Diagnostic accuracy of $p16^{INK4a}$ immunohistochemistry in oropharyngeal squamous cell carcinomas: a systematic review and meta-analysis

Reference:
Prigge Elena-Sophie, Arbyn Marc, von Doeberitz Magnus Knebel, Reuschenbach Miriam.- Diagnostic accuracy of $p16^{INK4a}$ immunohistochemistry in oropharyngeal squamous cell carcinomas: a systematic review and meta-analysis
Full text (Publisher’s DOI): http://dx.doi.org/doi:10.1002/IJC.30516
To cite this reference: http://hdl.handle.net/10067/1419870151162165141
Diagnostic accuracy of p16INK4a immunohistochemistry in oropharyngeal squamous cell carcinomas: a systematic review and meta-analysis

ES Prigge¹, M Arbyn²,³, M von Knebel Doeberitz¹, M Reuschenbach¹

Affiliations:
¹Department of Applied Tumor Biology, Institute of Pathology, University of Heidelberg, and Clinical Cooperation Unit Applied Tumor Biology, German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany
²Belgian Cancer Centre and Unit of Cancer Epidemiology, Scientific Institute of Public Health, B1050 Brussels, Belgium
³University of Antwerp, Antwerp, Belgium

Corresponding author:
Dr. Elena-Sophie Prigge
Department of Applied Tumor Biology, Institute of Pathology, University Hospital Heidelberg
Im Neuenheimer Feld 224, 69120 Heidelberg, Germany
Tel: +49-6221-5635594, Fax: +49-6221-565981
e-mail: elena.prigge@med.uni-heidelberg.de

Short title:
Meta-analysis on p16 immunohistochemistry in oropharyngeal cancer

Key words:
human papillomavirus, HPV, oropharyngeal squamous cell carcinoma, OPSCC, p16

List of abbreviations:
HNSCC: head and neck squamous cell carcinoma
HPV: human papillomavirus
IHC: immunohistochemistry
ISH: in-situ hybridization
OPSCC: oropharyngeal squamous cell carcinoma
PICOS: population, index test, comparator tests, outcomes, study design
RSe: relative sensitivity
RSp: relative specificity

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/ijc.30516

This article is protected by copyright. All rights reserved.
Article category
Tumor Markers and Signatures

Novelty and impact
Our systematic review and meta-analysis demonstrates that combined testing for p16$^{\text{INK4a}}$ overexpression on immunohistochemistry (IHC) and HPV DNA by PCR is highly accurate to identify HPV-induced oropharyngeal carcinomas. The test combination is as sensitive as either p16$^{\text{INK4a}}$ IHC or HPV DNA PCR alone but significantly more specific than either separate test. This observation may pave the way for a reliable diagnosis and subsequent therapy of affected patients, particularly in the context of de-escalated therapy.
The accurate diagnosis of human papillomavirus (HPV) causality in oropharyngeal squamous cell carcinomas (OPSCC) is likely to influence therapeutic decisions in affected patients in the near future. We conducted a systematic review and meta-analysis to determine the diagnostic accuracy of p16INK4a immunohistochemistry (IHC) to identify HPV-induced OPSCC. We identified all studies that performed p16INK4a IHC (index test) and HPV E6/E7 mRNA detection using an amplification-based method (gold standard to indicate a transforming relevance of HPV) in OPSCC. Testing with one or more comparator tests (HPV DNA PCR, HPV DNA in-situ hybridization (ISH) and p16INK4a IHC/HPV DNA PCR combined testing) was an optional criterion for inclusion. Among 1,636 retrieved studies 24 fulfilled the inclusion criteria. The pooled sensitivity of p16INK4a IHC, HPV DNA PCR, HPV DNA ISH and p16INK4a IHC/HPV DNA PCR combined testing was 94% (95%-confidence interval (CI) 91-97%), 98% (94-100%), 85% (76-92%) and 93% (87-97%), respectively. The pooled specificity was 83% (78-88%), 84% (74-92%), 88% (78-96%) and 96% (89-100%), respectively. p16INK4a IHC/HPV DNA PCR combined testing was as sensitive as either p16INK4a IHC or HPV DNA PCR alone but significantly more specific than either separate test. In conclusion, p16INK4a IHC is very sensitive but moderately specific to diagnose HPV-transformed OPSCC when used as a single test. Combined p16INK4a IHC and HPV DNA PCR testing significantly enhances specificity while maintaining high sensitivity. This diagnostic test combination thus represents an attractive testing strategy for the reliable diagnosis of HPV-induced OPSCC in the clinical setting and may constitute an inclusion criterion for future therapeutic trials.
INTRODUCTION

Human papillomavirus (HPV)-associated head and neck squamous cell carcinomas (HNSCC) represent an increasing health problem, particularly in the oropharyngeal tonsils. Current estimations imply that HPV-induced oropharyngeal squamous cell carcinomas (OPSCC) will surpass the number of incident cervical carcinoma cases in the United States by the year 2020. Rising incidence rates of HPV-induced HNSCC have been noted in several other Western countries. Patients with HPV-driven OPSCC generally demonstrate better survival compared to patients with HPV-negative OPSCC. Consequently, therapy de-escalation for patients with HPV-associated OPSCC is currently discussed and has given rise to the implementation of new clinical trials. Furthermore, HPV-specific therapeutics may become available in the near future. Specific treatment schemes for patients with HPV-induced OPSCC might alleviate therapy-related toxicity and reduce patient morbidity. At the same time, such a treatment regimen might undertreat patients with a tumor carrying HPV that did not induce carcinogenesis. Those considerations highlight the importance of reliable and accurate markers to diagnose truly HPV-induced OPSCC.

