Meta-Analysis of the Accuracy of p16 or p16/Ki-67 Immunocytochemistry Versus HPV Testing for the Detection of CIN2+/CIN3+ in Triage of Women With Minor Abnormal Cytology

Eliana Peeters, MSc [1]; Nicolas Wentzensen, MD, PhD [2]; Christine Bergeron, MD, PhD[3]; and Marc Arbyn, MD, PhD[1]

BACKGROUND: Women with atypical squamous cells of undetermined significance (ASC-US) can be triaged accurately with a high-risk human papillomavirus (hrHPV) test to identify those who need a referral. However, the triage of low-grade squamous intraepithelial lesion (LSIL) with hrHPV testing has very low specificity. Overexpression of p16, with or without Ki-67, indicates neoplastic transformation of human papillomavirus–infected cervical cells and may more accurately predict underlying cervical intraepithelial neoplasia of grade 3 or worse (CIN3+). **METHODS:** A literature search was conducted in 3 bibliographic databases. Studies were selected if they included women with ASC-US or LSIL who were triaged with dual staining (p16/Ki-67) and/or p16 staining and, if available, with a comparator hrHPV test to detect cervical intraepithelial neoplasia of grade 2 or worse (CIN2+) or CIN3+. **RESULTS:** Thirty-eight studies were eligible. The sensitivity of p16 staining for CIN3+ was significantly lower than that of hrHPV DNA testing (ratio for ASC-US, 0.87; 95% confidence interval [CI], 0.78-0.97; ratio for LSIL, 0.86; 95% CI, 0.80-0.93). In contrast, the specificity of p16 staining was substantially higher with relative specificities of 1.60 (95% CI, 1.35-1.88) and 2.29 (95% CI, 2.05-2.56) for ASC-US and LSIL respectively. Dual staining was as sensitive as hrHPV DNA testing but was more specific (ratio for ASC-US, 1.65; 95% CI, 1.42-1.92; ratio for LSIL, 2.45; 95% CI, 2.17-2.77). **CONCLUSIONS:** This meta-analysis confirms that p16 staining and p16/Ki-67 staining are more specific for CIN2+/CIN3+ than hrHPV DNA testing. Although p16 staining is less sensitive for CIN3+ than hrHPV DNA testing, dual staining has similar sensitivity. ***Cancer Cytopathol* 2019;127:169-180**. *© 2019 American Cancer Society.*

KEY WORDS: atypical squamous cells of undetermined significance (ASC-US); cervical cancer; diagnostic test accuracy; immunocytochemistry; low-grade squamous intraepithelial lesion (LSIL); meta-analysis; $p16^{INK4A}$; p16/Ki-67; triage.

INTRODUCTION

Human papillomavirus (HPV)–based cervical cancer screening with a validated assay is being introduced into several Western countries.[1] However, cytology-based screening still remains the main screening method in many countries. Women with high-grade squamous intraepithelial lesion screening results have a high risk of cervical intraepithelial neoplasia of grade 2 or worse (CIN2+) or cervical intraepithelial neoplasia of grade 3 or worse (CIN3+) and should be immediately referred for colposcopy for follow-up. For women with atypical squamous cells of undetermined significance (ASC-US), a triage step is recommended to determine whether the patient should be referred for colposcopy. Previous systematic reviews have revealed that triage

Corresponding author: Marc Arbyn, MD, PhD, Unit of Cancer Epidemiology, Belgian Cancer Centre, Sciensano, Juliette Wytsmanstreet 14, 1050 Brussels, Belgium; marc.arbyn@sciensano.be

[1]Unit of Cancer Epidemiology, Belgian Cancer Centre, Sciensano, Brussels, Belgium; [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; [3]Laboratoire Cerba, Cergy Pontoise, France.

Additional supporting information may be found in the online version of this article.

Received: December 10, 2018; **Accepted:** January 2, 2019

Published online February 27, 2019 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cncy.22103, wileyonlinelibrary.com

with HPV testing (pooled and/or separate detection of HPV genotypes)[2] is more sensitive and has equal specificity in comparison with repeat cytology in women with ASC-US.[3,4] Therefore, HPV-based triage of ASC-US has been recommended in many guidelines worldwide.[5-7] In contrast, HPV testing to triage women with low-grade squamous intraepithelial lesions (LSIL) is hardly useful because most LSILs are HPV-positive, resulting in very poor specificity.[4] This results in a posttest probability after a positive HPV test (a positive predictive value) that is only slightly higher than the pretest prevalence of CIN2+.[3]

Therefore, more specific biomarkers are needed to assist clinicians in the triage of LSIL.[2] One such biomarker is the tumor suppressor protein p16, which prevents normal cells from entering the S phase of the cell cycle.[8-10] However, when cervical cells are transformed through high-risk human papillomavirus (hrHPV) infections, the inactivation of the retinoblastoma protein by the viral oncogenic protein E7 and the subsequent release of transcription factor E2F lead to overexpression of p16 in the nuclei and cytoplasm of cervical cells.[2,9-15] Thus, overexpression of p16 can serve as an indicator of precancerous cervical lesions and cervical cancer.[9] Another useful biomarker is the human Ki-67 protein, which is expressed in the nuclei of proliferating cells during all phases of the cell cycle except for the G0 phase (quiescent cells). Therefore, Ki-67 is useful for determining the cell population's growth fraction and thus can serve as a marker for cell proliferation in normal and malignant cells.[16,17] Both proteins can be detected through immunocytochemistry (see the Materials and Methods section for a description of the index tests).

A previous systematic review compared p16 staining with the signal-based amplification assay HC2.[18] The results showed that for patients with ASC-US, the sensitivities of p16 staining and HC2 to detect CIN2+ were equal (0.95; 95% confidence interval [CI], 0.89-1.01), but the specificity of p16 staining was significantly higher (1.82; 95% CI, 1.57-2.12). With respect to LSIL, the sensitivity of p16 staining was significantly lower (0.87; 95% CI, 0.81-0.94), but the specificity was almost 3 times higher (2.74; 95% CI, 1.99-3.76). Therefore, p16 staining could be recommended for the triage of ASC-US, but for the triage of LSIL, additional follow-up is needed before a patient is referred back for routine screening after a negative p16 staining result.

In this systematic review, we update the current evidence regarding the accuracy of p16 staining and dual staining with p16 and Ki-67 for detecting CIN2+ or CIN3+ in the triage of women with ASC-US or LSIL. Furthermore, we also compare the accuracy of p16 staining and dual staining with the accuracy of hrHPV testing.

MATERIALS AND METHODS

Clinical Question

This systematic review assessed the absolute accuracy (in terms of sensitivity and specificity) of p16 staining and dual staining as well as the relative accuracy of these tests in comparison with hrHPV testing for detecting underlying CIN2+ or CIN3+ in women who had cervical cytology results of ASC-US or LSIL. The study protocol received a priori approval by the appropriate institutional review committee.

Index Test

Overexpression of p16 and co-expression of p16 and Ki-67 within the same cell of the squamous epithelium of the cervix can be visualized through immunocytochemistry.[19,20] This technique makes use of primary and secondary antibodies (labeled with alkaline phosphatase or horseradish peroxidase or biotinylated).[21] The primary antibody binds to the epitope of the protein of interest, whereas the secondary antibody produces a stain when it is bound to the Fc fragment of the primary antibody.[21,22] This stain is produced by the addition of 3,3′-diaminobenzidine brown (for p16) or Fast Red chromogen (for Ki-67), which can be visualized through light microscopy.[21,23] When p16 is overexpressed in the cells, a brown cytoplasmic and/or nuclear stain is produced, whereas a red nuclear stain is visible in cases of overexpression of Ki-67.[21]

Different positivity cutoffs are applied for p16 and/or Ki-67 immunocytochemistry.[19,24] Some p16 evaluations take the abnormal morphology of the cell into account by assessing the nucleocytoplasmic ratio, chromatin distribution, anisonucleosis, nuclear shape, and membrane structure.[25] Cervical slides are given a score of 0 if cells are not stained and a score of 1 to 4 if cells are stained and fulfill 1 or more morphological criteria (the more criteria fulfilled, the higher the score).[15,25] The evaluation of a dual stain is usually simplified: a slide is called positive when a single dual stain–positive cell is found on a slide.

