Modern Pathology (2017) 30, 1021–1031

Evaluation of p16/Ki-67 dual-stained cytology as triage test for high-risk human papillomavirus-positive women

Renée MF Ebisch[1], Judith van der Horst[2], Meyke Hermsen[2], L Lucia Rijstenberg[2], Judith EM Vedder[2], Johan Bulten[2], Remko P Bosgraaf[1], Viola MJ Verhoef[3], Daniëlle AM Heideman[3], Peter JF Snijders[3], Chris JLM Meijer[3], Folkert J van Kemenade[4], Leon FAG Massuger[1], Willem JG Melchers[5], Ruud LM Bekkers[1,6] and Albert G Siebers[2]

[1]Department of Obstetrics and Gynaecology, Radboud University Medical Center, Nijmegen, The Netherlands; [2]Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands; [3]Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands; [4]Department of Pathology, Erasmus University Medical Center, Rotterdam, The Netherlands; [5]Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, The Netherlands and [6]Department of Obstetrics and Gynaecology, Catharina Hospital, Eindhoven, The Netherlands

The aim of this study was to evaluate the clinical utility of p16/Ki-67 dual staining, for the identification of CIN in high-risk HPV-positive women from a non-responder screening cohort. P16/Ki-67 dual staining, Pap cytology, and HPV16/18 genotyping were performed on physician-taken liquid-based samples from 495 women who tested high-risk HPV positive on self-sampled material (PROHTECT-3B study). Different triage strategies involving p16/Ki-67 dual staining were evaluated for sensitivity, specificity, and predictive value for ≥CIN2 and ≥CIN3, and compared to Pap cytology with a threshold of atypical cells of undetermined significance. Centrally revised histology or an adjusted endpoint with combined high-risk HPV negative and cytology negative follow-up at 6 months was used as gold standard. Pap cytology (threshold atypical cells of undetermined significance) triage of high-risk HPV-positive samples showed a sensitivity of 93% (95% confidence interval: 85–98) with a specificity of 49% (95% confidence interval: 41–56) for ≥CIN3. Three triage strategies with p16/Ki-67 showed a significantly increased specificity with similar sensitivity. P16/Ki-67 triage of all high-risk HPV-positive samples had a sensitivity of 92% (95% confidence interval: 84–97) and a specificity of 61% (95% confidence interval: 54–69) for ≥CIN3. Applying p16/Ki-67 triage to only high-risk HPV-positive women with low-grade Pap cytology showed a similar sensitivity of 92% (95% confidence interval: 84–97), with a specificity for ≥CIN3 of 64% (95% confidence interval: 56–71). For high-risk HPV-positive women with low-grade and normal Pap cytology, triage with p16/Ki-67 showed a sensitivity of 96% (95% confidence interval: 89–99), and a specificity of 58% (95% confidence interval: 50–65). HPV16/18 genotyping combined with Pap cytology showed a sensitivity and specificity for ≥CIN3 similar to Pap cytology with an atypical cells of undetermined significance threshold. Because the quality of Pap cytology worldwide varies, and differences in sensitivity and specificity are limited between the three selected strategies, p16/Ki-67 triage of all high-risk HPV-positive samples would be the most reliable strategy in triage of high-risk HPV-positive women with an increased specificity and similar sensitivity compared with Pap cytology triage.
Modern Pathology (2017) **30,** 1021–1031; doi:10.1038/modpathol.2017.16; published online 17 March 2017

The introduction of cytology-based organized cervical cancer screening programs has contributed to decreased cervical cancer incidence and mortality in developed countries.[1–3] Compared to cytology, human papillomavirus (HPV) DNA testing has a higher sensitivity with a higher negative predictive value for detection of cervical intraepithelial neoplasia (CIN) and cancer. The high reassurance of a low risk of cervical cancer for high-risk HPV-negative women is one of the advantages of the HPV DNA test, which has resulted in a shift from cytology-based screening towards HPV DNA detection as primary screening method.[4,5]

Correspondence: RMF Ebisch, MD, Department of Obstetrics and Gynaecology, Radboud University Medical Center, PO Box 9101, Nijmegen 6500 HB, The Netherlands.
E-mail: renee.ebisch@radboudumc.nl
Received 13 December 2016; revised 2 February 2017; accepted 3 February 2017; published online 17 March 2017

However, the high sensitivity of HPV DNA testing is combined with a lower positive predictive value, due to the fact that most high-risk HPV infections clear spontaneously and do not result in cancer.[6] Additional triage of high-risk HPV-positive women is therefore required to limit the number of unnecessary referrals for follow-up procedures in women without clinically meaningful high-risk HPV infections.[7] Currently proposed triage strategies for high-risk HPV-positive women in HPV DNA-based screening programs are repeated Pap cytology and/or HPV16 and HPV18 genotyping. Pap cytology is a relatively subjective test for which high expertise is required. Owing to its limited sensitivity repeat cytology testing within 6–12 months is needed before returning high-risk HPV-positive women with normal cytology back to regular screening. This bears the risk of losing them during follow up.[8] HPV16/18 genotyping is an objective test to triage high-risk HPV-positive women, however, this strategy alone yields limited sensitivity as it only identifies cervical lesions associated with these two high-risk HPV types. Sensitivity of HPV16/18 genotyping can be improved by combining it with Pap cytology at the cost of a lower specificity.[9–11]

Another biomarker widely studied for triage is the p16/Ki-67 dual staining. P16 ^{INK4a} (p16) is a cell-cycle regulatory protein that induces cell-cycle arrest under normal physiological conditions, and Ki-67 is a marker expressed during cell proliferation.[12,13] The simultaneous detection of p16 and Ki-67 within the same cervical epithelial cell will not occur under physiological conditions and may be used as a surrogate marker of cell-cycle deregulation mediated by transforming high-risk HPV infections. P16/Ki-67 dual-stain has been previously studied as a potential primary screening test, as triage test for women with atypical cells of undetermined significance, for surveillance of women treated for high-grade CIN, and also in a limited number of studies as triage test in women with a positive high-risk HPV test.[14–22]

The aim of the current study was to evaluate the overall clinical performance of the p16/Ki-67 dual-stain test as triage method for high-risk HPV-positive women from a non-responder screening cohort.

