal squamous cell carcinoma patients from 1998 to 2007 was created as previously described.[15,16] According to the amount of available biopsied or resected tumor tissue, duplicate 2 mm punches (or if inadequate tumor tissue present, then 0.6 mm punches) were taken from each case. Approximately 75% of cases on the array had the larger (2 mm) punches. In all, 243 cases were evaluated in total, 44 of which were excluded because of missing tumor ('core loss') or insufficient evaluable tumor, leaving a total of 199 cases. Patients were treated without regard to their p16 status.[17]

p16 Immunohistochemistry

Immunohistochemistry was performed for p16 on formalin-fixed, paraffin-embedded tissue sections. Three different and individually optimized p16INK4a antibody clones were utilized (Table 1). The Bond Polymer Refine detection system was used for visualization. Slides were then dehydrated, cleared and coverslipped. Staining was interpreted by three study pathologists (JSL, BP and NC). Cases were considered suitable/sufficient for interpretation if at least 10% of the cross-sectional area across the two cores consisted of tumor cells. Both nuclear and cytoplasmic staining was required for a cell to be considered 'positive' and staining distribution was read in quartiles as 1–24% (1), 25–49% (2), 51–75% (3) and 76–100% (4). Intensity of staining was defined as: weak = 1, moderate = 2 and strong = 3. For simpler characterization, results were dichotomized in three separate ways: (A) no staining *vs* any staining of tumor cells at all ('any staining cutoff'); (B) no staining or staining in $<50\%$ of tumor cells *vs* in $>50\%$ ('50% cutoff'); (C) no staining or staining in $<75\%$ of tumor cells *vs* $>75\%$ ('75% cutoff'). For the E6H4 clone, H-scores were generated by one study pathologist (JSL) by assessing the exact fraction (in 5% increments) of tumor cells with nuclear and cytoplasmic staining, which was then multiplied by the intensity of staining (1, 2 or 3) to generate the score. As was previously suggested by Jordan *et al*, in their work, a cutoff of 60 was used to dichotomize patients into positive and negative results.[10]

RNA *In Situ* Hybridization for High-Risk HPV

In situ hybridization for high-risk HPV E6/E7 mRNA had been performed and interpreted as previously

Table 1 Antibody clones and details of immunohistochemical staining methods

Clone	*Manufacturer*	*Dilution*	*Automated staining instrument*	*Antigen retrieval*
E6H4	CINtec Histology, Ventana, Tucson, AZ, USA	Predilute	Leica Bond-Max	Epitope retrieval 2 solution for 20 min at pH 9.5
JC8	Santa Cruz Biotech., Santa Cruz, CA, USA	1:10	Leica Bond-Max	Epitope retrieval 1 solution for 20 min at pH 6.0
G175-405	BD Biosciences, Franklin Lakes, NJ, USA	1:10	Leica Bond-Max	Epitope retrieval 2 solution for 30 min at pH 9.5

described[16] by hand using the RNAscope HPV kit (Advanced Cell Diagnostics, Inc, Hayward, CA, USA) according to the manufacturer's instructions. Probes covered HPV types 16, 18, 31, 33, 35, 52 and 58. Positive staining was identified as brown, punctate dots present in the nucleus and/or cytoplasm. Control probes for the bacterial gene DapB (negative control) and for the housekeeping gene ubiquitin C (positive control evidence of adequate RNA) were also included on each case. Cases in which there was $<10\%$ surface area consisting of tumor were excluded. The array (and corresponding control) slides and were classified in a binary manner as either positive or negative. Positive cases had to have granular cytoplasmic and/or nuclear brown staining that was higher than the signal on the DapB-negative control slide.

Statistical Analysis

We assessed differences in disease-specific survival and overall survival in patient groups defined by immunohistochemistry findings in tumor tissue obtained at the time of initial diagnosis and treatment. Patients were considered to have died of their disease if their cancer had recurred and they died with known disease in their bodies. In the disease-specific survival analyses, patients who died without disease were censored at their time of death. Kaplan–Meier survival plots were used to estimate differences in survival between these groups. The significance of these differences was assessed with the log-rank test. Cox proportional hazards regression analysis was used to estimate the relative mortal hazard in patients with positive *vs* negative immunohistochemistry results. These relative risk estimates were adjusted for age at diagnosis.