Various different methodical approaches are currently applied to identify HPV-induced HNSCC. The goal of all those approaches must be the identification of a biologically relevant, i.e. transforming, HPV infection. The detection of HPV oncogene E6/E7 transcripts is currently regarded as gold standard in this context, as HPV-driven carcinomas critically depend on the carcinogenic action of the HPV E6 and E7 oncogenes. However, the detection of viral transcripts is comparatively laborious and may not be routinely feasible in all laboratories. This is particularly true for transcript detection from formalin-fixed paraffin-embedded (FFPE) specimens, which are frequently used during routine diagnostic work up. In those FFPE samples RNA is often found fragmented or chemically modified requiring specific sample preparation techniques to recover RNA that is suitable for down-stream applications such as reverse transcription and qPCR. HPV oncogene transcript detection can be applied in cervical cancer screening, e.g. using the commercially available APTIMA® assay, where analysis is generally performed from cells sampled into a preservative liquid. However, cancer samples from the head and neck are frequently stored and processed as FFPE tissue posing a challenge to the detection of HPV transcripts due to the reduced quality of RNA in this material. Thus, a reliable and practicable testing strategy to identify HPV-driven HNSCC, particularly within the oropharynx, is urgently needed.

HPV E7 oncogenic signaling induces strong overexpression of the cellular protein p16INK4a in HPV-transformed cells. The immunohistochemical detection of p16INK4a overexpression is now used as a surrogate biomarker of HPV-transformed epithelium at the uterine cervix, particularly in the triage of
Women with positive cervical cancer screening results and to identify cervical precancer in biopsies. Similarly, it is frequently applied to determine HPV association in HNSCC. Apart from p16\(^{INK4a}\) immunohistochemistry (IHC) other HPV testing methods are commonly used in HNSCC. Among them, viral DNA detection by polymerase chain reaction (PCR) or in-situ hybridization (ISH) as well as combined detection of p16\(^{INK4a}\) IHC and HPV DNA detection by PCR, are frequently applied. The reported test accuracy to identify a transforming HPV infection differs substantially among studies. This may in part be influenced by different methodical procedures and heterogeneous definitions of a positive test result. Moreover, test performance may depend on the anatomical location of the tumor in oropharyngeal or non-oropharyngeal regions.

Our meta-analysis aimed at estimating the diagnostic accuracy of p16\(^{INK4a}\) immunohistochemistry to detect a transforming HPV infection in carcinomas of oropharyngeal origin where the majority of HPV-induced HNSCC arise. HPV-induced transformation was defined by the presence of viral E6/E7 oncogene transcripts using an amplification-based method. The pooled sensitivity and specificity of p16\(^{INK4a}\) IHC were evaluated. In addition, the accuracy of HPV DNA detection by PCR, HPV DNA detection by ISH and combined testing of p16\(^{INK4a}\) IHC and HPV DNA by PCR were assessed and compared.

**MATERIALS AND METHODS**

**Clinical question**

The meta-analysis aimed at estimating the clinical sensitivity and specificity of markers to detect transforming HPV infections in OPSCC. The following markers were evaluated: a) p16\(^{INK4a}\) IHC, b) HPV DNA detection by PCR, c) HPV DNA detection by ISH and d) combined testing with p16\(^{INK4a}\) IHC and HPV DNA detection by PCR. The PICOS (Population - Index test - Comparator tests - Outcomes - Study design) components of the clinical question are detailed in Figure S1A in line with the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)” guidelines.

**Population**

The population comprised men or women with oropharyngeal cancer of squamous epithelial origin. In line with the International Classification of Diseases-10 (ICD-10) code, OPSCC were classified with respect to their exact anatomic origin within the oropharyngeal region: palatine tonsil (C09), base of tongue/lingual tonsil (C01/C02.4), soft palate (C05.1), uvula (C05.2), Waldeyer’s ring (C14.2), other oropharyngeal subsites (C10.0-8) and oropharynx unspecified (C10.9). If authors did not report the exact ICD-10 codes the nominal indication of tumor localization was considered acceptable.

**Index and comparator tests**
p16\textsuperscript{INK4a} IHC was chosen as index test. Three other markers (HPV DNA detection by PCR, HPV DNA detection by ISH and combined testing with p16\textsuperscript{INK4a} IHC and HPV DNA PCR) were evaluated as comparator tests. This meta-analysis specifically investigated whether combined testing for p16\textsuperscript{INK4a} IHC and HPV DNA PCR would increase specificity while maintaining high sensitivity to detect HPV-transformed OPSCC as compared to the individual tests. The co-test was considered positive when both tests (p16\textsuperscript{INK4a} IHC and HPV DNA PCR) were positive and was considered negative when one or both tests were negative.

**Outcomes**

A transforming HPV infection was assumed if E6 and/or E7 mRNA was detected in the tumors, which is currently regarded as gold standard to indicate a transforming HPV infection in HNSCC. The meta-analysis aimed assessing the absolute accuracy (sensitivity and specificity) of the index and comparator tests as well as the relative accuracy of tests compared to other tests.

**Inclusion criteria**

The manuscripts of interest comprised retrospective or prospective studies that performed p16\textsuperscript{INK4a} IHC and analyzed HPV E6 and/or E7 mRNA expression by an amplification-based method in squamous cell carcinomas of the oropharynx. Additional testing with one or more comparator tests was an optional criterion for inclusion but a conditional criterion to compute relative accuracy measures.

**Search strategy**

The search term (Figure S1B) was phrased broadly and entered into PUBMED to identify relevant references. Additionally, a manual exploration of reference lists of selected reports was performed. No restrictions regarding language or publication dates were applied. Extraction of the data was performed by EP and checked by MR. The last search was run on January 8\textsuperscript{th} 2016.

**Study selection, data extraction and statistical analyses**

The applied exclusion criteria are shown in the PRISMA flow chart (Figure 1). If the required raw data could not be extracted from the manuscripts the authors were contacted. If the authors did not respond the study was excluded from the main or respective subgroup meta-analysis. The following study characteristics potentially influencing study outcomes were extracted: anatomical subsite within the oropharyngeal region, country of patient recruitment, tissue material used for the tests, definition of p16\textsuperscript{INK4a} IHC positivity, p16\textsuperscript{INK4a} antibody and method of HPV oncogene transcript detection (Table 1 and Tables S1 and S2). The quality of the included studies was evaluated using the QUADAS-2 checklist, a tool to assess the risk of bias and applicability of diagnostic accuracy studies\textsuperscript{27,28} (Table S3).
Pooled sensitivity and specificity were computed for the tests p16\textsuperscript{INK4a} IHC, HPV DNA PCR, HPV DNA ISH, and combined testing of p16\textsuperscript{INK4a} IHC and HPV DNA PCR and visualized by forest plots. Pooled absolute sensitivity and specificity of the four analyzed tests were estimated using a bivariate normal model for the logit transforms of sensitivity and specificity, taking into account the intrinsic correlation between false- and true-positive rates as well as the variability between studies \textsuperscript{29-31}.

