



Cervical Cancer Screening: Comparison of Conventional Pap Smear Test, Liquid-Based Cytology, and Human Papillomavirus Testing as Stand-alone or Cotesting Strategies

Linda A. Liang¹, Thomas Einzmann², Arno Franzen³, Katja Schwarzer⁴, Gunther Schaubberger¹, Dirk Schriefer^{1,5}, Kathrin Radde¹, Sylke R. Zeissig⁶, Hans Ikenberg⁷, Chris J.L.M. Meijer⁸, Charles J. Kirkpatrick⁹, Heinz Kölbl¹⁰, Maria Blettner¹¹, and Stefanie J. Klug¹

ABSTRACT

Background: Some countries have implemented stand-alone human papillomavirus (HPV) testing while others consider cotesting for cervical cancer screening. We compared both strategies within a population-based study.

Methods: The MARZY cohort study was conducted in Germany. Randomly selected women from population registries aged ≥ 30 years ($n = 5,275$) were invited to screening with Pap smear, liquid-based cytology (LBC, ThinPrep), and HPV testing (Hybrid Capture2, HC2). Screen-positive participants [ASC-US+ or high-risk HC2 (hrHC2)] and a random 5% sample of screen-negatives were referred to colposcopy. *Post hoc* HPV genotyping was conducted by GP5+/6+ PCR-EIA with reverse line blotting. Sensitivity, specificity (adjusted for verification bias), and potential harms, including number of colposcopies needed to detect 1 precancerous lesion (NNC), were calculated.

Results: In 2,627 screened women, cytological sensitivities (Pap, LBC: 47%) were lower than HC2 (95%) and PCR (79%)

for CIN2+. Cotesting demonstrated higher sensitivities (HC2 cotesting: 99%; PCR cotesting: 84%), but at the cost of lower specificities (92%–95%) compared with HPV stand-alone (HC2: 95%; PCR: 94%) and cytology (97% or 99%). Cotesting versus HPV stand-alone showed equivalent relative sensitivity [HC2: 1.06, 95% confidence interval (CI), 1.00–1.21; PCR: 1.07, 95% CI, 1.00–1.27]. Relative specificity of Pap cotesting with either HPV test was inferior to stand-alone HPV. LBC cotesting demonstrated equivalent specificity (both tests: 0.99, 95% CI, 0.99–1.00). NNC was highest for Pap cotesting.

Conclusions: Cotesting offers no benefit in detection over stand-alone HPV testing, resulting in more false positive results and colposcopy referrals.

Impact: HPV stand-alone screening offers a better balance of benefits and harms than cotesting.

See related commentary by Wentzensen and Clarke, p. 432

Introduction

With the implementation of cytologic Papanicolaou (Pap) smear as a detection method for cervical cell abnormality since the 1960s, overall

cervical cancer incidence and mortality rates in high income countries have fallen drastically (1). Lately however, incidence rates have remained stagnant in many of these settings (1, 2). Despite its successes, screening with cytology is resource-intensive and prone to poor reproducibility with a widely ranging sensitivity of 43% to 96%, even in high-resource countries such as Germany (3). In addition, since the discovery of the causative agent human papillomavirus (HPV) in almost all cervical cancers, prophylactic vaccines that target HPV types attributable in up to 90% of cervical cancers have been developed (4). Consequently, as HPV-vaccinated cohorts move toward screening eligibility, accuracy of cytology will be even further compromised because of the significant reduction in precancerous and cancerous lesions (5). Therefore, more objective detection methods are needed.

Molecular testing for HPV DNA has recently appeared as an alternative screening method, offering greater reproducibility and high-throughput benefits. These advantages led to U.S. FDA approval of HPV testing as an adjunct to cytology (reflex testing) or as a concomitant test (cotest). Since then, pooled studies and meta-analyses of several randomized controlled trials and observational studies have demonstrated superior detection of HPV-based screening (both stand-alone and cotesting) in comparison with cytology (3, 6, 7). These findings coupled with results from the ATHENA trial prompted regulatory approval of HPV testing as a stand-alone screening strategy in 2014 (8). As a result, stand-alone HPV testing has become the preferred strategy over cytology in European and U.S. guidelines among others (9, 10). In the Netherlands, cytology has already been replaced by stand-alone HPV testing at 5-year intervals (11).