Results

Rates of positive and negative results did not vary greatly between the three different cutoffs of 75%, 50% and 'any staining'. At the 75% cutoff, there were 161 positive (81%) and 38 negative (19%) cases with the E6H4 clone, 148 positive (74%) and 51 negative (26%) with JC8 and 135 positive (68%) and 64 negative (32%) with G175-405. At this cutoff, there were 165 concordant (83%) and 34 discrepant (17%) cases. Reduced intensity staining, defined as $<3+$, was seen more frequently with the G175-405 (33%) and JC8 (28%) clones as compared with E6H4 (15%) (Figures 1a-c). When reducing to the 50% cutoff, the results for the three clones were much more similar. Results for E6H4 changed minimally with 162 positive (81%) and 37 negative (19%) results, whereas the JC8 and G175-405 demonstrated higher numbers of positive and fewer negative patients, each clone showing 158 positive (79%) and 41 negative (21%) results. When reducing to the any staining cutoff, a small amount more patients were positive *vs* negative with each antibody. Although the numbers of positive results and negative results increased and decreased with lesser percentages of cells required for positivity for each antibody, the differences were least for the E6H4 antibody, slightly higher for the JC8 antibody, and highest for G175-405, demonstrating that the E6H4 clone has the least amount of partial reactivity. Nonspecific staining, defined as cases with any type of cytoplasmic only or nuclear only staining (in the absence of cells with nuclear and cytoplasmic staining), was present in 12% of all cases evaluated with the G175-405 clone and only 1% of all cases evaluated with both the E6H4 and JC8 clones.

HPV RNA *in situ* hybridization results were available for 195 of the E6H4-stained cases and 191 of both the JC8 and G175-405-stained cases. When p16 immunohistochemistry was correlated with HPV RNA *in situ* hybridization results (Table 2), sensitivity, specificity and positive predictive values were similar amongst all clones at all three cutoffs, although specificity was somewhat lower for the E6H4 clone at the 50 and 75% cutoffs. Negative predictive values were significantly lower for the JC8 and G175-405 clones at the 75% cutoff, but increased substantially at the 50% and any staining cutoffs. H-scores were calculated for the E6H4 clone only. When correlating with HPV RNA *in situ* hybridization status, H-score results were essentially identical to those of both the 50 and 75% cutoffs and slightly superior to the any staining cutoff (Table 3).

For the E6H4 antibody, only two cases (1%) were negative at the 75% cutoff but positive with the any staining cutoff, meaning that there is essentially no