Materials and methods

Study Population and Specimens

This study is a post-hoc analysis on physician-taken triage cervical scrapes of former non-responder women who were recruited into the screening program by offering self-sampling for high-risk HPV testing in the PROHTECT-3B study in 2011 and 2012. (PRotection by Offering HPV TEsting on self-sampled Cervicovaginal specimens Trial-3B).[23] In this trial, non-responders to the regular cervical screening program, aged 30–60 years, were invited to participate by returning self-sampled material to the laboratory for high-risk HPV testing (GP5+/6+ polymerase chain reaction; EIA HPV GP HR kit; Diassay, Rijswijk, The Netherlands). Women who tested high-risk HPV positive on their self-sample were advised to have an additional cervical smear taken by a physician for Pap cytology testing. Women with an abnormal Pap smear were referred to a gynecologist for colposcopic examination, and for women with a normal Pap smear a 6-month follow-up smear was performed for both high-risk HPV testing and Pap cytology. Further details of the PROHTECT-3B study design are reported elsewhere.[23] All women provided written informed consent. The Ministry of Health gave ethical approval for this study (No. 2010/WBO04), and the regional institutional review board approved the protocol for this *post-hoc* analysis.

From 495 of the total of 834 high-risk HPV-positive women in the PROHTECT-3B study, a study-endpoint was known, and a physician-taken triage liquid-based cytology cervical scrape was available. Women with abnormal cytology results (defined as $\geq$ atypical cells of undetermined significance) were referred for a colposcopy-directed biopsy, whereas women with normal cytology results (defined by negative for intraepithelial lesion or malignancy cytology result) were re-invited for an exit test with Pap cytology and high-risk HPV co-testing 6 months later. Women with a positive exit test, defined as atypical cells of undetermined significance or worse ($\geq$ atypical cells of undetermined significance) cytology and/or high-risk HPV-positive test results, were referred for a colposcopy-directed biopsy. Colposcopists were aware of the high-risk HPV-positive status and colposcopy was performed according to the Dutch national guidelines. If no abnormalities were seen at colposcopy, it was advised to take two random biopsies according to the study protocol. Women with a double-negative exit test (negative for intraepithelial lesion or malignancy cytology and negative high-risk HPV results) after 6 months were considered to have a minimal risk of $\geq$CIN2 lesions and were not referred for colposcopy; these women were classified as not having CIN2 ($<$CIN2). We included all results recorded before June 2013. At this point, the database was closed with a mean follow-up of 15 months (range: 6–18 months).

Pap Cytology

All liquid-based cytology samples were processed and reported in the laboratory of the department of Pathology, Radboud university medical center, Nijmegen, The Netherlands. The ThinPrep 3000 was used for processing, and cytological classification was performed according to the primarily used Dutch CISOE-A classification. For analysis of cervical smears, the CISOE-A classification system was translated into the Bethesda nomenclature; in which

borderline or mild dyskaryosis equals atypical cells of undetermined significance and low-grade squamous intraepithelial lesion, and worse than borderline or mild dyskaryosis equals high-grade squamous intraepithelial lesion.[24] All abnormal cytology was analyzed independently by two cytotechnicians who were aware of the high-risk HPV-positive status, but unaware of the p16/Ki-67 and HPV16/18 genotype results.

Histology and Centralized Revision

Histological results were obtained from records of the pathologists, and missing data were retrieved from the Dutch nationwide computerized registry of histopathology and cytopathology (PALGA).[25] The histology outcomes were classified as CIN2 or worse (≥CIN2) or CIN3 or worse (≥CIN3), CIN1, or no CIN, AIS was included in the CIN3 group. When multiple results were registered per woman, the most severe histological diagnosis was used for analysis. Histology samples collected during colposcopy procedures were subjected to central pathologist review. If the first pathologist disagreed with the initial diagnosis, a pathologist specialized in gynecologic oncology independently reviewed the case resulting in a final diagnosis. Majority consensus diagnoses were established on all available cervical tissue specimens. Pathologists were blinded to all other study results.

HPV16/18 Genotyping

Partial HPV genotyping was done by the Roche Cobas 4800 test, according to the manufacturer's recommendations in the laboratory of the Department of Medical Microbiology, Radboud university medical center.[26] This test provides separate result for HPV16 and 18, and a pool of 12 other high-risk HPV types (ie, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68). The HPV genotyping results were categorized as HPV16/18 positive when HPV16 and/or HPV18 were present, regardless of the presence or absence of 12 other HPV types. Other results were scored as non-16/18 HPV positive.

p16/Ki-67 Dual-Stain Cytology

After HPV testing and Pap cytology testing the residual liquid-based cytology material was stored for approximately 2 years in the original ThinPrep vials (Hologic UK, Crawley, West Sussex, UK) at room temperature (storage between 4 and 30° C is advised by the manufacturer). A slide for p16/Ki-67 was prepared from the residual PreservCyt material using the ThinPrep 2000 Processor (Hologic). The CINtec PLUS Cytology kit (Roche mtm Laboratories AG, Mannheim, Germany) was used according to the instructions of the manufacturer. Staining was performed on a Ventana benchmark ultra (Roche mtm Laboratories AG), and each run included one control specimen.

Six trained observers each independently reviewed one-third of the cases for the presence of dual-stained cells, resulting in two independent results for each case. A case was considered positive if one or more cervical epithelial cells were stained both with a red nuclear stain for Ki-67 and a brown cytoplasmatic stain for p16, independent of