If the HSROC regression analysis allowing for concomitant estimation of sensitivity and specificity did not converge, the sensitivity and specificity were pooled separately by METAPROP, which is a package in STATA, allowing for pooling of binomial data \textsuperscript{32}. Heterogeneity between the individual studies was analyzed using a chi\textsuperscript{2} test (Cochrane’s Q). The percentage of total variation across studies due to heterogeneity was assessed with the I\textsuperscript{2} index \textsuperscript{33}. The applied statistical tests were two-sided. A P-value of <0.05 was considered statistically significant. The statistical analyses were performed using STATA software (version 13.1). Relative sensitivity and specificity of one test compared to p16\textsuperscript{INK4a} IHC or compared to another comparator test were computed using the bivariate model including test as a co-variate \textsuperscript{31}. When less than four studies were available, ratios of the accuracy measures were pooled as ratios of proportions ignoring the correlation between sensitivity and specificity \textsuperscript{34}.

RESULTS

Identification of relevant manuscripts

1,636 articles were retrieved. The process of study selection is illustrated in the PRISMA flow-chart in Figure 1. Twenty-four manuscripts fulfilled all inclusion criteria and did not meet any exclusion criteria. Tables S1 and S2 give an overview of the general and technical characteristics of all included studies.

Accuracy of p16\textsuperscript{INK4a} IHC

Computation of the sensitivity and specificity of p16\textsuperscript{INK4a} IHC was feasible for 23 studies. For one study by Rietbergen and colleagues \textsuperscript{35} no individual p16\textsuperscript{INK4a} IHC data but only data for p16\textsuperscript{INK4a} IHC/HPV DNA PCR combined testing were available. The overall pooled sensitivity and specificity of p16\textsuperscript{INK4a} IHC among the included studies were 94% (95% confidence interval (CI) 91-97%) and 83% (95% CI 78-88%), respectively (Figure 2A; Table 2). A subgroup meta-analysis was conducted with respect to the application of the gold standard: HPV oncogene mRNA detection on all samples (general gold standard) or only on samples that had tested positive for HPV DNA in a preceding test (selective gold standard). There was no statistically significant heterogeneity between the two groups with respect to sensitivity (p=0.973; I\textsuperscript{2} = 57.67%) or specificity (p=0.101; I\textsuperscript{2} = 64.40%; Figure S2) of p16\textsuperscript{INK4a} IHC. This allowed grouping of all studies, irrespective of general or selective application of the gold standard test. The subgroup meta-analyses by
specific anatomic subsite (palatine tonsil, base of tongue or soft palate) revealed similar pooled sensitivity and specificity estimates (Figure 2B). Not all studies provided data at the level of subsites. Therefore, the overall pooled sensitivity and specificity differ slightly between the analyses for oropharynx, grouped (Figure 2A), and the analyses at the subsite level (Figure 2B).

Of note, the definition of a positive p16\textsuperscript{INK4a} IHC test result varied considerably among the included studies. Multiple definitions were applied (Table 1). ‘Strong and diffuse nuclear and cytoplasmic staining in ≥/>70% of the tumor’ was used in six studies,\textsuperscript{22, 36-40} the mere percentage of p16\textsuperscript{INK4a}-overexpressing cells was considered in six studies\textsuperscript{24, 41-45}, staining intensity in one study\textsuperscript{13}, and various combinations of those variables with or without consideration of the intracellular and intratumoral distribution were applied in eleven studies\textsuperscript{35, 46-55}. Two subgroup analyses were conducted to determine the influence of the definition of a positive p16\textsuperscript{INK4a} IHC test result on the sensitivity and specificity of this test. First, studies applying a definition of i) strong and ii) diffuse p16\textsuperscript{INK4a} expression in iii) ≥70% of tumor cells were compared to studies in which the definition for a positive test result did not fulfill all three criteria. The subgroup analysis revealed no significant differences between both groups with regard to sensitivity or specificity of p16\textsuperscript{INK4a} IHC to identify HPV-induced OPSCC (Fig. S3 A,B). In the second subgroup analysis studies applying a definition of ≥50% p16\textsuperscript{INK4a}-positive cells (irrespective of staining intensity, pattern or intracellular localization) were compared to studies using a definition of <50% p16\textsuperscript{INK4a}-positive cells or not considering the percentage of p16\textsuperscript{INK4a}-positive cells for a positive test result. No significant differences for sensitivity or specificity were found between both groups (Fig. S3 C,D). The significant heterogeneity of definitions for a positive p16\textsuperscript{INK4a} IHC test result precluded further subgroup analyses of the studies with regard to this variable at this point.

Accuracy of HPV DNA detection by PCR
The sensitivity of HPV DNA PCR to detect a transforming HPV infection in oropharyngeal tumors pooled from eleven eligible studies was 98% (95% CI 94-100%; Figure 3A; Table 2), whereas the pooled specificity was 84% (95% CI 74-92%; Figure 3A; Table 2). Subgroup meta-analyses by anatomical site were not feasible due to the small number of studies providing subsite data.

Accuracy of HPV DNA detection by ISH
The sensitivity and specificity of HPV DNA ISH pooled from eight studies were 85% (95% CI 76-92%) and 88% (95% CI 78-96%; Figure 3B; Table 2).

Accuracy of p16\textsuperscript{INK4a} IHC and HPV DNA PCR combined testing
Eleven studies were retrieved that assessed accuracy of combined testing for p16\textsuperscript{INK4a} IHC and HPV DNA by PCR. The pooled sensitivity and specificity were 93% (95% CI 87-97%) and 96% (95% CI 89-100%; Figure 3C; Table 2), respectively.