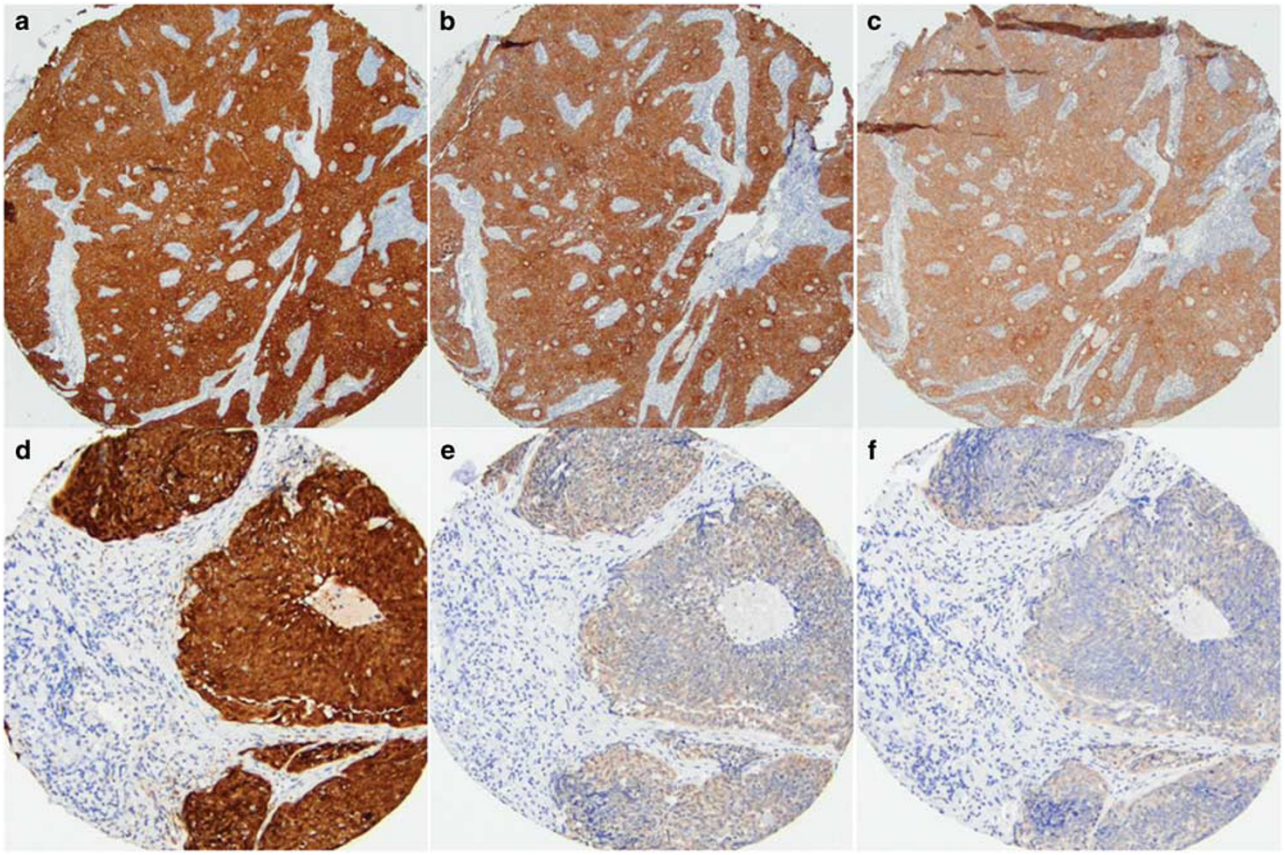

Figure 1 Examples of immunostaining with the different antibodies. (**a–c**) Concordant positive p16 immunohistochemical results (**a** = E6H4, **b** = JC8, **c** = G175-405). All show diffuse nuclear and cytoplasmic staining, but the intensity of staining is best with the E6H4 clone. (**d–f**) Discordant p16 immunohistochemical results (**d** = E6H4, **e** = JC8, **f** = G175-405). Staining would be positive for all three antibodies using the 'any staining' cutoff, but would only be positive for the E6H4 clone using the 50 and 75% cutoffs (all images × 20 magnification).

functional difference for patients which cutoff was utilized. However, rates were much higher for JC8 (17 cases or 8%) and G175-405 (32 cases or 16%), and HPV RNA *in situ* hybridization (in available cases) was positive in 50% (1/2) of the discrepant E6H4, 80% (12/15) of JC8 and 79% (23/29) of G175-405 cases.

p16 immunohistochemistry results were correlated with patient survival for all three clones at the various cutoffs. Follow-up for study subjects varied from 0.06 to 12.52 years with a median follow-up of 3.35 years. By simple disease recurrence (Table 4), performance of the different clones was very similar for each of the three antibody clones at each of the cutoffs. Although results were similar for all antibodies at all cutoffs, the differential between the fraction of patients with disease recurrence in the p16-positive *vs* -negative cohorts was widest (ie, best stratification) for the E6H4 clone. Interestingly, the differential was widest for the any staining cutoff for the E6H4 and JC8 antibodies, although not for G175-405.

There was a marked association between p16 immunohistochemistry status for all antibodies and survival regardless of the cutoff level. Figures 2 and 3 show the cumulative overall and disease-specific mortality associated with the three antibodies at all three cutoffs (HPV RNA *in situ* hybridization cumulative overall and disease-specific survival curves are provided as Supplementary Figure 1). The hazard ratios for death and for death from disease varied little between the E6H4 and JC8 clones, but were slightly lower (better) for the E6H4 clone. Hazard ratios were modestly higher for the G175-405 clone (Table 5). When adjusted for age, hazard ratios were nearly identical to crude results, meaning that p16 results were independent of patient age. Interestingly, hazard ratios were lowest (best) for the any staining cutoff for the E6H4 and JC8 antibodies, suggesting, from a statistical standpoint, that this cutoff would be the best choice for interpreting p16 immunohistochemistry as a prognostic test in routine clinical practice.

p16 immunohistochemistry was also subsequently reviewed by two additional pathologists (NC and BP) in order to assess interobserver variability (Table 6). With the any staining cutoff, there was perfect agreement among all pathologists. For the 75% cutoff, individual and combined agreement was only modest to good for each antibody. The numbers

Table 2 p16 immunohistochemistry performance correlated with HPV mRNA *in situ* hybridization by the various clones and cutoffs

Clone	*PPV*	*NPV*	*Sens*	*Spec*
	75% Cutoff			
E6H4	98% (156/159)	86% (31/36)	97% (156/161)	91% (31/34)
JC8	100% (142/142)	69% (34/49)	90% (142/157)	100% (34/34)
G175-405	99% (129/130)	56% (34/61)	83% (129/156)	97% (34/35)
	50% Cutoff			
E6H4	98% (157/160)	89% (31/35)	98% (157/161)	91% (31/34)
JC8	99% (151/152)	85% (33/39)	96% (151/157)	97% (33/34)
G175-405	99% (148/150)	80% (33/41)	95% (148/156)	94% (33/35)
	Any staining			
E6H4	98% (157/161)	88% (30/34)	98% (157/161)	88% (30/34)
JC8	97% (153/157)	91% (31/34)	98% (154/157)	91% (31/34)
G175-405	96% (153/159)	91% (29/32)	98% (153/156)	83% (29/35)

Abbreviations: HPV, human papillomavirus; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

Table 3 Comparison of H-scores to the different percentage cutoffs using the E6H4 antibody

	PPV	*NPV*	*Sens*	*Spec*
H-score	98% (157/160)	89% (31/35)	98% (157/160)	91% (31/34)
75% Cutoff	98% (156/159)	86% (31/36)	97% (156/161)	91% (31/34)
50% Cutoff	98% (157/160)	89% (31/35)	98% (157/161)	91% (31/34)
Any staining	98% (157/161)	88% (30/34)	98% (157/161)	88% (30/34)

Abbreviations: HPV, human papillomavirus; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