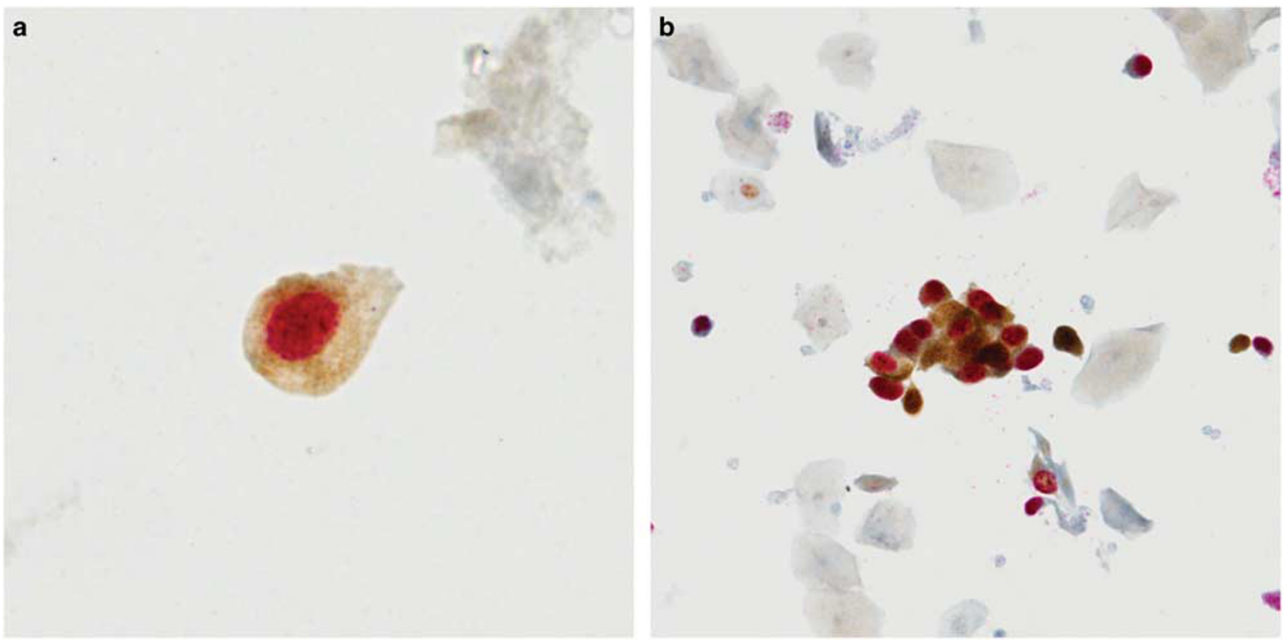

Figure 1 Example of p16/Ki-67 dual staining in cervical cytology. (**a**) A p16/Ki-67 dual-stain positive single cell. (**b**) A p16/Ki-67 dual-stained cluster of cells. A case is considered dual-stain positive if one or more cervical epithelial cells are stained with a red nuclear stain for Ki-67 and a brown cytoplasmatic stain for p16.

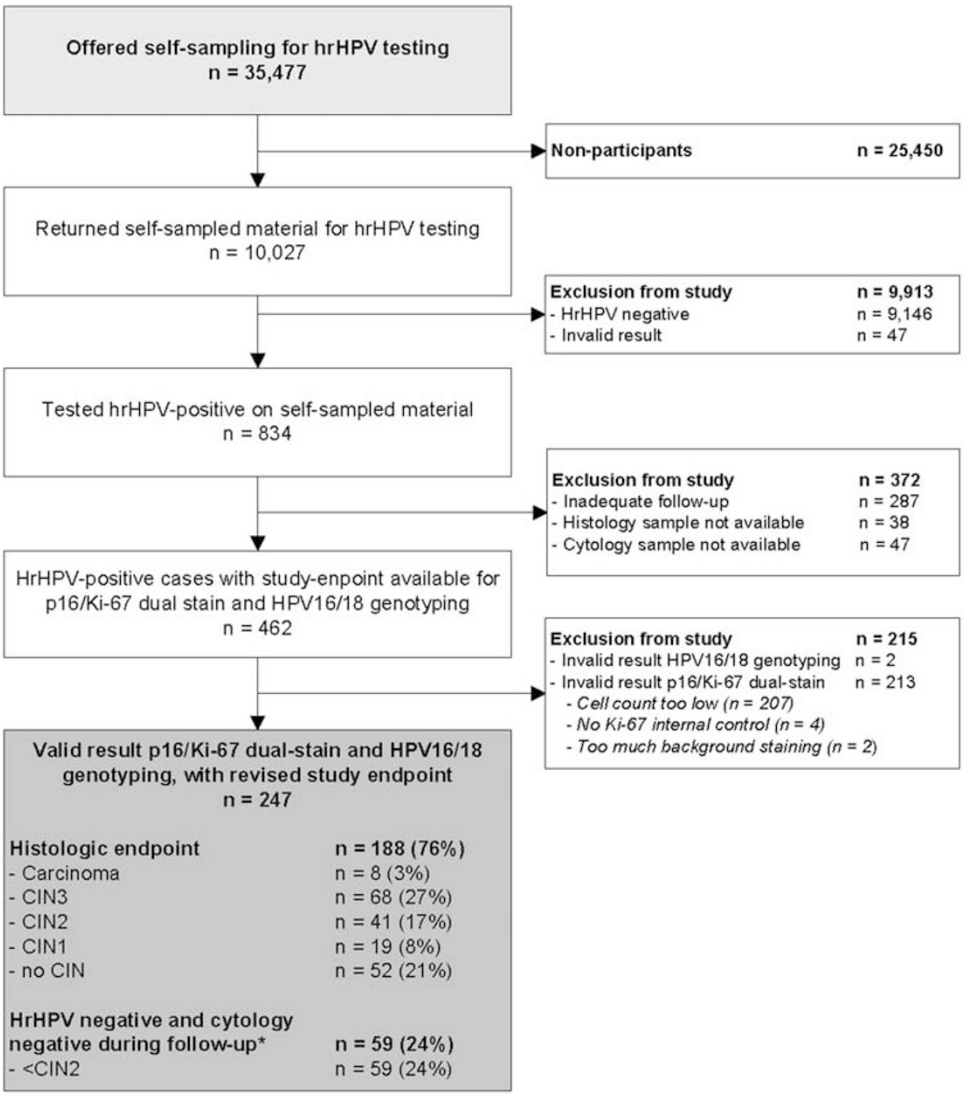

Figure 2 Trial Profile.*High-risk HPV negative and cytology negative during follow up; women with an initial cytology test result negative for intraepithelial lesion or malignancy, and after 6 months a negative for intraepithelial lesion or malignancy result and a high-risk HPV-negative test are considered to have less than CIN2 (< CIN2).

cellularity criteria (Figure 1). Cases were considered negative when no staining or only single staining of p16 or Ki-67 was observed in a single cell. A case was scored as inadequate if background staining prohibited adequate evaluation, no p16 and/or Ki-67 staining was visible as internal control, or if slides did not meet the squamous cellularity criteria as specified in the Hologic criteria (≥5000 cells per slide). Inadequate cases were excluded from evaluation. In case of disagreement between two observers, a consensus score using a multi-headed microscope was obtained. The observers were unaware of the Pap cytology result, HPV16/18 genotyping result, or follow-up data. Our data on p16/ki-67 triage of high-risk HPV-positive women was compared to previous studies found in a systematic search combining synonyms for p16/Ki-67 dual-stain and HPV. Information on sensitivity, specificity, number of participants, HPV status, cytology result, and p16/Ki-67 stain was obtained from relevant studies.