Literature Search

A literature search was performed in EMBASE, MEDLINE, and Scopus to retrieve studies that were published from January 2012 onward to build further on a previously conducted meta-analysis.[18] The last retrieval was executed on October 2017. For each database, a search string was constructed according to the population, index test, comparator test, and outcome method for diagnostic test assessments.[26] The criteria for study inclusion were the following: 1) a group of women who had a cervical cytology result of ASC-US or LSIL, with the 2 groups distinguished; 2) overexpression of p16 or a combination of p16 and Ki-67, with both identified through immunocytochemistry performed on cervical cell specimens; 3) an hrHPV assay identifying viral DNA or RNA or another triage test performed on cervical cell specimens; and 4) the presence of CIN2+ and/or CIN3+ verified by colposcopy and histology. The applied search strings for MEDLINE and EMBASE can be found in in the supporting information (text boxes 1 and 2).

Furthermore, through Scopus, articles were checked that cited at least 1 of the 3 previously conducted meta-analyses on the triage of women with cytological abnormalities with p16 or p16/Ki-67 immunocytochemistry.[18,27,28] The selection of the included studies was performed by E.P. and was verified systematically by M.A. Discordances were discussed until a consensus was reached, and if no consensus was reached, arbitration for inclusion or exclusion was submitted to N.W.

The quality of each included study was scored according to a checklist of 13 Quality Assessment of Diagnostic Accuracy Studies (QUADAS) criteria. These criteria could be further arranged into 4 groups (ie, criteria related to the patient selection, the index test, the reference standard, and the flow and timing of the tests under investigation) to assess the level of bias within each group.[29]

Statistical Analysis

To pool the absolute accuracy of p16 staining, dual staining, and hrHPV testing and construct summary receiver operating characteristic curves, a bivariate normal model was used through implementation of the Stata procedure metandi.[30] We investigated intertest differences between hrHPV DNA assays by including assays as a covariate in the bivariate normal model with the SAS macro metadas.[31] Because no significant differences were found ($P > .05$), all hrHPV assays could be pooled together.

We subsequently evaluated jointly the relative sensitivity and specificity of the following comparisons with a bivariate normal model[31]: 1) p16 staining versus hrHPV DNA tests, 2) dual staining versus hrHPV DNA tests, 3) p16 staining versus hrHPV messenger RNA (mRNA) tests, 4) dual staining versus hrHPV mRNA tests, and 5) p16 staining versus dual staining. Whenever a failure in convergence occurred, the comparison was made separately for the relative sensitivity and relative specificity with a random effect model for ratios of proportions.[32]

We demonstrated the clinical utility of 3 triage tests (ie, p16 staining, dual staining, and hrHPV DNA testing) with pretest-posttest probability (PPP) plots.[33,34] The posttest probabilities were computed from the prevalence of CIN3+ in ASC-US and LSIL, which was derived from a previous meta-analysis,[3] and the likelihood ratios of positive and negative tests were derived from the pooled sensitivity and specificity for each of the 3 triage methods. PPP plots distinguish risk regions (high, medium, and low, which are colored red, yellow, and green, respectively; they are defined by thresholds of CIN3+ probability at 1% and 10%) that suggest patient management decisions (referral for colposcopy, further surveillance, and release to routine screening, respectively).[33,34]

All analyses were performed in Stata (version 14; StataCorp LLC, College Station, Texas) with the exception of the metadas macro, which was performed in SAS (version 9.3; SAS Institute Inc, Cary, North Carolina). The P value for statistical significance was defined as ≤.05.

RESULTS

Selected Studies

We retrieved 458 articles with the search string for MEDLINE and 824 articles with the search string for EMBASE. Together with 6 citations of Kisser and Zechmeister-Koss,[27] 34 citations of Roelens et al,[18] and 182 citations of Tsoumpou et al,[28] we identified 1504 candidate-eligible studies. After the removal of duplicates, 957 references were maintained for further review. Two hundred seventy and 661 irrelevant

TABLE 1. Absolute Accuracy of p16 Staining, Dual Staining, and Validated hrHPV DNA Assays in the Triage of ASC-US or LSIL for the Outcomes CIN2+ and CIN3+

Test	Triage Group	Outcome	No. of Studies	Sensitivity, % (95% CI)[a]	Specificity, % (95% CI)[a]
p16 staining	ASC-US	CIN2+	17	82 (76-87)	71 (65-76)
		CIN3+	9	85 (73-92)	62 (58-65)
	LSIL	CIN2+	15	83 (76-88)	62 (52-71)
		CIN3+	8	86 (79-91)	49 (38-60)
p16/Ki-67 staining	ASC-US	CIN2+	13	84 (77-89)	77 (70-82)
		CIN3+	5	88 (58-98)	72 (67-76)
	LSIL	CIN2+	18	86 (82-89)	66 (59-72)
		CIN3+	6	96 (88-98)	47 (36-58)
Validated hrHPV DNA[b]	ASC-US	CIN2+	25	93 (91-95)	45 (38-53)
		CIN3+	14	98 (85-100)	47 (39-56)
	LSIL	CIN2+	25	95 (94-96)	27 (23-33)
		CIN3+	13	100 (95-100)	22 (19-25)

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; CI, confidence interval; CIN2+, cervical intraepithelial neoplasia of grade 2 or worse; CIN3+, cervical intraepithelial neoplasia of grade 3 or worse; hrHPV, high-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion.

[a]Absolute accuracy estimates were pooled with a binormal model.

[b]Pooled accuracy of 6 hrHPV DNA assays (Abbott RealTime hrHPV, BD Onclarity, Cervista, Cobas 4800, HC2, and Linear Array).

references were excluded on the basis of the title and the abstract, respectively. The reasons for exclusion are mentioned in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flowchart (see Supporting Fig. 1). From the remaining 26 eligible studies, 20 new studies published between January 2012 and October 2017 were included in this meta-analysis because for 6 references there was no response to an additional data request. Furthermore, 16 other eligible studies from Roelens et al published between January 2005 and December 2011 were also included, and this yielded 38 included studies altogether. The study from Denton et al[42] contributed three datasets.

The 2 index tests were compared with 6 clinically validated hrHPV DNA tests (ie, Abbott RealTime hrHPV assay, BD Onclarity, Cervista, Cobas 4800, HC2, and Linear Array [restricted to the 14 hrHPV types]) and 2 hrHPV RNA tests (Aptima and PreTect HPV-Proofer).[1,35] The comparison between dual staining and PreTect HPV-Proofer (with regard to the triage of ASC-US and LSIL) and between dual staining and Aptima (with regard to the triage of ASC-US) could not be analyzed due to a lack of eligible studies. With respect to the comparison of p16 staining and dual staining, data were available only for the outcome CIN2+. Twenty studies[25,36-52] reported the accuracy of p16, among which 1 study[51] also evaluated dual staining. Eighteen studies[21,53-69] evaluated the accuracy of dual staining only. The positivity criteria for p16 staining varied across the studies. In 11 studies[36-40,42-46], positivity for p16 was defined as having more than 1 cervical cell that was p16-immunoreactive and morphologically abnormal. Eight studies[25,41,42,47-50,52] used the nuclear scoring system of Wentzensen et al.[15] For 1 study[51], the p16 positivity cutoff was brown cytoplasmic staining for p16 in more than 10 cervical cells. With respect to dual staining, 19 studies[21,51,53-69] considered simultaneous red nuclear and brown cytoplasmic staining of at least 1 cervical cell as a positivity criterion.