Table 4 Simple, binary disease recurrence rates based on the p16 immunohistochemistry results for the different antibody clones and cutoffs

Clone			***Differential***
	75% Cutoff—positive	*75% Cutoff—negative*	
E6H4	26/170 (15.3%)	14/39 (35.9%)	20.6%
JC8	22/152 (14.5%)	19/53 (35.8%)	21.3%
G175-405	18/137 (13.1%)	23/67 (34.3%)	21.2%
	50% Cutoff—positive	*50% Cutoff—negative*	
E6H4	26/171 (15.2%)	14/38 (36.8%)	21.6%
JC8	25/163 (15.3%)	16/42 (38.1%)	22.8%
G175-405	26/160 (16.3%)	15/44 (34.1%)	17.8%
	Any staining—positive	*Any staining—positive*	
E6H4	26/172 (15.1%)	14/37 (37.8%)	22.7%
JC8	27/169 (16.0%)	14/36 (38.9%)	22.9%
G175-405	30/169 (17.8%)	11/35 (31.4%)	13.6%

dropped substantially, from 1.0 to as low as 0.42, despite only few disagreements, because the results already are binary (only two class outcomes) and a large degree of agreement is already expected. Agreement at the 75% cutoff was still best, although, for the E6H4 clone. Agreement improved for all antibody clones when using the 50% cutoff so that all three were essentially comparable.

Discussion

p16 immunohistochemistry has become a critical test for prognostication for patients with oropharyngeal squamous cell carcinoma.[1,2] Results are now relied upon by clinicians for counseling and patient education purposes regarding the etiology of their cancer.[4,11] Results are used to help guide initial clinical work up of patients and treatment decisions within the standard of care guidelines. Clinical trials are underway that will better define specific clinical management for these patients, distinct from the current standard oropharyngeal squamous cell carcinoma recommendations and approaches.[3,4,18] For all of these reasons, HPV/p16 testing needs to be standardized across practices. For p16 immunohistochemistry, an

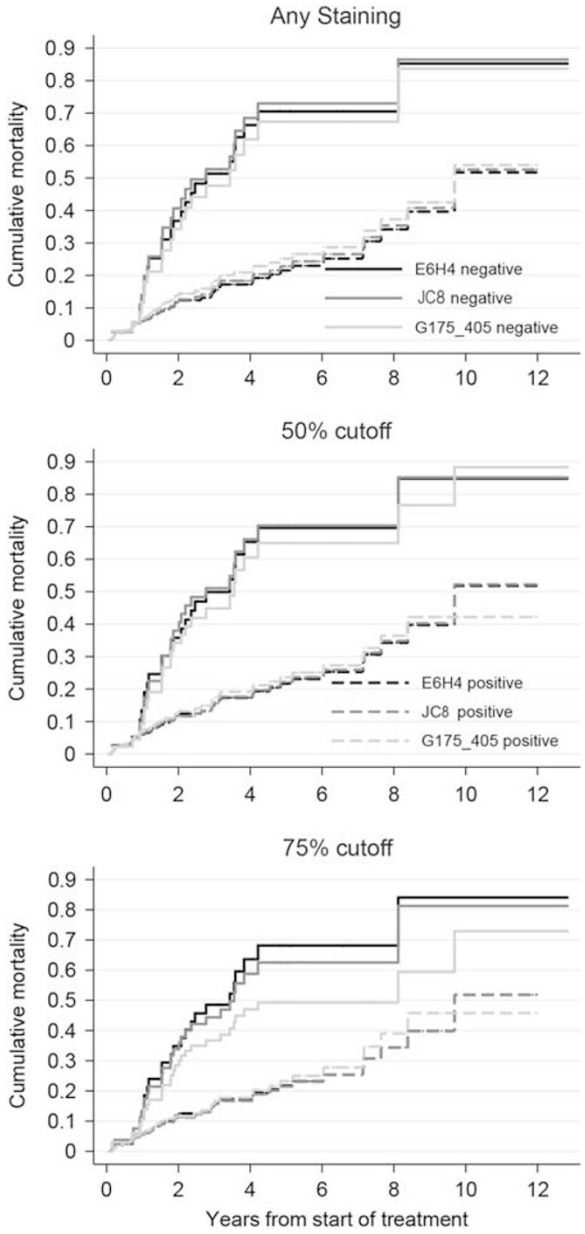

Figure 2 Kaplan–Meier curves for overall survival for the three antibody clones and at the three different cutoffs.