Statistical Analysis

Descriptive statistics were used to calculate sensitivities, specificities, predictive values with corresponding 95% confidence intervals, and referral rate. The extent of overdiagnosis was estimated by using the number of referrals needed to diagnose one lesion with endpoint ≥CIN2 or ≥CIN3, this equals by dividing one by the positive predictive value. Differences in sensitivity and specificity between the reference strategy and the other strategies were evaluated with McNemar's χ^2 test with continuity correction. Statistical analyses were conducted using SPSS (Statistical Package for the Social Sciences)

Table 1 Test results of p16/Ki-67 dual-stain in regard to Pap cytology and HPV16/18 genotyping

	Pap cytology			
	NILM	*ASC-US and LSIL*	*HSIL*	*Total*
p16/Ki-67 positive				
HPV16/18 positive	4	18	57	79 (32%)
HPV16/18 negative	9	23	25	57 (23%)
Total	13	41	82	136 (55%)
p16/Ki-67 negative				
HPV16/18 positive	10	7	3	20 (8%)
HPV16/18 negative	65	20	6	91 (37%)
Total	75	27	9	111 (45%)
Total	88 (36%)	68 (27%)	91 (37%)	247 (100%)

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; HPV, human papillomavirus; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignancy.

database, version 20.0.1 for Windows (Chicago, IL, USA). Differences in sensitivity and specificity were estimated with R, version 3.0.1, package DTComPair (Vienna, Austria).[27]

Results

Population

In the PROHTECT-3B study, 35 477 women were offered self-sampling for high-risk HPV testing. Out of 10 027 women who participated in the study, 834 (8.3%) were high-risk HPV positive on a self-sample. Of the 834 high-risk HPV-positive women, 287 were excluded because of inadequate follow up, 47 because no liquid-based cytology left-over sample was available, and 38 women because no histology sample was available for revision, leaving 462 cases available for p16/Ki-67 dual-stain cytology and high-risk HPV genotyping (Figure 2). The mean age of the study group was 42 years, with a median age of 38 years (range: 33–63 years).

From the total group of 462 cases available for p16/Ki-67 dual staining, 136 were positive, 113 negative, and 213 were inadequate. In 207 of the latter, the cell count was too low, in 4 cases no Ki-67 internal control was visible, and in 2 cases the background staining made adequate evaluation impossible. The low cell count in the majority of cases was probably due to the fact that too little of the sample material was left after previous tests. Of the remaining 249 women, 2 had an inadequate genotyping result; leaving 247 cases available for analysis. There was no difference in age between the included and excluded women with a median age of 38 years in both groups and a mean age of 42 years and 41 years, respectively. The excluded group contained more low-grade and normal histology results. Resulting in 47% ≥CIN2 and 71% CIN1 or less (≤CIN1) in the included group, and 29% ≥CIN2 and 53% ≤CIN1 in the excluded group.

Test Results and Endpoints

From the total of 247 women with a valid result for p16/Ki-67 dual-stain and HPV16/18 genotyping, 159 (64%) women had an abnormal cytology result of atypical cells of undetermined significance or worse, 136 women (55%) had a p16/Ki-67 positive score, and 99 women (40%) were positive for HPV16/18 (Table 1). Of the total group of 247 women, 188 women had a revised histological endpoint; 8 women had a cervical carcinoma, 68 were diagnosed with CIN3, 41 with CIN2, 19 with CIN1, and 52 with no CIN. The remaining 59 women were considered to have no CIN, based on both a high-risk HPV negative and negative for intraepithelial lesion or malignancy cytology result after 6 months (Figure 2). Revision of the 188 histology results yielded a similar result in 153 (81%), cases and a different result in 35 cases (19%) Revision leaded to shift of 3 cases scored as ≥CIN2 towards <CIN2, and 12 cases scored as <CIN2, towards ≥CIN2, resulting in a 3.6% higher ≥CIN2 prevalence after revision.

Performance of Triage Strategies in High-Risk HPV-Positive Women

In this *post-hoc* analysis of the PROHTECT-3b study, first the performance of six baseline triage strategies by p16/Ki-67 dual-stain, Pap cytology and HPV16/18 genotyping was explored with endpoints ≥CIN2 and ≥CIN3. Baseline triage with cytology testing with an atypical cells of undetermined significance threshold was used as reference strategy. This strategy yielded a sensitivity of 93% (95% confidence interval: 85–98) with a specificity of 49% (95% confidence interval: 41–56), positive predictive value of 45% (95% confidence interval: 37–53), and negative predictive value of 94% (95% confidence interval: 87–98) for ≥CIN3. Five women with CIN3 were missed with this strategy. P16/Ki-67 dual staining showed a similar sensitivity of 92% (95% confidence interval: 84–97), with an increased specificity of 61% (95% confidence interval: 54–69) for ≥CIN3. Six women with CIN3 were missed with this strategy. HPV16/18 genotyping alone showed a significantly lower sensitivity of 75% (95% confidence interval: 64–97), with 19 missed CIN3 cases, and a significant improvement in specificity of 75% (95% confidence interval: 68–82) for ≥CIN3 with a referral rate of only 40%.

Three strategies showed similar sensitivity with improved specificity, compared to the reference strategy for ≥CIN3. Similar was defined by a non-significant difference from Pap cytology triage. The first strategy was p16/Ki-67 triage of all high-risk

Table 2 Clinical performance of p16/Ki-67 dual-stain and HPV16/18 genotyping in triage or hrHPV-positive women to detect ≥CIN3

	TP	*FP*	*TN*	*FN*	*Sensitivity*		*Specificity*		*PPV*	*NPV*	*RR (%)*	*NRND ratio*
					% (95% CI)	P-*value*	*% (95% CI)*	P-*value*	*% (95% CI)*	*% (95% CI)*		
Pap cytology ≥ ASC-US	71	88	83	5	93 (85–98)	REF	49 (41–56)	REF	45 (37–53)	94 (87–98)	64	2.2
p16/Ki-67 dual-stain	70	66	105	6	92 (84–97)	1	61 (54–69)	< 0.01	52 (43–60)	95 (89–98)	55	1.9
HPV16/18 genotyping	57	42	129	19	75 (64–84)	< 0.01	75 (68–82)	< 0.01	58 (47–68)	87 (81–92)	40	1.7
p16/Ki-67 and/or HPV16/18 genotyping	74	82	89	2	97 (91–100)	0.37	52 (44–60)	0.45	47 (39–56)	98 (92–100)	63	2.1
Pap cytology, with p16/Ki-67 triage of ASC-US and LSIL[a]	70	62	109	6	92 (84–97)	1	64 (56–71)	< 0.01	53 (44–62)	95 (89–98)	53	1.9
Pap cytology, with p16/Ki-67 triage of NILM[b]	74	98	73	2	97 (91–100)	0.25	43 (35–51)	< 0.01	43 (36–51)	97 (91–100)	70	2.3
Pap cytology, with p16/Ki-67 triage of NILM, ASC-US, and LSIL[c]	73	72	99	3	96 (89–99)	0.62	58 (50–65)	0.01	50 (42–59)	97 (92–99)	59	2