**Relative accuracy**

Eleven studies compared p16\textsuperscript{INK4a} IHC with HPV DNA PCR. The tests did not differ significantly from unity with respect to sensitivity or specificity: relative sensitivity (RSe) of HPV DNA PCR compared to p16\textsuperscript{INK4a} IHC: 1.01 (95% CI: 0.98-1.03), relative specificity (RSp): 1.02 (95% CI 0.91-1.14; Figure 4A; Table 3). Among ten included studies, the combination of p16\textsuperscript{INK4a} IHC and HPV DNA PCR was as sensitive as either p16\textsuperscript{INK4a} IHC or HPV DNA PCR alone (RSe: 0.97 (95% CI 0.93-1.02) and 0.97 (95% CI 0.93-1.02), respectively) but significantly more specific than either separate test alone (RSp: 1.13 (95% CI 1.04-1.23) and 1.07 (95% CI 1.01-1.13); Figure 4B and C; Table 3).

**Quality assessment**

The risk of bias and concerns of applicability as assessed by the QUADAS-2 tool categories was judged overall low in the categories “index and comparator test” and “reference standard” (Table S3). It must be emphasized that the methodical characteristics of the reference standard varied considerably between the different studies; however, all studies met the formal reference standard requirements, i.e. HPV E6 and/or E7 mRNA detection in head and neck tumor samples. Some concern existed with regard to patient selection in three-quarters of the included studies. This mainly resulted from unavailable information on patient/material exclusion criteria. The risk of bias and concerns on “flow and timing” was judged low in the vast majority of studies. A minority (6/24, 25.0%) of studies was judged with moderate risk of bias/concerns in this category due to missing information on uninterpretable results for index and comparator tests.

**DISCUSSION**

Several studies have demonstrated the improved survival of patients with HPV-associated compared to HPV-negative OPSCC patients\textsuperscript{6-10}. Consequently, therapy de-escalation for HPV-associated oropharyngeal cancer patients has moved into the clinical focus and has recently been implemented in phase I-III clinical trials\textsuperscript{11}. Furthermore, HPV-specific cancer treatment such as therapeutic vaccines may become available in the near future\textsuperscript{12}. This development highlights the need for a reliable and practicable testing strategy to identify truly HPV-driven OPSCC and preclude false-positive or false-negative test results.

*Clinical utility of combined p16\textsuperscript{INK4a} IHC/HPV DNA PCR testing strategy*
From a biological point of view the detection of HPV oncogene transcripts constitutes the current gold standard to identify an etiological role of HPV in HNSCC \(^{13, 14}\). However, considering the comparatively laborious methodology an alternative testing method appears worthwhile. Among those alternative tests HPV DNA detection by PCR or ISH as well as p16\(^{\text{INK4a}}\) IHC are most commonly applied in HNSCC.

p16\(^{\text{INK4a}}\) becomes overexpressed upon HPV E7 oncoprotein signaling via induction of the demethylase KDM6B that removes repressive histone 3 lysine 27 trimethyl (H3K27me3) marks from the p16\(^{\text{INK4a}}\)-encoding CDKN2A promoter region \(^{18}\). In non-HPV-transformed cells p16\(^{\text{INK4a}}\) overexpression naturally induces cellular senescence and irreversible cell cycle arrest \(^{56-58}\). However, in HPV-transformed cells E7 also inactivates the tumor suppressor protein RB, a downstream player of p16\(^{\text{INK4a}}\), which prevents the tumor suppressive function of p16\(^{\text{INK4a}}\) in those cells \(^{59, 60}\). Consequently, p16\(^{\text{INK4a}}\) overexpression functions as a direct surrogate marker of HPV-transformed cells and is already comprehensively used as a diagnostic marker in cancer early detection at the uterine cervix. The pooled meta-analysis data confirmed the high sensitivity of p16\(^{\text{INK4a}}\) IHC to detect a transforming HPV infection in oropharyngeal carcinomas. However, specificity was only moderate (Figure 2A and B). This finding implies that p16\(^{\text{INK4a}}\) immunohistochemistry should not be applied as a single test to identify HPV-driven OPSCC since a proportion of tumors will be incorrectly attributed to HPV. The combination with a further test thus appears advisable.

In this context, HPV DNA detection by PCR proved to be a valuable complementing test in our meta-analysis: Both high sensitivity and specificity were demonstrated when p16\(^{\text{INK4a}}\) IHC was combined with HPV DNA detection by PCR (Figures 3C and 4C). HPV DNA PCR represents a widely used HPV detection method in HNSCC (reviewed in \(^{61}\)). The method is routinely available in most surgical laboratories and provides results quickly. Our meta-analysis also demonstrated high sensitivity of sole HPV DNA detection by PCR for HPV-transformed OPSCC (98%; Figure 3A). However, specificity was only moderate (84%; Figure 3A). In consequence, HPV DNA PCR may detect low HPV genome copy numbers without an etiological relevance in the respective head and neck tumors and thus produce a false-positive test result. This could be serious if therapeutic decisions solely rely on PCR-based HPV DNA detection in those tumors. Our meta-analysis supports the recommendation to combine HPV DNA detection by PCR with p16\(^{\text{INK4a}}\) IHC to improve specificity in OPSCC and maintain high sensitivity and corroborates the previous findings of an individual study by Smeets and colleagues \(^{13}\).

In-situ hybridization (ISH) technique is used by many authors to detect HPV DNA in cancers of the head and neck. This method allows for direct visualization of the virus within the tumor cells and
minimizes the risk for a false-positive test result that may derive from tissue contamination with viral DNA. Thus, HPV DNA ISH has been suggested as a particularly specific test for the detection of a biologically meaningful HPV infection in those tumors. However, our meta-analysis data demonstrate that the specificity of HPV DNA ISH lies within the range of p16\textsuperscript{INK4a} IHC and HPV DNA PCR when used as single tests (88% vs. 83% and 84%, respectively) and is substantially lower compared to p16\textsuperscript{INK4a}/HPV DNA PCR combined testing (96%; Table 2). At the same time, the sensitivity of HPV DNA ISH is low compared to the other analyzed tests (85% compared to 94%, 98% and 93% for p16\textsuperscript{INK4a} IHC, HPV DNA PCR and p16\textsuperscript{INK4a} IHC/HPV DNA PCR combined testing; Table 2). Based on this data, the sole use of HPV DNA ISH to detect a transforming HPV infection in OPSCC cannot be recommended. Furthermore, combined testing with another HPV test (p16\textsuperscript{INK4a} IHC or HPV DNA PCR) appears inadvisable since the sensitivity of HPV DNA ISH is already low when used as a single test and would further decrease upon combination with another test on condition that the individual tests have to be positive to rate the overall test result as positive.