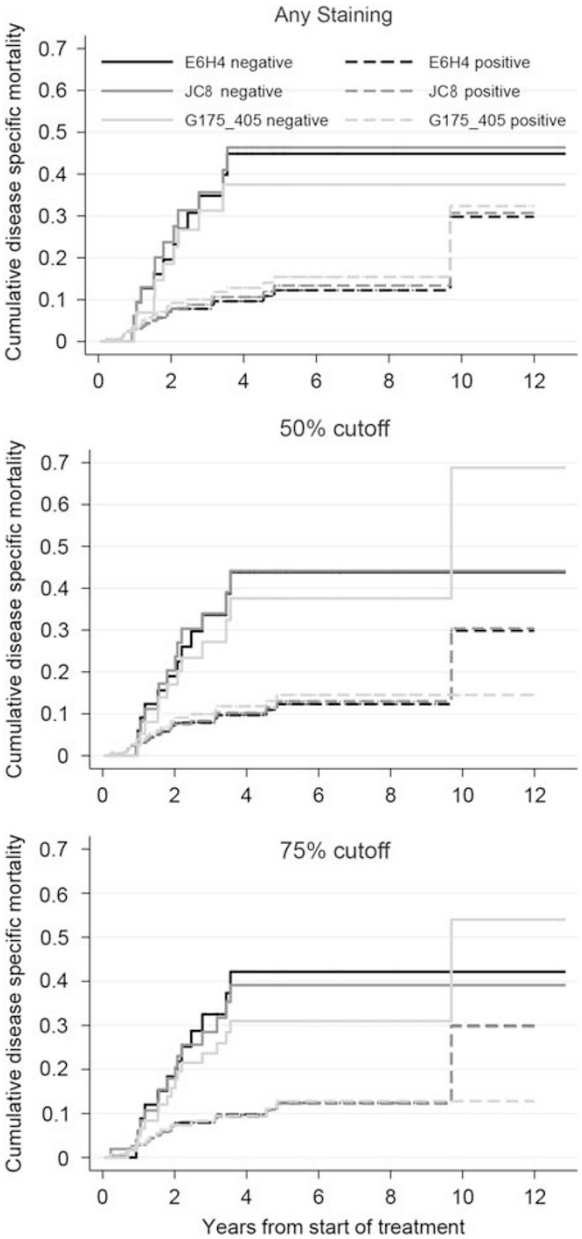

Figure 3 Kaplan–Meier curves for disease-specific survival for the three antibody clones and at the three different cutoffs.

antibody of choice is needed, and testing platforms and methods need to be more clearly defined. In light of the changing AJCC/UICC guidelines, it is also expected that proficiency testing may soon follow.

Although correlation with HPV mRNA status is one way to specifically evaluate the performance of different p16 antibodies and test platforms (many call HPV mRNA detection the 'gold standard' test[10]), our opinion is that patient outcomes are the ultimate standard by which these methods should be judged.[2,3,15,17] The best test or tests are the ones that most widely differentiate favorable and unfavorable survival rates.

Although there are, as of yet, no head to head, direct antibody comparisons in oropharyngeal squamous cell carcinoma, the results of several studies with different antibody clones and platforms has suggested differing performance characteristics.[12,14] Although most E6H4-based studies have shown low rates of partial or equivocal staining,[5] and low background nonspecific nuclear or cytoplasmic only staining,[10,15] G175-405 clone-based studies have shown significant rates of weaker and more frequent partial staining.[12,14] Interestingly, this has led to several studies attempting to address how to clarify equivocal results with the latter antibody.[12,14] These have ranged from trying to define specific staining

Table 5 Regression analysis hazard ratios for survival outcomes for the three antibodies at the three respective cutoffs

	Any staining	*50% Cutoff*	*75% Cutoff*
Overall survival			
E6H4	0.240 (0.143–0.403)	0.248 (0.148–0.418)	0.258 (0.153–0.434)
JC8	0.237 (0.141–0.399)	0.245 (0.147–0.410)	0.275 (0.165–0.457)
G175-405	0.314 (0.183–0.538)	0.306 (0.183–0.511)	0.438 (0.266–0.722)
Disease-specific survival			
E6H4	0.239 (0.114–0.503)	0.248 (0.118–0.523)	0.260 (0.124–0.547)
JC8	0.244 (0.117–0.511)	0.256 (0.123–0.530)	0.271 (0.132–0.557)
G175-405	0.375 (0.171–0.823)	0.334 (0.160–0.699)	0.341 (0.165–0.704)

Table 6 Interobserver agreement for the three study pathologists for the various antibody clones and cutoffs

Clone and staining cutoff	*Reviewer 1 vs 2*		*Reviewer 1 vs 3*		*Reviewer 2 vs 3*		*Combined*
	Agreement (%)	*Kappa*	*Agreement (%)*	*Kappa*	*Agreement (%)*	*Kappa*	*Kappa*
E6H4 75%	100	1.00	91.30	0.47	91.30	0.47	0.57
E6H4 50%	100	1.00	100.00	1.00	100.00	1.00	1.00
E6H4 any	100	1.00	100.00	1.00	100.00	1.00	1.00
JC8 75%	100	1.00	78.26	0.18	78.26	0.18	0.35
JC8 50%	95.7	0.65	100.00	1.00	95.65	0.65	0.73
JC8 any	100	1.00	100.00	1.00	100.00	1.00	1.00
G175-405 75%	82.6	0.28	86.96	0.59	86.96	0.36	0.42
G175-405 50%	100	1.00	100.00	1.00	100.00	1.00	1.00
G175-405 any	100	1.00	100.00	1.00	100.00	1.00	1.00

patterns such as confluence *vs* not or intensity of reactivity, to combining HPV-specific testing with the p16 results.[3,7,12,14]