Abbreviations: ≥ ASC-US, atypical squamous cells of undetermined significance or worse; CI, confidence interval; ≥ CIN3, cervical intraepithelial neoplasia grade 3 or worse; FN, false negatives; FP, false positives; hrHPV, high-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignancy; NPV, negative predictive value; NRND, number of referrals needed to diagnose one ≥ CIN3 lesion; PPV, positive predictive value; REF, reference standard; RR, referral rate; TN, true negatives; TP, true positives.
[a]Pap cytology as triage method for hrHPV-positive women, with p16/Ki-67 triage of ASC-US and LSIL Pap cytology results. ASC-US and LSIL Pap cytology, p16/Ki-67 positive samples were scored as positive, and ASC-US and LSIL Pap cytology, p16/Ki-67 negative were scored as negative.
[b]Pap cytology as triage method for hrHPV-positive women, with p16/Ki-67 triage of NILM Pap cytology results. NILM Pap cytology, p16/Ki-67 positive samples were scored as positive, and NILM Pap cytology, p16/Ki-67 negative were scored as negative.
[c]Pap cytology as triage method for hrHPV-positive women, with p16/Ki-67 triage of NILM, ASC-US and LSIL Pap cytology results. NILM, ASC-US and LSIL Pap cytology, p16/Ki-67 positive samples were scored as positive, and NILM, ASC-US, and LSIL Pap cytology, p16/Ki-67 negative were scored as negative.

HPV-positive women. The second strategy was baseline cytology with p16/Ki-67 triage restricted to high-risk HPV-positive women with atypical cells of undetermined significance or low-grade squamous intraepithelial lesion cytology, which showed a sensitivity of 92% (95% confidence interval: 84–97), with the highest specificity of 64% (95% confidence interval: 56–71), and a false negative result for 6 women with CIN3. The third strategy was 16/Ki-67 triage of high-risk HPV-positive women with negative for intraepithelial lesion or malignancy, atypical cells of undetermined significance, or low-grade squamous intraepithelial lesion cytology, resulted in an equal sensitivity of 96% (95% confidence interval: 89–99) with a specificity of 58% (95% confidence interval: 50–65), and three false negative cases for CIN3. Combined p16/Ki-67 dual staining with HPV16/18 genotyping showed a similar sensitivity, with only two missed CIN3 lesions, and a similar specificity, compared with Pap cytology triage. Adding p16/Ki-67 triage to high-risk HPV-positive women with normal cytology detected four women with CIN3, and another eight with CIN2, however, at the cost of 32 and 24 unnecessary referrals, respectively (Table 2).

For ≥CIN2 the reference cytology strategy showed a sensitivity of 94% (95% confidence interval: 88–98) with a specificity of 62% (95% confidence interval: 53–71), positive predictive value of 69% (95% confidence interval: 61–76) and negative predictive value of 92% (95% confidence interval: 84–97). The strategy combining p16/Ki-67 triage of atypical cells of undetermined significance and low-grade squamous intraepithelial lesion cytology also yielded a similar sensitivity but with a statistically significant increase in specificity for ≥CIN2, compared to Pap cytology triage. The sensitivity of this strategy was 89% (95% confidence interval: 82–94) with a specificity of 79% (95% confidence interval: 70–85; Table 3).

Our systematic search yielded five studies that previously reported on triage of high-risk HPV-positive women with all cytology categories, representing 3270 women, results are summarized in Table 4.[22,28–31] Results of these studies show sensitivities ranging from 83 to 93% for ≥CIN2 and 87 to 95% for ≥CIN3. Specificities range from 53 to 75% for ≥CIN2 and 48 to 57% for ≥CIN3. Only the specificity for ≥CIN3 found in this study falls outside the range, but is not significantly higher compared with previously published numbers.

Evaluation of Reproducibility

The overall κ-value of dual-stain cytology for comparing scores of two independent evaluators was 0.67 (95% confidence interval: 0.59–0.76), which is considered substantial according to the classification of Landis and Koch.[32]

Discussion

In this study we evaluated the clinical utility of p16/Ki-67 dual staining, either or not combined with

Table 3 Clinical performance of p16/Ki-67 dual-stain and HPV16/18 genotyping in triage of hrHPV-positive women to detect ≥CIN2

	TP	*FP*	*TN*	*FN*	*Sensitivity*		*Specificity*		*PPV*	*NPV*	*RR (%)*	*NRND ratio*
					% (95% CI)	P-*value*	*% (95% CI)*	P-*value*	*% (95% CI)*	*% (95% CI)*		
Pap cytology ≥ ASC-US	110	49	81	7	94 (88–98)	REF	62 (53–71)	REF	69 (61–76)	92 (84–97)	64	1.5
p16/Ki-67 dual-stain	101	35	95	16	86 (79–92)	0.04	73 (65–81)	0.03	74 (66–81)	86 (78–92)	55	1.4
HPV16/18 genotyping	77	22	108	40	66 (57–74)	< 0.01	83 (76–89)	< 0.01	78 (68–86)	73 (65–80)	40	1.3
p16/Ki-67 and/or HPV16/18 genotyping	110	46	84	7	94 (88–98)	1	65 (56–73)	0.75	71 (63–78)	92 (85–97)	63	1.4
Pap cytology, with p16/Ki-67 triage of ASC-US and LSIL[a]	104	28	102	13	89 (82–94)	0.41	79 (70–85)	< 0.01	79 (71–85)	89 (82–94)	53	1.3
Pap cytology, with p16/Ki-67 triage of NILM[b]	113	59	71	4	97 (92–99)	0.25	55 (46–63)	< 0.01	66 (58–73)	95 (87–99)	70	1.5
Pap cytology, with p16/Ki-67 triage of NILM, ASC-US, and LSIL[c]	107	38	92	10	92 (85–96)	0.51	71 (62–78)	0.07	74 (66–81)	90 (83–95)	59	1.4

Abbreviations: ≥ ASC-US, atypical squamous cells of undetermined significance or worse; CI, confidence interval; ≥CIN2, cervical intraepithelial neoplasia grade 2 or worse; FN, false negatives; FP, false positives; hrHPV, high-risk human papillomavirus; LSIL, low-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignancy; NPV, negative predictive value; NRND, number of referrals needed to diagnose one ≥CIN2; PPV, positive predictive value; REF, reference standard; RR, referral rate; TN, true negatives; TP, true positives.
[a]Pap cytology as triage method for hrHPV-positive women, with p16/Ki-67 triage of ASC-US and LSIL Pap cytology results. ASC-US and LSIL Pap cytology, p16/Ki-67 positive samples were scored as positive, and ASC-US and LSIL Pap cytology, p16/Ki-67 negative were scored as negative.
[b]Pap cytology as triage method for hrHPV-positive women, with p16/Ki-67 triage of NILM Pap cytology results. NILM Pap cytology, p16/Ki-67 positive samples were scored as positive, and NILM Pap cytology, p16/Ki-67 negative were scored as negative.
[c]Pap cytology as triage method for hrHPV-positive women, with p16/Ki-67 triage of NILM, ASC-US and LSIL Pap cytology results. NILM, ASC-US and LSIL Pap cytology, p16/Ki-67 positive samples were scored as positive, and NILM, ASC-US, and LSIL Pap cytology, p16/Ki-67 negative were scored as negative.