Twenty-two studies provided accuracy data for the outcome CIN2+, whereas 16 studies provided separate data for CIN2+ and CIN3+. In 5 studies, ASC-US was the only triage group, and for 6 studies, this was LSIL. The remaining 27 studies considered both ASC-US and LSIL as triage groups. In total, 4113 women with ASC-US and 5990 with LSIL were included.

A detailed summary of the included studies published from January 2005 to December 2011 can be found in the article of Roelens et al[18]; a summary of the remaining included studies can be found in Supporting Tables 1 and 2.

Assessment of Study Quality

Supporting Table 3 shows that the included studies fulfilled the vast majority of the 13 QUADAS items with a positive quality score ranging from 54% (7 of 13) to 100%. Eight percent (1 of 13) to 46% (6 of 13) of the items received an unclear quality score, and 8% (1 of 13) to 23% (3 of 13) did not fulfill certain items. The risk of bias with respect to patient selection was low in 35 studies and moderate in 1 study. The risk of bias with respect to reporting of the index test was low in 34 studies and

TABLE 2. Relative Accuracy of p16 Staining and Dual Staining Versus hrHPV DNA Tests and hrHPV RNA Testing with Aptima or PreTect HPV-Proofer and Dual Staining Versus p16 Staining in the Triage of ASC-US and LSIL for the Outcomes CIN2+ and CIN3+

Index Test	Comparator Test	Triage Group	Outcome	No. of Studies[a]	Relative Sensitivity (95% CI)[b]	Relative Specificity (95% CI)[b]
p16 staining	Validated hrHPV DNA[d]	ASC-US	CIN2+	13	0.90 (0.84-0.96), *P* = .002	1.60 (1.35-1.88), *P* < .0001
			CIN3+	10	0.87 (0.78-0.97), *P* = .0111	1.28 (1.06-1.55), *P* = .0093
		LSIL	CIN2+	13	0.84 (0.80-0.89), *P* < .0001	2.29 (2.05-2.56), *P* < .0001
			CIN3+	10	0.86 (0.80-0.93), *P* < .0001	2.47 (2.20-2.78), *P* < .0001
p16 staining	PreTect HPV-Proofer[c]	ASC-US	CIN2+	2	0.98 (0.84-1.15), *P* = .813	0.91 (0.81-1.03), *P* = .124
			CIN3+	2	0.92 (0.79-1.08), *P* = .319	0.83 (0.75-0.93), *P* = .001
		LSIL	CIN2+	2	0.95 (0.79-1.13), *P* = .541	0.82 (0.71-0.95), *P* = .009
			CIN3+	2	1.13 (0.95-1.35), *P* = .169	0.73 (0.60-0.90), *P* = .003
p16 staining	Aptima[c]	ASC-US	CIN2+	2	0.82 (0.67-1.01), *P* = .068	1.46 (0.98-2.16), *P* = .061
			CIN3+	2	0.91 (0.74-1.12), *P* = .382	1.48 (1.01-2.15), *P* = .043
		LSIL	CIN2+	2	0.79 (0.72-0.87), *P* = .000	1.87 (1.67-2.09), *P* = .000
			CIN3+	2	0.86 (0.78-0.96), *P* = .005	1.92 (1.73-2.13), *P* = .000
p16 staining	p16/Ki-67[c]	ASC-US	CIN2+	3	0.98 (0.89-1.07), *P* = .59	0.87 (0.76-0.99), *P* = .036
		LSIL	CIN2+	3	0.96 (0.89-1.04), *P* = .338	0.85 (0.70-1.05), *P* = .133
p16/Ki-67 staining	Validated hrHPV DNA[d]	ASC-US	CIN2+	11	0.90 (0.84-0.97), *P* = .004	1.65 (1.42-1.92), *P* < .0001
			CIN3+	5	0.96 (0.81-1.12), *P* = .5752	1.48 (1.29-1.72), *P* < .0001
		LSIL	CIN2+	13	0.90 (0.87-0.94), *P* < .0001	2.45 (2.17-2.77), *P* < .0001
			CIN3+	5	0.96 (0.92-1.00), *P* = .061	2.56 (2.20-2.98), *P* < .0001
p16/Ki-67 staining	Aptima[c]	LSIL	CIN2+	1	0.96 (0.90-1.04), *P* = .336	1.41 (1.06-1.86), *P* = .017
			CIN3+	1	0.98 (0.94-1.03), *P* = .510	1.40 (1.18-1.68), *P* = .0002

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; CI, confidence interval; CIN2+, cervical intraepithelial neoplasia of grade 2 or worse; CIN3+, cervical intraepithelial neoplasia of grade 3 or worse; hrHPV, high-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion.

[a]These studies provided accuracy data for the index and comparator test.

[b]Relative accuracy measures were computed from a binormal model.

[c]Comparisons were performed separately for relative sensitivity and relative specificity because the low number of studies did not allow the binormal model to be run.

[d]Pooled accuracy of 6 different hrHPV DNA assays (Abbott RealTime hrHPV polymerase chain reaction, BD Onclarity, Cervista, Cobas 4800, HC2, and Linear Array).

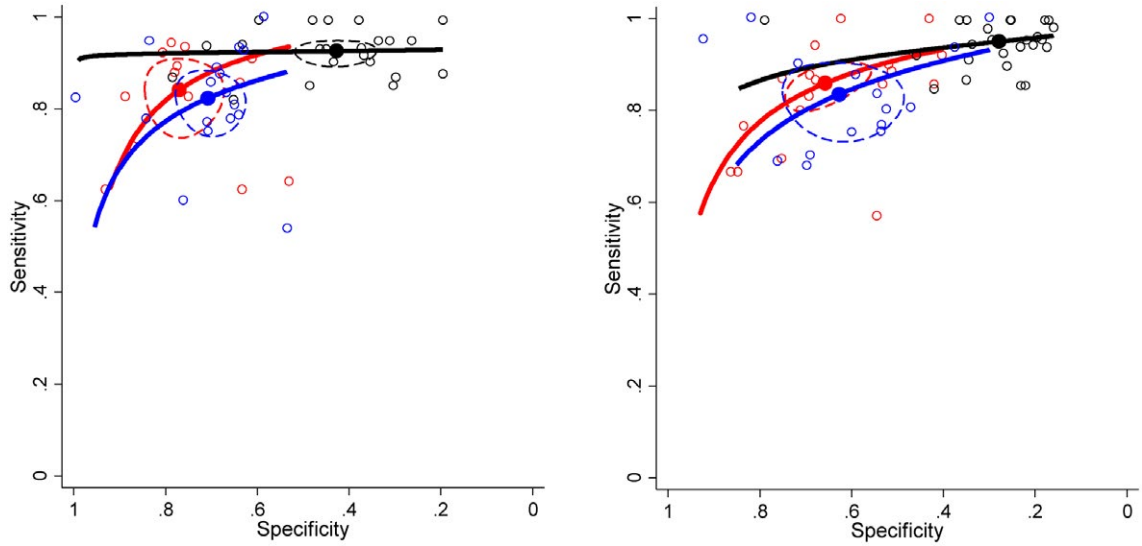


Figure 1. Summary receiver operating characteristic curves of the sensitivity and specificity of dual staining (red), p16 immunocytochemistry (blue), and high-risk human papillomavirus DNA testing (black) for cervical intraepithelial neoplasia of grade 2 or worse in the triage of (*Left*) atypical squamous cells of undetermined significance and (*Right*) low-grade squamous intraepithelial lesions. The filled circles represent summary values of sensitivity and specificity. The hollow circles represent estimates of sensitivity and specificity for each individual study. The dashed lines represent the 95% confidence regions around each summary point. The solid lines represent the summary receiver operating characteristic curves.

moderate in 2 studies. The risk of bias with respect to reporting of the reference standard was low in 32 studies and moderate in 4 studies. We did not find a high risk of bias in the 3 categories.