In this study, we compared three different antibody clones in a large, well-characterized oropharyngeal squamous cell carcinoma patient cohort. Results showed that the E6H4 clone provided the most consistent and intense staining with the lowest rates of partial staining and lowest nonspecific background reactivity (Figure 1). Nonspecific staining was a significant issue for the G175-405 clone only. The JC8 clone performed well, similar to the E6H4 clone, although intensity of staining was lower and the fraction of cases with partial staining was higher.

Correlation with high-risk HPV RNA *in situ* hybridization results showed significant false-negative rates (low negative predictive values or NPV) for both the JC8 (NPV = 69%) and G175-405 (NPV = 56%) clones at the 75% cutoff. This improved substantially for both clones with the 50% cutoff and even more so with the any staining cutoff. These findings are not surprising given the higher rates of partial and low intensity staining that were observed with these antibodies. Some have suggested H-scores as better for interpreting p16 immunohistochemical results.[10] We calculated these for the E6H4 clone and used the cutoff of 60 that Jordan *et al*[10] proposed based on ROC curve analysis in their work. However, in our study, H-scores showed no benefit over percentage-based cutoffs (Table 3) with almost identical performance to all three cutoffs. Evaluating H-scores for the JC8 and G175-405 may prove beneficial as these clones are more prone to partial and weak intensity staining, although this was not evaluated for these two antibodies.

Interobserver variability in the assessment of p16 staining is also important for consistent results across clinical practices and settings. This has previously been at least partially evaluated. Jordan *et al*.[10] demonstrated kappa values of >0.90 amongst three pathologists reviewing p16 immunohistochemistry results in 231 oropharyngeal squamous cell carcinoma patients, but using only their H-score with a cutoff of 60 (combination of intensity × distribution on scale of 0 to 300). This actually is a low threshold for p16 positivity and is not agreed upon, used, or recommended by most pathologists and organizations for testing. A cutoff of 70–75% is the most often utilized and recommended in practice based on large studies validating it as highly prognostic. We found interobserver agreement to be very high (not surprisingly) for the any staining cutoff (Table 3). There was 100% agreement across all pathologists and specimens. It was worst for the 75% cutoff, but still was what could be considered good (or at least modest) agreement, and was still best for the E6H4 antibody.

The antibodies are ultimately best judged based on patient outcomes, particularly disease-specific survival and the differential between rates between positive and negative results. The results show that

E6H4 provided equivalent or better performance than the other two antibodies, with equivalent or better hazard ratios for overall and disease-specific survival (Table 5 and Figures 2 and 3). An important second question is 'what cutoff should be used for a positive result?' Most publications, committee proceedings and review articles recommend a high threshold for percentage of positive cells, usually 70 or 75%.[1,11,17,19,20] The good news is that the E6H4 clone, which is the most commonly utilized antibody in clinical practices, provided the best prognostication and performance at this cutoff (Tables 4 and 5; Figures 2 and 3). The bad news, although, was that we found in the current study that the any staining cutoff was the best criterion for prognostication, with the widest survival differences, lowest hazard ratios for death and with 100% agreement between reviewers for all three antibodies. So why did we find that an any staining cutoff was equivalent to, or better than, 75%? This is because most (between 50 and 80%) of the patients with these partial p16 results have high-risk HPV-related tumors. Interestingly, many of the largest studies using p16 as definitional have simply used 'strong' staining or have reported 'diffuse, strong' in almost all cases, without a percentage cutoff.[6,7,21–23] This approach is supported by our results and is probably a function of the high overall HPV-positive patient fraction in our study population. The lack of improvement of results using the H-score is also not surprising as it is functionally like a low percentage cutoff. Most of the partial staining patients were high-risk HPV positive. In this population, the cutoff used for p16 positivity seemed to make little difference.

Despite our findings, a high percentage cutoff may still be better for patient care. Even if most patients with partial staining have HPV-related tumors, lowering the threshold for positivity will increase the likelihood of a false-positive result, labeling some additional patients as HPV-positive when they are not. Above the 70–75% cutoff, virtually all patients are HPV RNA positive, whereas it is 50–80% of the any staining and H-score positive patients. In addition, p16 immunohistochemistry is, and will be, used around the world in patient populations where HPV rates are much lower. The latter means there will be much more 'nonspecific' p16 expression that is not related to HPV and, as such, the PPV for p16 expression will be lower.