Pap cytology and/or HPV16/18 genotyping, for the identification of ≥CIN2 and ≥CIN3 in high-risk HPV-positive women from a non-responder screening cohort. Three of the proposed strategies for triaging high-risk HPV-positive women showed increased specificity with similar sensitivity for ≥CIN3, compared to Pap cytology. These strategies were; p16/Ki-67 triage of all high-risk HPV-positive women, p16/Ki-67 triage of high-risk HPV-positive women with atypical cells of undetermined significance or low-grade squamous intraepithelial lesion cytology, and p16/Ki-67 triage of high-risk HPV-positive women with negative for intraepithelial lesion or malignancy, atypical cells of undetermined significance or low-grade squamous intraepithelial lesion cytology. With a ≥CIN2 threshold only the strategy with p16/Ki-67 triage of high-risk HPV-positive women with atypical cells of undetermined significance or low-grade squamous intraepithelial lesion cytology showed increased specificity with similar sensitivity compared to Pap cytology triage of high-risk HPV-positive women.

Our findings on sensitivity and specificity of p16/Ki-67 triage of high-risk HPV-positive women independent of cytology result are comparable with previously published studies, with only small differences in sensitivity and specificity between Pap cytology and p16/Ki-67 dual-stain. These studies have been performed with different high-risk HPV tests and different cell collection medium. Also, the populations in these studies are different, some are performed in a general population, others in a outpatient population, and none were previously performed in a non-responder population, which could explain slightly different results. A recent large study on p16/Ki-67 performance in high-risk HPV-positive women by Wentzensen and colleagues shows an increased specificity with a maintained sensitivity for ≥CIN3 detection, compared with Pap cytology in triage of high-risk HPV-positive women.[29] This was also confirmed by Luttmer *et al*[22] who show a good clinical performance of p16/Ki-67 dual-stained cytology as triage method for high-risk HPV-positive women with an increase in sensitivity and specificity for ≥CIN3. A previously performed study on p16/Ki-67 triage of high-risk HPV-positive women with normal cytology showed that p16/Ki-67 dual-stained cytology detects more than 70% of underlying ≥CIN3 lesions in high-risk HPV-positive women with normal cytology at baseline. They conclude this strategy is suitable for triaging these women to colposcopy.[33] Our study also confirms the additional detection of high-grade lesions in high-risk HPV-positive women with a normal cytology result, however at the cost of additional colposcopy referrals.

Previous studies have also analyzed the clinical value of p16/Ki-67 triage of women with low-grade or normal cytology in high-risk HPV-positive cohorts. To our knowledge, none of them show results on sensitivity, specificity and predictive values of overall triage strategies with p16/Ki-67 triage of certain Pap cytology subgroups in high-risk HPV-positive women. This approach gives an

Table 4 Characteristics of the included studies

Study	Included patients	Cytology technique	HPV test	p16/Ki-67 (≥ CIN2)		Pap cytology (≥ CIN2)		p16/Ki-67 (≥ CIN3)		Pap cytology (≥ CIN3)	
				Sensitivity % (95% CI)	*Specificity % (95% CI)*	*Sensitivity % (95% CI)*	*Specificity % (95% CI)*	*Sensitivity % (95% CI)*	*Specificity % (95% CI)*	*Sensitivity % (95% CI)*	*Specificity % (95% CI)*
Allia *et al*[30]	477	LBC ThinPrep	HC2	NA	64 (57–70)	NA	NA	NA	NA	NA	NA
Gustinucci *et al*[28]	375	Cobas PCR medium	Cobas	88 (76–94)	75 (69–79)	78 (65–87)	73 (67–77)	92 (75–99)	NA	96 (81–100)	NA
Luttmer *et al*[22]	446	LBC ThinPrep	GP5+/6+	86 (80–91)	60 (54–66)	87 (82–92)	54 (49–60)	94 (89–99)	51 (46–56)	88 (80–95)	45 (40–50)
Wentzensen *et al*[29]	1509	LBC Surepath	HC2	83 (77–89)	59 (56–62)	77 (70–83)	50 (47–52)	87 (79–93)	57 (54–60)	84 (75–91)	49 (46–51)
Yu *et al* [31]	463	LBC ThinPrep	Cobas	93 (88–95)	53 (46–59)	95 (91–97)	54 (47–60)	95 (91–97)	48 (69–80)	98 (95–99)	49 (43–55)
Ebisch	247	LBC ThinPrep	GP5+/6+	86 (79–92)	73 (65–81)	94 (88–98)	62 (53–71)	92 (84–97)	61 (54–69)	93 (85–98)	49 (41–56)

Abbreviations: CI, confidence interval; CIN, cervical intraepithelial neoplasia; HC, Hybrid Capture 2; HPV, human papillomavirus; LBC, liquid-based cytology; NA, not available; PCR, polymerase chain reaction.

overview of the whole triage strategy, instead of only results on triage of a certain cytology subgroup. Our data show that adding p16/Ki-67 triage to either low-grade only, or alternatively, low-grade and normal Pap cytology in high-risk HPV-positive women would improve the specificity of the triage step, while maintaining similar sensitivity. With improved specificity by additional p16/Ki-67 triage of these low-grade and/or normal cytology results, unnecessary colposcopy referrals could be prevented. By adding p16/Ki-67 triage for women with atypical cells of undetermined significance or low-grade squamous intraepithelial lesion cytology, the referral rate can be lowered from 64 to 53% with a decrease in number of referrals needed to diagnose one ≥CIN3 lesion from 2.2 to 1.9. These strategies can easily be combined with Pap cytology triage of high-risk HPV-positive women which will be used in the new Dutch cervical cancer screening program which started in January 2017.