Absolute Accuracy

The pooled sensitivity and specificity estimates for triaging ASC-US and LSIL patients with p16 staining and dual staining for the outcomes CIN2+ and CIN3+ are listed in Table 1. Loghavi et al.'s study[51] was excluded because of the outlying specifcity of 16 staining for both triage groups (27% [19%-37%] and 14% [8%-24%] with an absolute difference of 44% and 48% compared to the pooled specifcity reported in Table 1 for ASC-US and LSIL, respectively).

The hrHPV DNA assays could be further divided into 2 groups according to their method of amplification. Four tests could be grouped together as target-based amplification assays (ie, Abbott RealTime hrHPV, BD Onclarity, Cobas 4800, and Linear Array), and the other 2 assays (ie, Cervista and HC2) could be grouped together as signal-based amplification assays.[70] The relative accuracy data (target vs signal amplification) can be found in Supporting Table 4. Because there was no significant difference in accuracy between the 2 methods of amplification, the data for hrHPV DNA assays could be pooled together and are shown in Table 1.

The absolute accuracy of mRNA HPV assays targeting 5 (PreTect HPV-Proofer) or 14 hrHPV types (Aptima) and the relative accuracy with p16 immunocytochemistry were already evaluated in previously published meta-analyses, to which we refer the reader for more details.[71,72]

Triage of ASC-US

The sensitivity of triage with p16 staining was 82% (95% CI, 76%-87%) for detecting CIN2+ and 85% (95% CI, 73%-92%) for detecting CIN3+. Its specificity for determining the absence of CIN2+ was 71% (95% CI, 65%-76%). The sensitivity of dual staining was similar: 84% (95% CI, 77%-89%) and 88% (95% CI, 58%-98%) for CIN2+ and CIN3+, respectively. The specificity for cervical intraepithelial neoplasia of less than grade 2 (<CIN2) was 77% (95% CI, 70%-82%). The sensitivity was highest for hrHPV DNA testing with values of 93% (95% CI, 91%-95%) and 98% (95% CI, 85%-100%) for CIN2+ and CIN3+, respectively. On the contrary, hrHPV DNA testing had the lowest specificity with a value of 45% (95% CI, 38%-53%) for <CIN2 (see Table 1). PreTect HPV-Proofer had low sensitivity for CIN2+ (75.4%; 68.1%-82.7%; absolute difference of 17.6% compared with the pooled sensitivity of hrHPV DNA reported in Table 1), but high specificity for <CIN2 (77.9%; 70.1%-85.7%; absolute difference of 32.9% compared with the pooled specificity of hrHPV DNA reported in Table 1).[71] Aptima showed good performance for both sensitivity for CIN2+ (95.7%; 91.5%-97.2%; absolute difference of 2.7% compared with the pooled sensitivity of hrHPV DNA reported in Table 1) and specificity for <CIN2 (56.4%; 44.7%-67.5%; absolute difference of 11.4% compared with the pooled specificity of hrHPV DNA reported in Table 1).[72]

Triage of LSIL

For the triage test p16, the absolute sensitivity was 83% (95% CI, 76%-88%) for detecting CIN2+ and 86% (95% CI, 79%-91%) for detecting CIN3+. The absolute specificity for determining the absence of CIN2+ was 62% (95% CI, 52%-71%). With respect to dual staining, the absolute sensitivity was higher with values of 86% (95% CI, 82%-89%) and 96% (95% CI, 88%-98%) for CIN2+ and CIN3+, respectively. The absolute specificity was slightly higher for <CIN2 (66%; 95% CI, 59%-72%). The absolute sensitivity was highest for hrHPV DNA testing with values of 95% (95% CI, 94%-96%) and 100% (95% CI, 95%-100%) for CIN2+ and CIN3+, respectively. Conversely, hrHPV DNA testing had the lowest specificity with a value of 27% (95% CI, 23%-33%) for <CIN2 (see Table 1). PreTect HPV-Proofer had low sensitivity for CIN2+ (76.2%; 68.3%-76.9%; absolute difference of 19% compared with the pooled sensitivity of hrHPV DNA reported in Table 1), but high specificity for <CIN2 (74.2%; 70.1%-85.7%; absolute difference of 47% compared with the pooled specificity of hrHPV DNA reported in Table 1).[71] Aptima showed good performance for both sensitivity for CIN2+ (91.0%; 85.2%-94.7%; absolute difference of 4% compared with the pooled sensitivity of hrHPV DNA reported in Table 1) and specificity for <CIN2 (42.5%; 33.3%-52.3%; absolute difference of 15.5% compared with the pooled specificity of hrHPV DNA reported in Table 1).[72]

Relative Accuracy

The relative sensitivity and specificity estimates and their 95% CIs for triaging ASC-US and LSIL patients with p16 staining and dual staining versus hrHPV testing for the outcomes CIN2+ and CIN3+ are listed in Table 2. The relative study-specific and pooled accuracy values of triage with p16 staining or dual staining versus the comparator tests for the outcome CIN2+ are shown in forest plots (Supporting Figs. 2-5), whereas the absolute accuracy values of the index and comparator test are displayed in summary receiver operating characteristic plots (Fig. 1). Nine studies were not taken into account when we evaluated the relative accuracy because either they did not include a comparator test (n = 8) or a nonclinically validated hrHPV DNA assay (n = 1) was evaluated.

Triage of ASC-US

The sensitivity for detecting CIN2+ and CIN3+ was significantly lower with p16 staining than hrHPV DNA testing: the ratios were 0.90 (95% CI, 0.84-0.96) and 0.87 (95% CI, 0.78-0.97), respectively. The specificity for determining the absence of CIN2+ was significantly higher for p16 staining than hrHPV DNA testing (ratio, 1.60; 95% CI, 1.35-1.88). Dual staining was also less sensitive than hrHPV DNA, but this was significant only for the detection of CIN2+ (ratio, 0.90; 95% CI, 0.84-0.97). The specificity of dual staining for <CIN2 was significantly higher than the specificity of hrHPV DNA (ratio, 1.65; 95% CI, 1.42-1.92; see Fig. 1 [left]). The CIs around relative accuracy values of p16 staining versus PreTect HPV-Proofer always included unity with the exception of the relative specificity for excluding CIN3+ (ratio, 0.83; 95% CI, 0.75-0.93). p16 staining was also not different from Aptima except for the specificity for excluding CIN3+, which was 48% higher for p16 staining (ratio, 1.48; 95% CI, 1.01-2.15). Dual staining and p16 staining were equally sensitive for detecting CIN2+. However, p16 staining was 13% less specific than dual staining for <CIN2 (ratio, 0.87; 95% CI, 0.76-0.99; see Fig. 1 [left]).