All of the current results would seem to support a practical approach that has been suggested for p16 immunohistochemistry for dealing appropriately with equivocal results.[13] It suggests that results could be dichotomized as $<50\%$ nuclear and cytoplasmic staining (of any intensity) = negative and >70–75% nuclear and cytoplasmic staining = positive. For those patients with 50–75% nuclear and cytoplasmic staining, however, regardless of intensity or confluence, one should reflex to an HPV-specific test (ideally mRNA based, such as RNA *in situ* hybridization, or at least to DNA *in situ* hybridization or PCR, as the arbitrating result. If HPV positive, one then considers the patient to be HPV-positive/related and thus in the good prognosis and biology oropharyngeal squamous cell carcinoma group.

This study has a few limitations that bear mentioning. It is TMA-based, so only consists of small punches of tumor. Over 80% of these consisted of 2 mm punches and all were duplicate. This means that there is probably the equivalent amount of tumor present to many routine clinical practice oropharyngeal biopsies and to neck needle core biopsy specimens. Although there may be concern for lack of representativeness of the overall tumor by these small samples, they were taken at random from each donor tumor block, the results on this microarray have shown very high prognostic value for p16 and HPV mRNA status in this and other studies,[15] and, most compelling of all, is that small biopsies have been clearly proven as totally functional for p16 testing via the large number of literature studies on patients treated with primary chemoradiation after initial diagnosis on such specimens.[1,6,20,23,24] In addition, Ma *et al.*[25] specifically showed that small biopsy specimens reliably indicate p16 status in comparison with surgical resection specimens. This study is retrospective and thus has variation in how patients were managed, the length of follow-up, and who was censored, and therefore it may have unperceived selection biases. It is certainly possible that differences in staining intensity and nonspecific staining between the antibodies may be partially because of staining conditions. We did not attempt to modify the staining conditions to alter performance with the various antibodies in this work. However, each antibody clone was individually optimized to provide the best staining before use in this study. The E6H4 clone comes 'ready-to-use' with no dilution step being required, making it the simplest to perform.

Also, an evaluation of testing methods would be incomplete without mentioning cost comparison. In the current medical economic climate, cost savings are increasingly important. There is potentially a wide cost differential between antibody clones, and the temptation to pursue the least expensive available option. E6H4 performs best here, but it is actually the most expensive of the three antibodies, per patient.[3] However, determination of cost effectiveness is not entirely straightforward given that partial (or equivocal) staining that may occur more frequently with a suboptimal p16 immunohistochemistry test may trigger repeat testing and/or may necessitate expensive follow-up HPV-specific testing. Further, any small fraction of false-negative or -positive results is significant. Patients may be classified incorrectly and thus, may be treated in a manner that is not appropriate for their cancer's biology, or may suffer from either progressive disease or treatment-related morbidity when these

could have been avoided. Cost is important, but it has to be carefully considered as the impact of equivocal or incorrect diagnoses is so significant.

In summary, this evaluation of three commercially available p16 antibodies for performance in prognostication and correlation with high-risk HPV mRNA status in oropharyngeal squamous cell carcinoma shows that all three perform well as prognostic markers and surrogate markers of HPV status. However, there are modest, but potentially significant, differences. The E6H4 clone provided the cleanest and most intense staining, had equivalent or best correlation with high-risk HPV mRNA status and patient outcomes, and highest interobserver agreement. Although a 70–75% cutoff has been widely utilized for p16 immunohistochemistry in the studies defining it as a prognostic marker,[1] our results in a high HPV incidence US patient population suggest that results with the E6H4 antibody may be even better with either a 50% or even an any staining cutoff. If the JC8 and G175-405 antibodies are used, they appear to function best at a lower cutoff, such as 50% cutoff. One needs to be aware, although, particularly with the G175-405 antibody, that partial nuclear and cytoplasmic, as well as nonspecific nuclear or cytoplasmic only staining, are not uncommon. Future studies may focus on further refinement of the technical methods for these antibodies in order to optimize one specific antibody and platform for routine clinical use.