P16/Ki-67 dual-stain was also combined with HPV16/18 genotyping. An attractive feature of HPV16/18 genotyping is that this triage method could be implemented without additional costs, as most high-risk HPV tests also have the ability of immediate HPV16/18 genotyping. In this study, triage with HPV16/18 genotyping shows an increase in specificity, however at the cost of a lower sensitivity.

An advantage of p16/Ki-67 dual-stain over Pap cytology is the reduced role of morphology. It has been previously shown that p16/Ki-67 dual-stain shows substantial to good reproducibility with almost identical performance by novice evaluators compared with reference evaluations, indicating that it can be implemented in clinical practice with limited training.[34,35] The Kappa found in this study was lower as expected, this might be caused by the overall low quality of the samples. A disadvantage of the technique is that it cannot be reliably used on self-sampled material, as cellularity of cervical indicator cells in these samples is too limited resulting in a low sensitivity.[36] Women testing high-risk HPV positive on self-samples would therefore still need to visit their doctor for an additional clinician-taken cervical smear for triage. As such, in this setting direct triage strategies applicable to self-samples like HPV16/18 or other biomarkers such as DNA methylation analysis are preferred.[7,37] Dual staining neither rules-out the need for follow up, and in case the technique is used as additional triage tool after Pap-stained cytology triage, this would result in additional triage-costs. A wide variety of novel triage tests for high-risk HPV-positive women is currently being extensively studied for triage purposes. Molecular techniques based on host- and viral DNA methylation markers, and differences in gene-expression can be used as triage method in the future, possibly even with improved predictive characteristics.[38–40] Most of these markers have not yet been sufficiently validated to be ready for

implementation in screening program, but among them p16/Ki-67 dual-stained cytology or host cell DNA methylation analysis, with or without additional HPV16/18 genotyping, are attractive options for the near future.[41]

An important strength of the study is that the study design is comparable to future cervical screening programs with primary high-risk HPV testing. The potential bias of HPV knowledge increases cytology sensitivity and decreases cytology specificity, which gives more reason for adding p16/Ki-67 to triage of high-risk HPV-positive women.[9,42] A limitation of this study is the large number of samples that had to be excluded because of a limited cell count, most likely due to insufficient left-over material, or because of the longer shelf-life used in this study than advised by the manufacturer, which was within six weeks of collection.

As most of these samples are expected to be negative (because if a dual-stained cell was visible, the sample was scored as positive), the specificities found in this study might be an underestimation, with a slight overestimation of sensitivity. Another potential bias could be the fact that this study was performed in a non-responder population, this might result in slightly different results in a responder population with an expected lower high-risk HPV positivity rate.

The Netherlands and Australia will be among the first countries to initiate full high-risk HPV-based organized cervical cancer screening. It is expected that an increasing number of countries will also replace Pap cytology-based screening with high-risk HPV-based screening. The high sensitivity, reproducibility of the test, and possibility of high-throughput testing, are advantages of high-risk HPV-based screening. Triage with cytology is an obvious option because of the widespread knowledge on this technique. In line with others, our results indicate that the specificity of triage in high-risk HPV-based screening programs can be increased by replacing Pap cytology with p16/Ki-67 dual-stain, or adding dual-stain cytology as an additional triage step for low-grade Pap cytology.

Because the quality of Pap cytology worldwide varies, and two of the selected strategies are based on quality of Pap cytology, and differences in sensitivity and specificity are limited, it would be preferable to choose for primary p16/Ki-67 dual-stain triage of high-risk HPV-positive women. We therefore conclude that p16/Ki-67 dual-stained cytology of high-risk HPV-positive women shows increased specificity for $\geq$CIN3 with a maintained adequate sensitivity compared to Pap cytology in triage of high-risk HPV-positive women.

Acknowledgments

Screening Organizations Midden-West and Oost funded the PROHTECT-3B trial. We are grateful to all PROHTECT participants who contributed to this study. We would like to acknowledge all PROHTECT study team members, general practitioners, gynecologists, pathologists, technicians, secretarial services and the nationwide network and registry of histopathology and cytopathology in The Netherlands (PALGA) for their contribution. We would also like to acknowledge Lia van Zuylen-Manders, Rietje Huberts-Manders, and Marc Tomassen, who evaluated the p16/Ki-67 dual-stained samples. Roche Diagnostics Nederland, Almere, The Netherlands, provided the kits for p16/Ki-67 dual-stain and high-risk HPV genotyping. The source of funding did not have any influence on the design and the analysis of the results.

Disclosure/conflict of interest

DAMH has minority stake in Self-Screen B.V., a spin-off company of VU University Medical Center Amsterdam. PJFS has been on the speaker's bureau of Roche, Gen-Probe, Abbott, Qiagen, and Seegene, and is minority shareholder of Self-Screen B.V., a spin-off company of VUmc. CJLMM has received speakers fee from GSK, Qiagen, SPMSD/Merck, Roche, Menarini and Seegene, served occasionally on the scientific advisory board (expert meeting) of GSK, Qiagen, SPMSD/Merck, Roche and Genticel, and by occasion as consultant for Qiagen and Genticel. He holds minority stock of Self-Screen B. V., a spin-off company of VUMC and until March 2016 of Diassay B.V. Until 2014, he held a small number of certificates of shares in Delphi Biosciences, which went into receivership in 2014.

References

1 Peto J, Gilham C, Fletcher O, *et al.* The cervical cancer epidemic that screening has prevented in the UK. Lancet 2004;364:249–256.
2 Macgregor JE, Campbell MK, Mann EM, *et al.* Screening for cervical intraepithelial neoplasia in north east Scotland shows fall in incidence and mortality from invasive cancer with concomitant rise in preinvasive disease. BMJ 1994;308:1407–1411.
3 Louhivuori K. Effect of a mass screening program on the risk of cervical cancer. Cancer Detect Prev 1991;15: 471–475.
4 Rijkaart DC, Berkhof J, Rozendaal L, *et al.* Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. Lancet Oncol 2012;13:78–88.
5 Saslow D, Solomon D, Lawson HW, *et al.* American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin 2012;62:147–172.