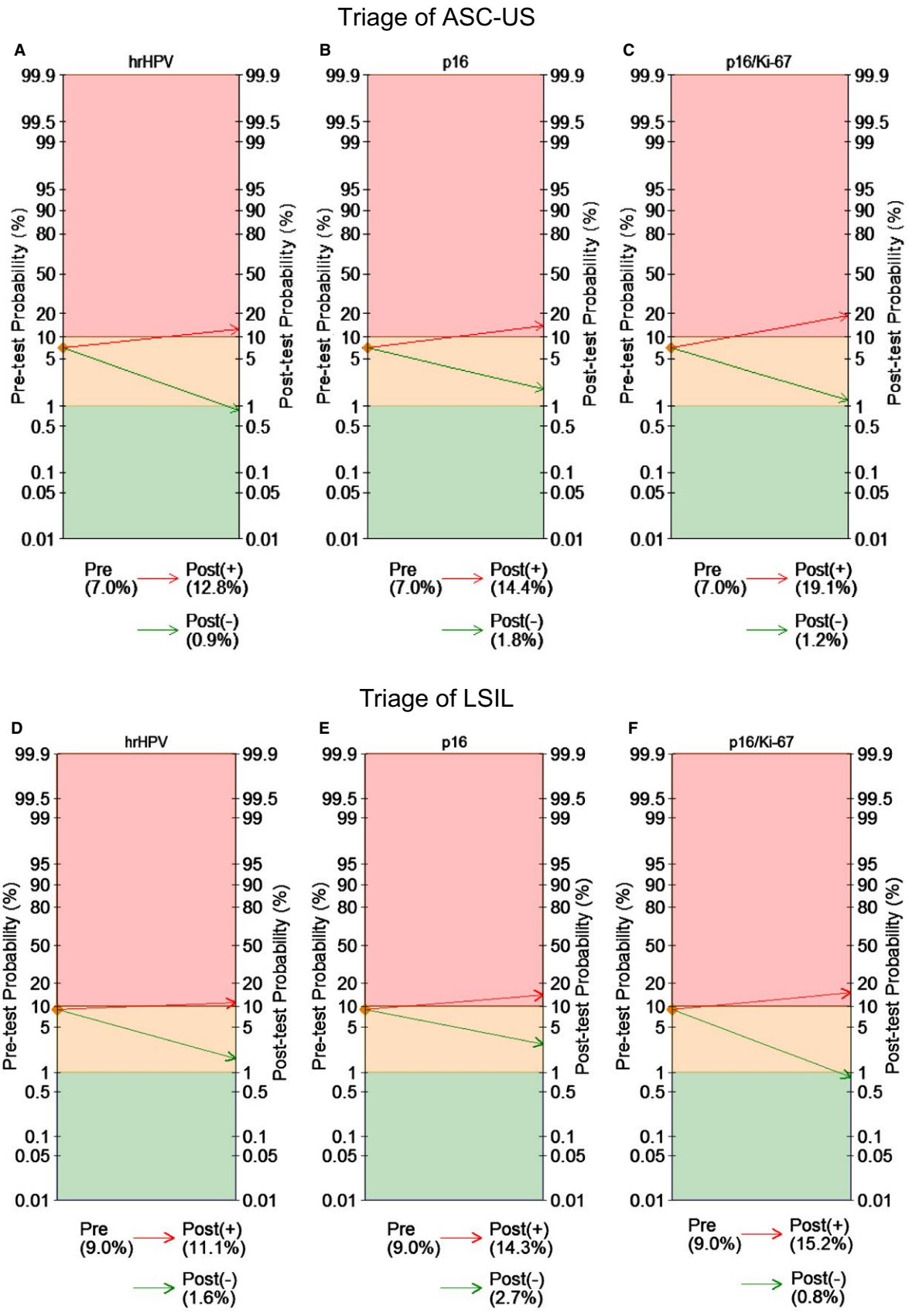
Triage of LSIL

The sensitivity for detecting CIN2+ and CIN3+ was significantly lower with p16 staining than hrHPV DNA testing. The relative sensitivity was 0.84 (95% CI, 0.80-0.89) for CIN2+ and 0.86 (95% CI, 0.80-0.93) for CIN3+. p16 staining was more specific in excluding CIN2+ than hrHPV DNA testing (ratio, 2.29; 95% CI, 2.05-2.56). Dual staining was also less sensitive for CIN2+ than hrHPV DNA testing (ratio, 0.90; 95% CI, 0.87-0.94) but more specific for <CIN2 (ratio, 2.45; 95% CI, 2.17-2.77; see Fig. 1 [right]). p16 staining was not found to be significantly different from PreTect HPV-Proofer in sensitivity for CIN2+ and CIN3+. However, the specificity for CIN2+ was significantly lower (ratio, 0.82; 95% CI, 0.71-0.95). The sensitivity of p16 staining for CIN2+ and CIN3+ was significantly lower than the sensitivity of Aptima (ratio for CIN2+, 0.79; 95% CI, 0.72-0.87; ratio for CIN3+, 0.86; 95% CI, 0.78-0.96), but the specificity was higher for <CIN2 (ratio, 1.87; 95% CI, 1.67-2.09). The CIs around the relative sensitivity of dual staining versus Aptima always included unity, whereas the relative specificity for excluding CIN2+ was significantly higher with dual staining than Aptima (ratio, 1.41; 95% CI, 1.06-1.86). Dual staining and p16 staining were equally sensitive and specific for detecting and excluding CIN2+, respectively (see Fig. 1 [right]).

PPP Plots

The PPP plot in Figure 2A shows that for a patient with ASC-US and a positive hrHPV DNA test, the risk of underlying CIN3+ is 12.8%. If the patient has a negative hrHPV DNA test, this risk is reduced to 0.9%, which lies in the green region. These risks are slightly increased when immunocytochemistry is applied as a triage method. The risks of having CIN3+ with positive p16 staining and dual staining are 14.4% and 19.1%, respectively, whereas with negative p16 staining and dual staining, the risks become 1.8% and 1.2%, respectively (see Fig. 2B,C).

For the triage of LSIL, a positive hrHPV DNA test exceeds the threshold of 10% with 1.1 percentage points (see Fig. 2D), whereas a negative hrHPV DNA test leads to a posttest risk of 1.6%, which lies in the yellow region, where one is indifferent between releasing the patient for routine screening and referring the patient for colposcopy. The average risk of CIN3+ for patients with a positive p16 staining or dual staining result is above the decision threshold for colposcopy referral with a risk of 14.3% or 15.2%, respectively. The risk of CIN3+ after negative p16 staining still lies in the yellow region with a risk of 2.7% and falls into the green region with a risk of 0.8% after negative dual staining (see Fig. 2E,F).



DISCUSSION

Triage of ASC-US

This meta-analysis demonstrated that for the triage of ASC-US, p16 staining was 13% less sensitive for detecting CIN3+ but was 60% more specific than hrHPV DNA testing for CIN2+. Dual staining was significantly less sensitive than hrHPV DNA testing in detecting CIN3+. However, it was 65% more specific in excluding CIN2+ in comparison with hrHPV DNA testing. p16 tended to be less sensitive for CIN2+ than Aptima and PreTect HPV-Proofer. The specificity of p16 for excluding CIN2+ tended to be higher in comparison with Aptima but not in comparison with PreTect HPV-Proofer. However, the differences between p16 and mRNA testing were not significant, and this might be ascribed to the low number of cases and studies. Therefore, no strong conclusions can be made regarding the comparison of p16 and mRNA testing.