Acknowledgments

We would like to thank Donna M Posey for her wonderful assistance with clerical support for the various aspects of this study.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Ang KK, Harris J, Wheeler R, *et al.* Human papillomavirus and survival of patients with oropharyngeal cancer. N Engl J Med 2010;363:24–35.
2 Bhatia A, Burtness B. Human papillomavirus-associated oropharyngeal cancer: defining risk groups and clinical trials. J Clin Oncol 2015;33:3243–3250.
3 Lewis JS Jr. p16 immunohistochemistry as a standalone test for risk stratification in oropharyngeal squamous cell carcinoma. Head Neck Pathol 2012;6(Suppl 1): S75–S82.
4 Masterson L, Moualed D, Liu ZW, *et al.* De-escalation treatment protocols for human papillomavirus-associated oropharyngeal squamous cell carcinoma: a systematic review and meta-analysis of current clinical trials. Eur J Cancer 2014;50:2636–2648.
5 Gao G, Chernock RD, Gay HA, *et al.* A novel RT-PCR method for quantification of human papillomavirus transcripts in archived tissues and its application in oropharyngeal cancer prognosis. Int J Cancer 2013;132: 882–890.
6 Hong AM, Martin A, Armstrong BK, *et al.* Human papillomavirus modifies the prognostic significance of T stage and possibly N stage in tonsillar cancer. Ann Oncol 2013;24:215–219.
7 Nasman A, Attner P, Hammarstedt L, *et al.* Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? Int J Cancer 2009;125: 362–366.
8 O'Sullivan B, Huang SH, Su J, *et al.* Development and validation of a staging system for HPV-related oropharyngeal cancer by the International Collaboration on Oropharyngeal cancer Network for Staging (ICON-S): a multicentre cohort study. Lancet Oncol 2016;17: 440–451.
9 Gondim DD, Haynes W, Wang X, *et al.* Histologic typing in oropharyngeal squamous cell carcinoma: a 4-year prospective practice study with p16 and high-risk HPV mRNA testing correlation. Am J Surg Pathol 2016;40:1117–1124.
10 Jordan RC, Lingen MW, Perez-Ordonez B, *et al.* Validation of methods for oropharyngeal cancer HPV status determination in US cooperative group trials. Am J Surg Pathol 2012;36:945–954.
11 O'Sullivan B, Lydiatt WM, Haughey BH, *et al.* HPV-mediated (p16+) oropharyngeal cancer. In: Amin MB, Edge S, Greene F, *et al.* (eds). AJCC Cancer Staging Manual. Eighth edn.Springer Nature: Switzerland, 2016, p 113–121.
12 Barasch S, Mohindra P, Hennrick K, *et al.* Assessing p16 status of oropharyngeal squamous cell carcinoma by combined assessment of the number of cells stained and the confluence of p16 staining: a validation by clinical outcomes. Am J Surg Pathol 2016;40: 1261–1269.
13 Lewis JS Jr., Chernock RD, Ma XJ, *et al.* Partial p16 staining in oropharyngeal squamous cell carcinoma: extent and pattern correlate with human papillomavirus RNA status. Mod Pathol 2012;25:1212–1220.
14 Schlecht NF, Brandwein-Gensler M, Nuovo GJ, *et al.* A comparison of clinically utilized human papillomavirus detection methods in head and neck cancer. Mod Pathol 2011;24:1295–1305.
15 Ukpo OC, Flanagan JJ, Ma XJ, *et al.* High-risk human papillomavirus E6/E7 mRNA detection by a novel in situ hybridization assay strongly correlates with p16 expression and patient outcomes in oropharyngeal squamous cell carcinoma. Am J Surg Pathol 2011;35: 1343–1350.
16 Lewis JS Jr., Ukpo OC, Ma XJ, *et al.* Transcriptionally-active high-risk human papillomavirus is rare in oral cavity and laryngeal/hypopharyngeal squamous cell carcinomas–a tissue microarray study utilizing E6/E7 mRNA *in situ* hybridization. Histopathology 2012;60: 982–91.
17 Lewis JS Jr., Thorstad WL, Chernock RD, *et al.* p16 positive oropharyngeal squamous cell carcinoma:an entity with a favorable prognosis regardless of tumor HPV status. Am J Surg Pathol 2010;34: 1088–96.
18 Gillison ML, Restighini C. Anticipation of the impact of human papillomavirus on clinical decision making for the head and neck cancer patient. Hematol Oncol Clin North Am 2015;29:1045–60.

19 Fakhry C, Zhang Q, Nguyen-Tan PF, *et al.* Human papillomavirus and overall survival after progression of oropharyngeal squamous cell carcinoma. J Clin Oncol 2014;32:3365–73.
20 Schache AG, Liloglou T, Risk JM, *et al.* Validation of a novel diagnostic standard in HPV-positive oropharyngeal squamous cell carcinoma. Br J Cancer 2013;108: 1332–9.
21 Huang SH, Xu W, Waldron J, *et al.* Refining American Joint Committee on Cancer/Union for International Cancer Control TNM stage and prognostic groups for human papillomavirus-related oropharyngeal carcinomas. J Clin Oncol 2015;33:836–45.
22 Huang SH, Patel S, O'Sullivan B, *et al.* Longer survival in patients with human papillomavirus-related head and neck cancer after positive postradiation planned neck dissection. Head Neck 2015;37:946–52.
23 Rischin D, Young RJ, Fisher R, *et al.* Prognostic significance of p16INK4A and human papillomavirus in patients with oropharyngeal cancer treated on TROG 02.02 phase III trial. J Clin Oncol 2010;28: 4142–8.
24 Lin BM, Wang H, D'Souza G, *et al.* Long-term prognosis and risk factors among patients with HPV-associated oropharyngeal squamous cell carcinoma. Cancer 2013;119:3462–71.
25 Ma C, Lewis J Jr. Small biopsy specimens reliably indicate p16 expression status of oropharyngeal squamous cell carcinoma. Head Neck Pathol 2012;6: 208–15.

Supplementary Information accompanies the paper on Modern Pathology website (http://www.nature.com/modpathol)