From a theoretical point of view the order of testing (1) p16\textsuperscript{INK4a} IHC followed by HPV DNA PCR on the p16\textsuperscript{INK4a}-positive cases or 2) HPV DNA PCR followed by p16\textsuperscript{INK4a} IHC on the DNA-positive cases) does not impact on the sensitivity or specificity of the test combination p16\textsuperscript{INK4a} IHC and HPV DNA PCR. Both tests have to be positive in order to define a result as “positive” and if one or both tests are negative, the sample is considered to be “negative”, irrespective of the testing order. However, from a practical point of view it appears worthwhile to adopt the first option in the clinical setting, as proposed by Smeets and co-workers\textsuperscript{13}. p16\textsuperscript{INK4a} IHC is an easily applicable test with high-throughput capacity, available in virtually all routine laboratories and can be interpreted readily. The more laborious and time-consuming analysis of HPV DNA using PCR can thus be restricted to the p16\textsuperscript{INK4a}-positive cases, minimizing processing time and workload.

Considerations on the applied gold standard
The detection of HPV oncogene transcripts using an amplification-based method was selected as reference standard for a transforming HPV infection in OPSCC in this meta-analysis. HPV E6/E7 oncogene transcription is the pivotal prerequisite for cellular transformation by HPV. It thus represents the currently accepted gold standard to causally attribute a head and neck carcinoma to HPV\textsuperscript{13, 14}. At the same time this reference standard requires critical discussion with regard to its technical execution. In an ideal setting, the desirable technique to identify HPV-transformed OPSCC would comprise the detection of HPV E6 and/or E7 oncogene transcripts a) from all known and putative HR-HPV types, b) in the form of all splice transcript variants from c) fresh-frozen tumor tissue. Furthermore, it would be preferable to
perform those analyses on d) isolated tumor cells, e.g. by means of tumor microdissection. The latter proposition considers the fact that the existence of a permissive HPV infection stage in the head and neck epithelium as it occurs at the uterine cervix during the normal viral life cycle can presently not be definitely ruled out. During the permissive HPV infection stage the HPV E6/E7 oncogenes are expressed at low levels to allow for the extension of the pool of cells carrying viral genomes. Despite this very limited expression of the HPV E6/E7 transcripts they can be detected by sensitive techniques such as qPCR. Thus, there is a principle risk for contamination of tumor tissue with E6/E7 transcripts from neighboring epithelial cells harboring a permissive HPV infection. Furthermore, if the E6/E7 transcripts shall be analyzed quantitatively, the inclusion of non-tumor cells will distort the proportion of the E6/E7 transcripts measured against the housekeeping gene. However, the suggested ideal detection strategy is considerably laborious and consequently not realizable in the routine diagnostic setting.

None of the included studies in this meta-analysis met all of the above outlined demands. All of the included studies detected E6 and/or E7 transcripts from HR-HPV 16, whereas half (12/24) of studies detected oncogene transcripts exclusively from this HPV type (Table S2). The HR-HPV-type 16 is found in about 90% of HPV-associated OPSCC. While the vast majority of HPV-transformed OPSCC will thus be correctly identified by the exclusive detection of HPV-16 transcripts, a small minority of HPV-induced head and neck tumors may be missed in those studies. Furthermore, the specificity of the index and comparator tests may be underestimated in studies with a single type-specific reference standard. This problem in the evaluation of HPV-related diagnostic markers could be overcome in future studies if samples positive for non-HPV-16 types are excluded from the analysis via a preceding HPV DNA analysis.

Several studies in this meta-analysis used a selective gold standard proceeding where HPV E6/E7 transcripts were only analyzed in HPV DNA-positive samples (Table S2). This triaging approach could actually facilitate the detection of E6/E7 transcripts in OPSCC in laboratory routine. The detection of viral DNA as a first step to analyze causal HPV involvement in HNSCC is technically more practicable than detection of viral RNA which is sensitive to degradation. Consequently, only a proportion of tumors with a high pre-test probability of HPV-induced transformation have to be tested for E6/E7 mRNA in subsequent steps. At the same time, HPV DNA detection by PCR is a very sensitive procedure, minimizing the risk of missing HPV involvement in the tumor material in the first place. In line with this consideration our results demonstrated no significant difference between p16\(^{INK4a}\) accuracy tested against the general gold standard or the selective gold standard (Figure S2).
Irrespective of the HPV type, the detection of transcript variants like E6*I using splice site-specific primers may be favored over the detection of the full length E6-E7 transcripts: non-splice site-specific primers can also amplify analogue DNA that may remain after incomplete DNAse digestion and thereby generate false-positive test results. The splice variant E6*I represents the most abundantly expressed transcript from the HR-HPV E6-E7 open reading frame (ORF) in HPV-transformed tissue at the uterine cervix. Furthermore, the short amplicon length of E6*I (60-86 base pairs) also facilitates its detection from FFPE tissue where RNA is often found fragmented. The limitations and advantages of the various transcript detection techniques should be critically discussed in the community and considered in the design of future diagnostic studies.

**Strengths and limitations**

The diagnostic performance of p16INK4a IHC in OPSCC was of major interest to this meta-analysis. Thus, the search term was designed to identify all studies that conducted p16INK4a IHC in OPSCC. Studies that did not perform p16INK4a IHC were excluded from the analyses. Consequently, individual studies that only tested for HPV DNA by PCR or ISH but not for p16INK4a may have been missed. However, the vast majority of published studies that assessed the gold standard “HPV oncogene transcripts” and simultaneously analyzed HPV DNA by PCR or ISH also performed p16INK4a IHC. Consequently, this meta-analysis allowed diagnostic accuracy computation not only for p16INK4a but also for further HPV-related markers from the majority of relevant published studies.