On the basis of the risks of having CIN3+ after a positive or negative test, we can suggest that for the triage of ASC-US, hrHPV DNA testing is suitable and is consistent with the results of previously conducted systematic reviews. p16 staining or dual staining can reduce the burden of follow-up, but it is not sufficiently sensitive as a standalone triage technique to bring the risk of CIN3+ below 1%. Nevertheless, it should be recognized that the risks of CIN3+ after a negative hrHPV DNA test or after negative dual staining are comparable in the PPP plots.

Triage of LSIL

In the triage of LSIL, p16 staining was 14% less sensitive for CIN3+, but it was more than 2-fold more specific in comparison with hrHPV DNA testing for the absence of CIN2+. The sensitivity of dual staining for CIN3+ was not significantly lower than the sensitivity of hrHPV DNA testing, but it was more than 2-fold more specific in excluding CIN2+. p16 tended to be less sensitive for CIN2+ and CIN3+ than Aptima but was similarly sensitive in comparison with PreTect HPV-Proofer. p16 showed lower specificity than PreTect HPV-Proofer but higher specificity than Aptima. Again, the comparisons with mRNA tests must be interpreted with caution because they are based on a small number of studies.

Our findings confirm the rather low performance of hrHPV DNA testing in LSIL triage. The positive predictive value of hrHPV DNA testing for CIN3+ is hardly different from the pretest risk,[3] and a negative hrHPV DNA test result does not reduce the risk below 1%. Dual staining in LSIL appears to be more efficient in LSIL triage with a positive predictive value for CIN3+ clearly higher than 10% and a complement of the negative predictive value less than 1%.

Comparison of p16 Staining and Dual Staining

Dual staining and p16 staining are equally sensitive with respect to ASC-US and LSIL triage. However, p16 staining was found to be less specific than dual staining: 13% significantly less specific in ASC-US triage and 15% less specific in LSIL triage although not significantly. These findings are consistent with the results from 2 studies that compared p16 staining and dual staining directly.[51,53] Both studies found similar sensitivity and lower specificity for p16 staining in the triage of ASC-US and LSIL.

p16 positivity can be caused by p16 expression in normal squamous metaplastic, endocervical, and atrophic cells without underlying intraepithelial neoplasia.[18,37,73] Their impact on the accuracy can be limited not only by a focus on the detection of at least 1 stained cervical cell (or more) but also by an investigation of the morphology of cervical cells.[25] This 2-fold analysis enables us to discriminate between p16-positive normal and dysplastic cells and further enhances the specificity of the staining method, although it can hamper the interobserver reproducibility through differences in morphological interpretation.[18,25,37,53] In this meta-analysis, 35% of the studies in which p16 staining was analyzed (7 of 20; see Supporting Table 2) considered, in addition to the detection of at least 1 stained cervical cell, morphological criteria for defining p16 positivity. Furthermore, the simultaneous detection of p16 and Ki-67, regardless of the morphology of the cervical cell, can also enhance accuracy, and this is also proven by our analysis.[53]

Figure 2. Pretest-posttest probability plots for the triage of (A-C) ASC-US and (D-F) LSIL with (*Left*) hrHPV DNA testing, (*Middle*) p16 staining, and (*Right*) dual staining to detect cervical intraepithelial neoplasia of grade 3 or worse. ASC-US indicates atypical squamous cells of undetermined significance; hrHPV, high-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion.

Comparison With Previous Meta-Analyses

Our findings for the comparison of p16 staining and hrHPV DNA testing in the triage of LSIL for CIN3+ were in agreement with the results of Roelens et al[18] (p16 staining was less sensitive but more specific). Concerning ASC-US triage, our findings also support the higher specificity of p16 staining in comparison with hrHPV DNA testing. This meta-analysis showed that p16 staining resulted in a loss of sensitivity in comparison with hrHPV DNA testing, whereas Roelens et al found similar sensitivity. Nonetheless, we must note that HC2 was the only hrHPV DNA test under investigation in the study of Roelens et al.

In conclusion, this meta-analysis has demonstrated that there is a loss in sensitivity for CIN3+ with p16 staining, but not with dual staining, in comparison with hrHPV DNA testing in triaging ASC-US and LSIL patients. Both staining methods showed higher specificity than hrHPV DNA testing, especially in LSIL triage. This can limit the burden of overdetection by preventing unnecessary health care costs and potential adverse events caused by overtreatment of nondiseased women.[74]

FUNDING SUPPORT

Eliana Peeters and Marc Arbyn were supported by the 7th Framework Programme of DG Research and Innovation of the European Commission via the Comparing health services interventions for the prevention of HPV-related cancer (COHEAHR) Network (grant 603019).

CONFLICT OF INTEREST DISCLOSURES

Nicolas Wentzensen is employed by the National Cancer Institute, which has received cervical cancer screening assays in kind or at reduced cost from BD and Roche for studies on which he is working. Eliana Peeters and Marc Arbyn are researchers at Sciensano, which receives funding from the VALidation of HPV GENotyping Tests (VALGENT) and VALidation of HUman papillomavirus assays and collection DEvices for HPV testing on Self-samples and urine samples (VALHUDES) projects[75,76] to evaluate human papillomavirus assays; the researchers do not receive any material or financial advantages from these projects. The other author made no disclosures.

AUTHOR CONTRIBUTIONS

Eliana Peeters: Study retrieval after 2011, data extraction, and writing of the manuscript. **Nicolas Wentzensen:** Evaluation of the inclusion of studies and critical review of the manuscript. **Christine Bergeron:** Critical review of the manuscript. **Marc Arbyn:** Study concept, data retrieval until 2017, methods of statistical analysis, and critical review of the manuscript.

REFERENCES

1. Arbyn M, Snijders PJ, Meijer CJLM, et al. Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening? *Clin Microbiol Infect*. 2015;21:817-826.
2. Cuschieri K, Wentzensen N. Human papillomavirus mRNA and p16 detection as biomarkers for the improved diagnosis of cervical neoplasia. *Cancer Epidemiol Biomarkers Prev*. 2008;17:2536-2545.
3. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. *Vaccine*. 2012;30(suppl 5):F88-F99.
4. Arbyn M, Roelens J, Simoens C, et al. Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. *Cochrane Database Syst Rev*. 2013;3:1-201.
5. Arbyn M, Anttila A, Jordan J, et al. European guidelines for quality assurance in cervical cancer screening. Second edition—summary document. *Ann Oncol*. 2010;21:448-458.
6. Wright TC Jr, Massad LS, Dunton CJ, et al. 2006 consensus guidelines for the management of women with cervical intraepithelial neoplasia or adenocarcinoma in situ. *Am J Obstet Gynecol*. 2007;197:340-345.
7. McLachlin CE, Mai E, Murphy J, et al. Ontario cervical cancer screening clinical practice guidelines. *J Obstet Gynaecol Can*. 2007;29:344-353.
8. Sherr CJ. The INK4a/ARF network in tumour suppression. *Nat Rev Mol Cell Biol*. 2001;2:731-737.
9. Munger K, Gwin TK, McLaughlin-Drubin ME. p16 in HPV-associated cancers. *Oncotarget*. 2013;4:1864-1865.
10. Klaes R, Friedrich T, Spitkovsky D, et al. Overexpression of p16(INK4A) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. *Int J Cancer*. 2001;92:276-284.
11. Sano T, Masuda N, Oyama T, Nakajima T. Overexpression of p16 and p14ARF is associated with human papillomavirus infection in cervical squamous cell carcinoma and dysplasia. *Pathol Int*. 2002;52:375-383.
12. Sarian LO, Derchain SF, Yoshida A, et al. Expression of cyclooxygenase-2 (COX-2) and Ki67 as related to disease severity and HPV detection in squamous lesions of the cervix. *Gynecol Oncol*. 2006;102:537-541.
13. Yim EK, Park JS. The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. *Cancer Res Treat*. 2005;37:319-324.
14. Evangelou K, Bramis J, Peros I, et al. Electron microscopy evidence that cytoplasmic localization of the p16(INK4A) "nuclear" cyclin-dependent kinase inhibitor (CKI) in tumor cells is specific and not an artifact. A study in non–small cell lung carcinomas. *Biotech Histochem*. 2004;79:5-10.
15. Wentzensen N, Bergeron C, Cas F, et al. Evaluation of a nuclear score for p16(INK4a)-stained cervical squamous cells in liquid-based cytology samples. *Cancer*. 2005;105:461-467.
16. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. *J Cell Physiol*. 2000;182:311-322.
17. Gerdes J, Li L, Schlueter C, et al. Immunobiochemical and molecular biologic characterization of the cell proliferation–associated nuclear antigen that is defined by monoclonal antibody Ki-67. *Am J Pathol*. 1991;138:867-873.
18. Roelens J, Reuschenbach M, von Knebel-Doeberitz M, et al. p16INK4a immunocytochemistry versus HPV testing for triage of women with minor cytological abnormalities: a systematic review and meta-analysis. *Cancer Cytopathol*. 2012;120:294-307.
19. Darragh TM, Colgan TJ, Cox JT, et al. The Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *Arch Pathol Lab Med*. 2012;136:1266-1197.