6 Arbyn M, Sasieni P, Meijer CJ, *et al.* Chapter 9: Clinical applications of HPV testing: a summary of meta-analyses. Vaccine 2006;24:S3/78–89.
7 Ebisch RM, Siebers AG, Bosgraaf RP, *et al.* Triage of high-risk HPV positive women in cervical cancer screening. Expert Rev Anticancer Ther 2016;16: 1073–1085.
8 Castle PE, Sideri M, Jeronimo J, *et al.* Risk assessment to guide the prevention of cervical cancer. J Low Genit Trac Dis 2008;12:1–7.
9 Ebisch RM, de Kuyper-de Ridder GM, Bosgraaf RP, *et al.* The clinical value of HPV genotyping in triage of women with high-risk-HPV-positive self-samples. Int J Cancer 2016;139:691–699.
10 Dijkstra MG, van Niekerk D, Rijkaart DC, *et al.* Primary high-risk HPV DNA testing in cervical cancer screening: how to manage screen-positive women? A POBASCAM trial substudy. Cancer Epidemiol Biomarkers Prev 2014;23:55–63.
11 Rijkaart DC, Berkhof J, van Kemenade FJ, *et al.* Evaluation of 14 triage strategies for HPV DNA-positive women in population-based cervical screening. Int J Cancer 2012;130:602–610.
12 von Knebel Doeberitz M. New markers for cervical dysplasia to visualise the genomic chaos created by aberrant oncogenic papillomavirus infections. Eur J Cancer 2002;38:2229–2242.
13 Bosgraaf RP, Siebers AG, De Hullu JA, *et al.* The current position and the future perspectives of cervical cancer screening. Expert Rev Anticancer Ther 2014;14:75–92.
14 Ikenberg H, Bergeron C, Schmidt D, *et al.* Screening for cervical cancer precursors with p16/Ki-67 dual-stained cytology: results of the PALMS study. J Natl Cancer Inst 2013;105:1550–1557.
15 Petry KU, Schmidt D, Scherbring S, *et al.* Triaging Pap cytology negative, HPV positive cervical cancer screening results with p16/Ki-67 Dual-stained cytology. Gynecol Oncol 2011;121:505–509.
16 Schmidt D, Bergeron C, Denton KJ, *et al.* 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.
17 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.
18 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.
19 Killeen JL, Dye T, Grace C, *et al.* Improved abnormal Pap smear triage using cervical cancer biomarkers. J Low Genit Tract Dis 2014;18:1–7.
20 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.
21 Polman NJ, Uijterwaal MH, Witte BI, *et al.* Good performance of p16/Ki-67 dual-stained cytology for surveillance of women treated for high-grade CIN. Int J Cancer 2017;140:423–430.
22 Luttmer R, Dijkstra MG, Snijders PJ, *et al.* p16/Ki-67 dual-stained cytology for detecting cervical (pre)cancer in a HPV-positive gynecologic outpatient population. Mod Pathol 2016;29:870–878.
23 Bosgraaf RP, Verhoef VM, Massuger LF, *et al.* Comparative performance of novel self-sampling methods in detecting high-risk human papillomavirus in 30,130 women not attending cervical screening. Int J Cancer 2015;136:646–655.
24 Bulk S, Van Kemenade FJ, Rozendaal L, *et al.* The Dutch CISOE-A framework for cytology reporting increases efficacy of screening upon standardisation since 1996. J Clin Pathol 2004;57:388–393.
25 Casparie M, Tiebosch AT, Burger G, *et al.* Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol 2007;29:19–24.
26 Heideman DA, Hesselink AT, Berkhof J, *et al.* Clinical validation of the cobas 4800 HPV test for cervical screening purposes. J Clin Microbiol 2011;49:3983–3985.
27 R Core Team. R: a language and environment for statistical computing. Available at: http://CRAN.R-project.org/package=DTComPair. Accessed on 24 August 2016.
28 Gustinucci D, Giorgi Rossi P, Cesarini E, *et al.* Use of cytology, E6/E7 mRNA, and p16INK4a-Ki-67 to define the management of human papillomavirus (HPV)-positive women in cervical cancer screening. Am J Clin Pathol 2016;145:35–45.
29 Wentzensen N, Fetterman B, Castle PE, *et al.* p16/Ki-67 dual stain cytology for detection of cervical precancer in HPV-positive women. J Natl Cancer Inst 2015;107: djv257.
30 Allia E, Ronco G, Coccia A, *et al.* Interpretation of p16 (INK4a) /Ki-67 dual immunostaining for the triage of human papillomavirus-positive women by experts and nonexperts in cervical cytology. Cancer Cytopathol 2015;123:212–218.
31 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.
32 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33: 159–174.
33 Uijterwaal MH, Polman NJ, Witte BI, *et al.* Triaging HPV-positive women with normal cytology by p16/Ki-67 dual-stained cytology testing: baseline and longitudinal data. Int J Cancer 2015;136:2361–2368.
34 Wentzensen N, Fetterman B, Tokugawa D, *et al.* Interobserver reproducibility and accuracy of p16/Ki-67 dual-stain cytology in cervical cancer screening. Cancer Cytopathol 2014;122:914–920.
35 Benevolo M, Allia E, Gustinucci D, *et al.* Interobserver reproducibility of cytologic p16INK4a /Ki-67 dual immunostaining in human papillomavirus-positive women. Cancer 2017.
36 Brink AA, Meijer CJ, Wiegerinck MA, *et al.* High concordance of results of testing for human papillomavirus in cervicovaginal samples collected by two methods, with comparison of a novel self-sampling device to a conventional endocervical brush. J Clin Microbiol 2006;44:2518–2523.
37 Steenbergen RD, Snijders PJ, Heideman DA, *et al.* Clinical implications of (epi)genetic changes in HPV-induced cervical precancerous lesions. Nat Rev Cancer 2014;14:395–405.
38 Verhoef VM, Bosgraaf RP, van Kemenade FJ, *et al.* Triage by methylation-marker testing versus cytology

in women who test HPV-positive on self-collected cervicovaginal specimens (PROHTECT-3): a randomised controlled non-inferiority trial. Lancet Oncol 2014;15:315–322.
39 Mirabello L, Sun C, Ghosh A, *et al.* Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. J Natl Cancer Inst 2012;104:556–565.
40 Luttmer R, De Strooper LM, Dijkstra MG, *et al.* FAM19A4 methylation analysis in self-samples compared with cervical scrapes for detecting cervical (pre)cancer in HPV-positive women. Br J Cancer 2016;115:579–587.
41 Luttmer R, De Strooper LM, Steenbergen RD, *et al.* Management of high-risk HPV-positive women for detection of cervical (pre)cancer. Expert Rev Mol Diagn 2016;16:961–974.
42 Moriarty AT, Nayar R, Arnold T, *et al.* The Tahoe Study: bias in the interpretation of Papanicolaou test results when human papillomavirus status is known. Arch Pathol Lab Med 2014;138:1182–1185.