We observed a high variability of definitions for p16INK4a IHC-positivity among the included studies. Two subgroup meta-analyses were conducted stratifying studies by the applied definitions of a positive p16INK4a IHC test result. In particular, we analyzed whether studies applying more rigorous definitions (high percentage of positive cells and extensive expression) would differ from studies with a lower threshold for a positive test result with respect to sensitivity and specificity. While we did not observe significant differences of both accuracy measures in those subgroup analyses (Fig. S3), it has to be taken into account that definitions still differed considerably between studies within the compiled subgroups. Therefore, it appears worthwhile to incorporate the detailed comparison of different p16INK4a cut-offs to detect HPV-transformed OPSCC in the design of future diagnostic studies.

While the included studies were heterogeneous regarding the technical execution of the index/comparator tests and the reference standard they all met the review question of this meta-analysis. For the vast majority of studies concerns of bias and applicability in the categories “index/comparator test”, “reference standard” and “flow and timing” were low (Table S3). Some concerns may exist in the...
category “patient selection” as exclusion criteria were not specifically indicated in some manuscripts. It has to be considered that a large proportion of the included studies were not designed as diagnostic studies. Consequently, the quality measures for diagnostic studies as assessed in the QUADAS analysis are only partially applicable to the included studies in this meta-analysis.

In addition to absolute accuracy computations this meta-analysis also provided data for relative accuracy analyses. In particular, the relative sensitivity and specificity of the test combination p16$^{\text{INK4a}}$ IHC and HPV DNA PCR could be compared to the diagnostic performance of the individual tests, revealing maintained high sensitivity and significantly increased specificity of the test combination compared to the single tests (Figure 4B and C).

This meta-analysis provided data for tumors arising in different anatomic subsites of the oropharynx. While the majority of HPV-induced OPSCC originate in the palatine tonsil, a non-negligible proportion of OPSCC outside this subsite is also causally attributable to HPV. This is particularly true for the base of tongue, which contains the lingual tonsil\textsuperscript{24, 36, 39, 42-45, 48, 54}. Thus, the correct distinction of HPV-induced OPSCC arising in oropharyngeal subsites other than the palatine tonsil is nevertheless clinically relevant. The data of this meta-analysis promote the accurate distinction of HPV-induced from HPV-unrelated tumors in distinct subsites of the oropharynx.

Finally, a small subset of HNSCC arising from non-oropharyngeal locations, such as the oral cavity or larynx, is etiologically driven by HPV\textsuperscript{25, 47, 49, 62, 71, 72}. This meta-analysis focused on carcinomas arising from the oropharynx, where the identification of a transforming HPV infection has major clinical importance. However, a systematic assessment of markers to identify HPV-driven HNSCC in non-oropharyngeal locations may be targeted in future studies to reliably determine the small group of HPV-induced cancers in those anatomic sites.

**Conclusion**

Our meta-analysis demonstrates high sensitivity but only moderate specificity of p16$^{\text{INK4a}}$ and HPV DNA PCR when used as single tests to detect a transforming HPV infection in oropharyngeal squamous cell carcinomas. However, by combining the two tests, specificity can be significantly optimized without affecting the sensitivity. This diagnostic proceeding may support the selection of OPSCC patients for de-escalated and HPV-specific therapeutic regimens.

**ACKNOWLEDGEMENTS**
We are very grateful for the support and provision of additional data by the authors Dr. Dana Holzinger, Dr. Ruud Brakenhoff, Dr. Margaret Brandwein-Gensler, Dr. Camille Ragin, Dr. Andrew Schache, Dr. Richard J Shaw, Dr. James Lewis, Dr. Elgar Susanne Quabius, Dr. Ruth Tachezy, Dr. Francesco Bussu, Dr. Lisa Licitra, Dr. Federica Perrone, Dr. Laura Cerezo Padellano, Dr. Fei-Fei Liu and Dr. Wei Shi. MA was supported by the 7th Framework program of DG Research & Innovation of the European Commission through the COHEAHR Network (coordinated by the University of Amsterdam; grant No. 603019); and the HPV-AHEAD Network (coordinated by IARC, Lyon, France; grant No. FP7-HEALTH-2011-282562).

MA was supported by the 7th Framework program of DG Research & Innovation of the European Commission through the COHEAHR Network (coordinated by the University of Amsterdam; grant No. 603019); and the HPV-AHEAD Network (coordinated by IARC, Lyon, France; grant No. FP7-HEALTH-2011-282562).

**DISCLOSURE**

Magnus von Knebel Doeberitz was co-inventor of various patents related to the diagnostic use of p16\textsuperscript{INK4a} antibodies. He was co-founder, shareholder and member of the supervisory board of mtm Laboratories, Heidelberg, Germany, a company that developed and marketed p16\textsuperscript{INK4a} related reagents and that was acquired by Roche in 2011. Magnus von Knebel Doeberitz received honoraria as scientific advisor and research funds from Oryx GmbH und Co KG. The regular salaries of Miriam Reuschenbach were partially covered by these funds. Magnus von Knebel Doeberitz holds a patent related to therapeutic use of p16\textsuperscript{INK4a} and Magnus von Knebel Doeberitz and Miriam Reuschenbach are inventors of a patent related to therapeutic use of p16\textsuperscript{INK4a}. The other authors report no potential conflict of interest.
REFERENCES