20. Yu LL, Chen W, Lei XQ, et al. Evaluation of p16/Ki-67 dual staining in detection of cervical precancer and cancers: a multicenter study in China. *Oncotarget*. 2016;7:21181-21189.
21. Wentzensen N, Schwartz L, Zuna RE, et al. Performance of p16/Ki-67 immunostaining to detect cervical cancer precursors in a colposcopy referral population. *Clin Cancer Res*. 2012;18:4154-4162.
22. Burry RW. Controls for immunocytochemistry: an update. *J Histochem Cytochem*. 2011;59:6-12.
23. Galgano MT, Castle PE, Atkins KA, et al. Using biomarkers as objective standards in the diagnosis of cervical biopsies. *Am J Surg Pathol*. 2010;34:1077-1087.
24. Bergeron C, Ordi J, Schmidt D, et al. Conjunctive p16INK4a testing significantly increases accuracy in diagnosing high-grade cervical intraepithelial neoplasia. *Am J Clin Pathol*. 2010;133:395-406.
25. Wentzensen N, Bergeron C, Cas F, Vinokurova S, von Knebel DM. Triage of women with ASCUS and LSIL cytology: use of qualitative assessment of p16INK4a positive cells to identify patients with high-grade cervical intraepithelial neoplasia. *Cancer*. 2007;111:58-66.
26. Bae JM. An overview of systematic reviews of diagnostic tests accuracy. *Epidemiol Health*. 2014;36:e2014016.
27. Kisser A, Zechmeister-Koss I. A systematic review of p16/Ki-67 immuno-testing for triage of low grade cervical cytology. *BJOG*. 2015;122:64-70.
28. Tsoumpou I, Arbyn M, Kyrgiou M, et al. p16INK4a immunostaining in cytological and histological specimens from the uterine cervix: a systematic review and meta-analysis. *Cancer Treat Rev*. 2009;35:210-220.
29. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:1-13.
30. Harbord RM, Whiting P. metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J*. 2009;9:211-229.
31. Takwoingi Y, Deeks J. MetaDAS: a SAS macro for meta-analysis of diagnostic accuracy studies. Quick reference and worked example. https://methods.cochrane.org/sites/methods.cochrane.org.sdt/files/public/uploads/MetaDAS%20Quick%20Reference%20v1.3%20May%202012.pdf. Accessed January 15, 2019.
32. Harris R, Bradburm M, Deeks J, et al. metan: fixed- and random-effects meta-analysis. *Stata J*. 2008;8:3-26.
33. Arbyn M, Xu L, Verdoodt F, et al. Genotyping for human papillomavirus types 16 and 18 in women with minor cervical lesions: a systematic review and meta-analysis. *Ann Intern Med*. 2017;166:118-127.
34. Arbyn M, Redman CWE, Verdoodt F, et al. Incomplete excision of cervical pre-cancer as predictor of treatment failure: a systematic review and meta-analysis. *Lancet Oncol*. 2017;18:1665-1679.
35. Xu L, Ostrbenk A, Poljak M, Arbyn M. Assessment of the Roche Linear Array HPV genotyping test within the VALGENT framework. *J Clin Virol*. 2018;98:37-42.
36. Nieh S, Chen SF, Chu TY, et al. Is p16(INK4A) expression more useful than human papillomavirus test to determine the outcome of atypical squamous cells of undetermined significance-categorized Pap smear? A comparative analysis using abnormal cervical smears with follow-up biopsies. *Gynecol Oncol*. 2005;97:35-40.
37. Holladay EB, Logan S, Arnold J, Knesel B, Smith GD. A comparison of the clinical utility of p16(INK4a) immunolocalization with the presence of human papillomavirus by Hybrid Capture 2 for the detection of cervical dysplasia/neoplasia. *Cancer*. 2006;108:451-461.
38. Meyer JL, Hanlon DW, Andersen BT, Rasmussen OF, Bisgaard K. Evaluation of p16(INK4a) expression in ThinPrep cervical specimens with the CINtec p16(INK4a) assay: correlation with biopsy follow-up results. *Cancer*. 2007;111:83-92.
39. Monsonego J, Pollini G, Evrard MJ, et al. P16(INK4a) immunocytochemistry in liquid-based cytology samples in equivocal Pap smears: added value in management of women with equivocal Pap smear. *Acta Cytol*. 2007;51:755-766.
40. Schledermann D, Andersen BT, Bisgaard K, et al. Are adjunctive markers useful in routine cervical cancer screening? Application of p16(INK4a) and HPV-PCR on ThinPrep samples with histological follow-up. *Diagn Cytopathol*. 2008;36:453-459.
41. Szarewski A, Ambroisine L, Cadman L, et al. Comparison of predictors for high-grade cervical intraepithelial neoplasia in women with abnormal smears. *Cancer Epidemiol Biomarkers Prev*. 2008;17:3033-3043.
42. Denton KJ, Bergeron C, Klement P, et al. The sensitivity and specificity of p16INK4a cytology vs HPV testing for detecting high-grade cervical disease in the triage of ASC-US and LSIL Pap cytology results. *Am J Clin Pathol*. 2010;134:12-21.
43. Guo M, Warriage I, Mutyala B, et al. Evaluation of p16 immunostaining to predict high-grade cervical intraepithelial neoplasia in women with Pap results of atypical squamous cells of undetermined significance. *Diagn Cytopathol*. 2011;39:482-488.
44. Passamonti B, Gustinuci D, Rechia P, et al. Expression of p16 in abnormal Pap-tests as an indicator of CIN2+ lesions: a possible role in the low grade ASC/US and L/Sil (lg) cytologic lesions for screening prevention of uterine cervical tumours. *Pathologica*. 2010;102:6-11.
45. Samarawardana P, Dehn DL, Singh M, et al. p16(INK4a) is superior to high-risk human papillomavirus testing in cervical cytology for the prediction of underlying high-grade dysplasia. *Cancer Cytopathol*. 2010;118:146-156.
46. Sung CO, Kim SR, Oh YL, Song SY. The use of p16(INK4A) immunocytochemistry in "atypical squamous cells which cannot exclude HSIL" compared with "atypical squamous cells of undetermined significance" in liquid-based cervical smears. *Diagn Cytopathol*. 2010;38:168-171.
47. Alameda F, Piujan L, Lioveras B, et al. The value of p16 in ASCUS cases: a retrospective study using frozen cytologic material. *Diagn Cytopathol*. 2011;39:110-114.
48. Ma YY, Cheng XD, Zhou CY, et al. Value of P16 expression in the triage of liquid-based cervical cytology with atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesions. *Chin Med J (Engl)*. 2011;124:2443-2447.
49. Nasioutziki M, Daniilidis A, Dinas K, et al. The evaluation of p16INK4a immunoexpression/immunostaining and human papillomavirus DNA test in cervical liquid-based cytological samples. *Int J Gynecol Cancer*. 2011;21:79-85.
50. Szarewski A, Mesher D, Cadman L, et al. A comparison of seven tests for high grade cervical intraepithelial neoplasia in women with abnormal smears: the Predictors 2 study. *J Clin Microbiol*. 2012;50:1867-1873.
51. Loghavi S, Walts AE, Bose S. CINtec(R) plus dual immunostain: a triage tool for cervical pap smears with atypical squamous cells of undetermined significance and low grade squamous intraepithelial lesion. *Diagn Cytopathol*. 2013;41:582-587.
52. Stanculescu R, Bratila E, Bausic V, Vladescu T. The triage of low-grade cytological abnormalities by the immunocytological expression of cyclin-dependent kinase inhibitor p16INK4a versus human papillomavirus test: a real possibility to predict cervical intraepithelial neoplasia CIN2 or CIN2+. *Rom J Morphol Embryol*. 2013;54:1061-1065.
53. Schmidt D, Bergeron C, Denton KJ, Ridder R. p16/ki-67 dual-stain cytology in the triage of ASCUS and LSIL Papanicolaou cytology: results from the European equivocal or mildly abnormal Papanicolaou cytology study. *Cancer Cytopathol*. 2011;119:158-166.
54. Dona MG, Vocaturo A, Giuliani M, et al. p16/Ki-67 dual staining in cervico-vaginal cytology: correlation with histology, human papillomavirus detection and genotyping in women undergoing colposcopy. *Gynecol Oncol*. 2012;126:198-202.