### TABLE 1 Definitions of p16\(^{INK4a}\) positivity assessed in OPSCC specimens among included studies

<table>
<thead>
<tr>
<th>Author, year</th>
<th>Definition of a p16(^{INK4a})-positive test result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smeets et al., 2007</td>
<td>Staining intensity (scored 0-3) above background of negative control (i.e. ≥1)</td>
</tr>
<tr>
<td>Shi et al., 2009</td>
<td>Strong signals in both the nuclei and the cytoplasm</td>
</tr>
<tr>
<td>Schache et al., 2011</td>
<td>Strong and diffuse nuclear and cytoplasmic staining in &gt;70% of the tumor</td>
</tr>
<tr>
<td>Schlecht et al., 2011</td>
<td>Mean intensity cut-off of ≥2 and diffuse (≥75%) staining in either the nuclei or cytoplasm</td>
</tr>
<tr>
<td>Rotnaglova et al., 2011</td>
<td>More than 50% of positive cells revealing nuclear and/or cytoplasmic staining</td>
</tr>
<tr>
<td>Perrone et al., 2011</td>
<td>Strong nuclear and cytoplasmic staining in &gt;50% of the tumor cells</td>
</tr>
<tr>
<td>Holzinger et al., 2012</td>
<td>Positivity in &gt;25% of the tumor cells (nuclear and/or cytoplasmic staining)</td>
</tr>
<tr>
<td>Jordan et al., 2012</td>
<td>Strong and diffuse nuclear and cytoplasmic staining in ≥70% of the tumor</td>
</tr>
<tr>
<td>Hoffmann et al., 2012</td>
<td>Strong nuclear as well as cytoplasmic staining in focal or diffuse distribution (defined as &quot;moderate&quot; (+) or &quot;strong&quot; (+++) expression)</td>
</tr>
<tr>
<td>Gao et al., 2013</td>
<td>More than 50% of cells showing nuclear and cytoplasmic staining</td>
</tr>
<tr>
<td>Rietbergen et al., 2013</td>
<td>Moderate to strong diffuse nuclear and cytoplasmic immunoreactivity in more than 70% of the carcinoma tissue</td>
</tr>
<tr>
<td>Bussu et al., 2013</td>
<td>Strong and diffuse nuclear and cytoplasmic staining in ≥70% of the tumor</td>
</tr>
<tr>
<td>Cerezo et al., 2014</td>
<td>Strong and diffuse nuclear and cytoplasmic staining in ≥70% of the tumor</td>
</tr>
<tr>
<td>Bussu et al., 2014</td>
<td>Strong and diffuse nuclear and cytoplasmic staining in ≥70% of the tumor</td>
</tr>
<tr>
<td>Salazar et al., 2014</td>
<td>&gt;50% of tumor cells presenting with a strong nuclear stain</td>
</tr>
<tr>
<td>Mehrad et al., 2014</td>
<td>Nuclear and cytoplasmic staining, regardless of intensity, in ≥76% of tumor cells</td>
</tr>
<tr>
<td>Prigge et al., 2014</td>
<td>Expression in more than 50% of tumor cells (irrespective of staining intensity and localization)</td>
</tr>
<tr>
<td>Deng et al., 2014</td>
<td>Intense nuclear and/or cytoplasmic reactivity in ≥40% of tumor cells</td>
</tr>
<tr>
<td>Isayeva et al., 2014</td>
<td>Both nuclear and cytoplasmic expression with ≥+2 intensity and ≥75% distribution</td>
</tr>
<tr>
<td>Lukesova et al., 2014</td>
<td>More than 50% of positive cells revealing nuclear and/or cytoplasmic staining</td>
</tr>
<tr>
<td>Quäbius et al., 2014</td>
<td>Strong nuclear as well as cytoplasmic staining in &gt;25% of the cells</td>
</tr>
<tr>
<td>Masterson et al., 2015</td>
<td>&gt;70% tumor cell staining</td>
</tr>
<tr>
<td>Mirghani et al., 2015</td>
<td>Strong and diffuse nuclear and cytoplasmic staining in &gt;70% of the tumor</td>
</tr>
<tr>
<td>Vojtechova et al., 2016</td>
<td>More than 50% of positive cells revealing nuclear and/or cytoplasmic staining</td>
</tr>
</tbody>
</table>

Studies using the same definition are marked by identical grey shadowing.
Table 2 Pooled sensitivity and specificity of p16^{INK4a} IHC, HPV DNA PCR, HPV DNA ISH and p16^{INK4a} IHC/HPV DNA PCR combined testing to identify a transforming HPV infection in OPSCC

<table>
<thead>
<tr>
<th>Anatomical site</th>
<th>Test No of studies</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oropharynx grouped</td>
<td>p16^{INK4a} IHC</td>
<td>23</td>
<td>94% (91-97%)</td>
</tr>
<tr>
<td></td>
<td>HPV DNA PCR</td>
<td>11</td>
<td>98% (94-100%)</td>
</tr>
<tr>
<td></td>
<td>HPV DNA ISH</td>
<td>8</td>
<td>85% (76-92%)</td>
</tr>
<tr>
<td></td>
<td>Combi p16^{INK4a} IHC &amp; HPV DNA PCR</td>
<td>11</td>
<td>93% (87-97%)</td>
</tr>
<tr>
<td>Palatine tonsil</td>
<td>p16^{INK4a} IHC</td>
<td>13</td>
<td>93% (87-98%)</td>
</tr>
<tr>
<td>Base of tongue</td>
<td>p16^{INK4a} IHC</td>
<td>9</td>
<td>95 (87-100%)</td>
</tr>
<tr>
<td>Soft palate</td>
<td>p16^{INK4a} IHC</td>
<td>2</td>
<td>100 (67-100%)</td>
</tr>
</tbody>
</table>

Results are shown for oropharynx grouped and stratified by subsite. HPV DNA PCR: human papillomavirus DNA detection using polymerase chain reaction, HPV DNA ISH: human papillomavirus DNA detection using in-situ hybridization, OPSCC: oropharyngeal squamous cell carcinoma, p16^{INK4a} IHC: p16^{INK4a} immunohistochemistry, 95% CI: 95% confidence interval.

Table 3 Relative accuracy of HPV DNA PCR vs. p16^{INK4a} IHC and of combined testing with HPV DNA PCR and p16^{INK4a} IHC vs. each separate test to identify a transforming HPV infection in OPSCC

<table>
<thead>
<tr>
<th>Compared tests</th>
<th>No of studies</th>
<th>Relative sensitivity (95% CI)</th>
<th>Relative specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV DNA PCR vs. p16^{INK4a} IHC</td>
<td>11</td>
<td>1.01 (0.98-1.03)</td>
<td>1.02 (0.91-1.14)</td>
</tr>
<tr>
<td>Combi p16^{INK4a} IHC &amp; HPV DNA PCR vs. p16^{INK4a} IHC</td>
<td>10</td>
<td>0.97 (0.93-1.02)</td>
<td>1.13 (1.04-1.23)</td>
</tr>
<tr>
<td>Combi p16^{INK4a} IHC &amp; HPV DNA PCR vs. HPV DNA PCR</td>
<td>10</td>
<td>0.97 (0.93-1.02)</td>
<td>1.07 (1.01-1.13)</td>
</tr>
</tbody>
</table>