55. Edgerton N, Cohen C, Siddiqui MT. Evaluation of CINtec PLUS(R) testing as an adjunctive test in ASC-US diagnosed SurePath(R) preparations. *Diagn Cytopathol*. 2013;41:35-40.
56. Killeen JL, Dye T, Grace C, Hiraoka M. Improved abnormal Pap smear triage using cervical cancer biomarkers. *J Low Genit Tract Dis*. 2014;18:1-7.
57. Vrdoljak-Mozetic D, Krasevic M, Ostojic VD, et al. HPV16 genotype, p16/Ki-67 dual staining and koilocytic morphology as potential predictors of the clinical outcome for cervical low-grade squamous intraepithelial lesions. *Cytopathology*. 2015;26:10-18.
58. Waldstrom M, Christensen RK, Ornskov D. Evaluation of p16(INK4a)/Ki-67 dual stain in comparison with an mRNA human papillomavirus test on liquid-based cytology samples with low-grade squamous intraepithelial lesion. *Cancer Cytopathol*. 2013;121:136-145.
59. Fujii T, Saito M, Hasegawa T, et al. Performance of p16/Ki-67 immunocytochemistry for identifying CIN2+ in atypical squamous cells of undetermined significance and low-grade squamous intraepithelial lesion specimens: a Japanese Gynecologic Oncology Group study. *Int J Clin Oncol*. 2015;20:134-142.
60. Ordi J, Sagasta A, Munmany M, et al. Usefulness of p16/Ki67 immunostaining in the triage of women referred to colposcopy. *Cancer Cytopathol*. 2014;122:227-235.
61. Uijterwaal MH, Witte BI, van Kemenade FJ, et al. Triaging borderline/mild dyskaryotic Pap cytology with p16/Ki-67 dual-stained cytology testing: cross-sectional and longitudinal outcome study. *Br J Cancer*. 2014;110:1579-1586.
62. Bergeron C, Ikenberg H, Sideri M, et al. Prospective evaluation of p16/Ki-67 dual-stained cytology for managing women with abnormal Papanicolaou cytology: PALMS study results. *Cancer Cytopathol*. 2015;123:373-381.
63. Possati-Resende JC, Fregnani JH, Kerr LM, et al. The accuracy of p16/Ki-67 and HPV test in the detection of CIN2/3 in women diagnosed with ASC-US or LSIL. *PLoS One*. 2015;10:e0134445.
64. White C, Bakhiet S, Bates M, et al. Triage of LSIL/ASC-US with p16/Ki-67 dual staining and human papillomavirus testing: a 2-year prospective study. *Cytopathology*. 2016;27:269-276.
65. Yu LL, Guo HQ, Lei XQ, et al. p16/Ki-67 co-expression associates high risk human papillomavirus persistence and cervical histopathology: a 3-year cohort study in China. *Oncotarget*. 2016;7:64810-64819.
66. Robial R, Longatto-Filho A, Roteli-Martins CM, et al. Frequency of *Chlamydia trachomatis* infection in cervical intraepithelial lesions and the status of cytological p16/Ki-67 dual-staining. *Infect Agent Cancer*. 2017;12:3.
67. Ziemke P. p16/Ki-67 immunocytochemistry in gynecological cytology: limitations in practice. *Acta Cytol*. 2017;61:230-236.
68. Yu J, Zhu HT, Zhao JJ, Su JZ, Xia YD. Sorting role of p16(INK4a)/Ki-67 double immunostaining in the cervical cytology specimens of ASCUS and LSIL cases [in Chinese]. *Zhonghua Bing Li Xue Za Zhi*. 2017;46:323-326.
69. Moleriu LC, Sas I, RaTiu AC, et al. HPV genotyping and p16/Ki-67 dual staining in the detection of high-grade cervical lesion in patients with LSIL on Pap smear. *Rom J Morphol Embryol*. 2017;58:433-437.
70. Abreu AL, Souza RP, Gimenes F, Consolaro ME. A review of methods for detect human Papillomavirus infection. *Virol J*. 2012; 9:262.
71. Verdoodt F, Szarewski A, Halfon P, Cuschieri K, Arbyn M. Triage of women with minor abnormal cervical cytology: meta-analysis of the accuracy of an assay targeting messenger ribonucleic acid of 5 high-risk human papillomavirus types. *Cancer Cytopathol*. 2013;121:675-687.
72. Arbyn M, Roelens J, Cuschieri K, et al. The APTIMA HPV assay versus the Hybrid Capture II test in triage of women with ASC-US or LSIL cervical cytology: a meta-analysis of the diagnostic accuracy. *Int J Cancer*. 2013;132:101-108.
73. Trunk MJ, Dallenbach-Hellweg G, Ridder R, et al. Morphologic characteristics of p16INK4a-positive cells in cervical cytology samples. *Acta Cytol*. 2004;48:771-782.
74. Arbyn M, Kyrgiou M, Simoens C, et al. Peri-natal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: a meta-analysis. *BMJ*. 2008;337:a1284.
75. Arbyn M, Depuydt C, Benoy I, et al. VALGENT: a protocol for clinical validation of human papillomavirus assays. *J Clin Virol*. 2016;76(suppl 1):S14-S21.
76. Arbyn M, Peeters E, Benoy I, et al. VALHUDES: a protocol for validation of human papillomavirus assays and collection devices for HPV testing on self-samples and urine samples. *J Clin Virol*. 2018;117:52-56.