Results are shown for oropharynx grouped (not stratified by subsite). HPV DNA PCR: human papillomavirus DNA detection using polymerase chain reaction, HPV DNA ISH: human papillomavirus DNA detection using in-situ hybridization, OPSCC: oropharyngeal squamous cell carcinoma, p16^{INK4a} IHC: p16^{INK4a} immunohistochemistry, 95% CI: 95% confidence interval.
FIGURE LEGENDS

Figure 1: PRISMA flow-chart
The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow-chart shows the process of study selection. *The tissue of interest comprised oropharyngeal cancer of squamous epithelial origin (OPSCC). Papers were excluded if the analyzed tissue did not meet these criteria or if analyses were restricted to specific histological variants of OPSCC (e.g. spindle cell variant) or if the tissue was processed as ex-vivo cultures or xenografts. †Studies were excluded if HPV mRNA analyses were only performed on samples with partial p16^{INK4a}-staining, if HPV mRNA analyses were only performed on p16^{INK4a}-positive, HPV DNA-negative samples or if p16^{INK4a} IHC was only performed on HPV DNA or RNA-positive samples.

Figure 2: Meta-analysis of the sensitivity and specificity of p16^{INK4a} IHC to identify a transforming HPV infection in OPSCC
A: oropharynx grouped; B: stratified by subsite (palatine tonsil, base of tongue or soft palate).
A transforming HPV infection was defined by E6 and/or E7 mRNA detection using an amplification-based method. IHC: immunohistochemistry, 95% CI: 95% confidence interval.

Figure 3: Meta-analysis of the sensitivity and specificity of HPV DNA PCR (A), HPV DNA ISH (B) and p16^{INK4a} IHC/HPV DNA PCR combined testing (C) to identify a transforming HPV infection in OPSCC
A transforming HPV infection was defined by E6 and/or E7 mRNA detection using an amplification-based method. For p16^{INK4a} IHC/HPV DNA PCR combined testing a test result was regarded as “positive” if the tumor showed both presence of HPV DNA and p16^{INK4a} overexpression. In contrast, the test result was considered “negative” if the tumor was negative for HPV DNA detection by PCR, p16^{INK4a} IHC expression or both. HPV DNA ISH: human papillomavirus DNA detection using in-situ hybridization, HPV DNA PCR: human papillomavirus DNA detection using polymerase chain reaction, IHC: immunohistochemistry, OPSCC: oropharyngeal squamous cell carcinoma, 95% CI: 95% confidence interval.

Figure 4: Meta-analysis of the relative accuracy of HPV DNA PCR vs. p16^{INK4a} immunohistochemistry (A), of combined testing of both tests vs. p16^{INK4a} immunohistochemistry (B) and of combined testing of both tests vs. HPV DNA PCR (C) to identify a transforming HPV infection in OPSCC
A transforming HPV infection was defined by E6 and/or E7 mRNA detection using an amplification-based method. For p16^{INK4a} IHC/HPV DNA PCR combined testing a test result was regarded as “positive” if the tumor showed both presence of HPV DNA and p16^{INK4a} overexpression. In contrast, the test result was considered “negative” if the tumor was negative for HPV DNA detection by PCR, p16^{INK4a} IHC expression or...
both. HPV DNA PCR: human papillomavirus DNA detection using polymerase chain reaction, IHC: immunohistochemistry, OPSCC: oropharyngeal squamous cell carcinoma, 95% CI: 95% confidence interval.

Figure S1: PICOS and search term
The PICOS (Population - Index test - Comparator tests - Outcomes - Study design) components of the clinical question in line with the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)” guidelines (A) and the search term (B) are detailed. HPV DNA ISH: human papillomavirus DNA detection using in-situ hybridization, HPV DNA PCR: human papillomavirus DNA detection using polymerase chain reaction, IHC: immunohistochemistry, OPSCC: oropharyngeal squamous cell carcinoma.

Figure S2: Meta-analysis on the sensitivity (A) and specificity (B) of p16\textsuperscript{INK4a} IHC to identify a transforming HPV infection in OPSCC, stratified by gold standard
The superior group contains all studies (n=19) in which HPV E6 and/or E7 mRNA detection was performed on all samples (general gold standard). The lower group contains all studies (n=4) in which HPV E6 and/or E7 mRNA detection was exclusively performed on samples that had tested positive for HPV DNA by PCR in a preceding analysis (selective gold standard). IHC: immunohistochemistry, OPSCC: oropharyngeal squamous cell carcinoma, 95% CI: 95% confidence interval.

Figure S3: Subgroup meta-analysis on the sensitivity (A and C) and specificity (B and D) of p16 IHC to identify a transforming HPV infection in OPSCC, stratified by the applied definition of a p16-positive test result
A and B: All studies applying a definition of i) strong and ii) diffuse p16\textsuperscript{INK4a} expression in iii) ≥70% of tumor cells (lower group) were compared to studies in which the definition for a positive test result did not fulfill all three criteria (upper group). C and D: All studies applying a definition of ≥50% p16\textsuperscript{INK4a}-positive cells irrespective of staining intensity, pattern or intracellular localization (lower group) were compared to studies using a definition of <50% p16\textsuperscript{INK4a}-positive cells or not considering the percentage of p16\textsuperscript{INK4a}-positive cells (upper group) for a positive test result.
References retrieved (n=1636)

Relevant abstracts (n=1122)

Exclusion of studies based on
• title (n=5)
• abstract (n=509)

Studies containing primary, p16\textsuperscript{INK4a} immunohistochemistry data (n=798)

Exclusion of studies based on full text (n=324)
• no p16\textsuperscript{INK4a} immunohistochemistry
• no primary data

Exclusion of studies based on full text (n=774)
• no analyses on tissue of interest*
• no HPV oncogene mRNA detection
• HPV oncogene mRNA detection using a non-amplification-based method
• pre-selected population†
• less than 10 samples analyzed
• no author’s response on required raw data

Inclusion of studies for analyses in final review (n=24)