¹Epidemiology, Department of Sports and Health Sciences, Technical University of Munich, Munich, Germany. ²Department of Obstetrics and Gynecology, University Hospital Cologne, Cologne, Germany. ³Department of Obstetrics and Gynecology, Kemperhof, Gemeinschaftsklinikum Mittelrhein, Koblenz, Germany. ⁴VivaQ MVZ, Mainz, Germany. ⁵Center of Clinical Neuroscience, University Hospital Carl Gustav Carus Dresden, Dresden, Germany. ⁶Cancer Registry of Rhineland-Palatinate, Mainz, Germany. ⁷Cytomol MVZ, Frankfurt, Germany. ⁸Amsterdam UMC, Vrije Universiteit Amsterdam, Pathology, Cancer Center Amsterdam, Amsterdam, the Netherlands. ⁹Institute of Pathology, University Medical Center, Johannes Gutenberg University, Mainz, Germany. ¹⁰Department of Obstetrics and Gynecology, Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria. ¹¹Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center, Johannes Gutenberg University, Mainz, Germany.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Stefanie J. Klug, Technical University of Munich, Munich 80992, Germany. Phone: 4989-2892-4951; Fax: 4989-2892-4953; E-mail: sekretariat.klug@tum.de

Cancer Epidemiol Biomarkers Prev 2021;30:474–84

doi: 10.1158/1055-9965.EPI-20-1003

©2020 American Association for Cancer Research.

There are still, however, several concerns of stand-alone HPV screening regarding lowered specificity, safety of extended screening intervals, testing in women under 30 years of age, and observations of HPV test-negative carcinomas (12, 13). These concerns have been frequently used to advocate cotesting over stand-alone HPV screening and have even bolstered cotesting as a screening modality alongside HPV testing and triennial cytology in the United States (10). While a large robust body of evidence supports HPV-based screening, there is an ongoing debate around cotesting versus stand-alone HPV testing (14). Few studies have compared accuracy of the two strategies (6, 15, 16), with some observing minor differences in detection, albeit based on retrospective analyses (17, 18). Moreover, separate comparisons between HPV testing and Pap or liquid-based cytology (LBC) are lacking, and few have compared Pap to LBC-based cotesting (19, 20). To our knowledge, no study has directly compared both Pap and LBC as cotesting strategies to stand-alone HPV testing with two standard HPV comparators. Findings from such analyses provide necessary evidence on optimal screening strategies, especially for countries considering HPV-based screening, such as Germany, which has implemented an organized screening program with cotesting only in 2020 (21). Therefore, in a large population-based sample of women within an opportunistic screening setting, we compared absolute and relative clinical test accuracy of stand-alone and cotesting strategies with conventional Pap, LBC, and two HPV tests.

Materials and Methods

Data stem from MARZY, a randomized prospective cohort study with a population-based sample of women eligible for cervical cancer screening in Germany between 2005 and 2012. Details on recruitment and intervention have been published in detail elsewhere (22). Briefly described, a random sample of 9,383 women selected from population registries were randomized into two intervention arms (sole invitation to screening, invitation with information brochure) and a non-invitation control arm to observe differences in screening attendance. At baseline, women randomized to both intervention arms ($n = 5,275$; eligible = 3,759) were invited to undergo screening with a conventional Pap smear, a LBC study swab, and HPV testing (Fig. 1). These analyses focus on baseline-screened participants between 2005 and 2007 ($n = 2,627$).

Participants

Inclusion criteria were women 30 years or older and residing within the urban and rural region of Mainz and Mainz-Bingen, Germany. Women with any previous cervical cancer diagnoses, hysterectomy, or pregnancy at baseline were excluded. To preserve real-world screening, all gynecological practices and general practitioners conducting routine cervical cancer screening within the study region or who were elected by participants outside the study region were contacted to cooperate ($n = 121$) and closely monitored for quality assurance (23). Participants provided written informed consent before undergoing screening.

Cytology

In line with the standard practice, gynecologists first obtained a conventional Pap smear and sent the specimen fixed onto a glass slide to their routine laboratory for assessment. Diagnostic results were relayed back to the study team. A second cytologic study swab was obtained using an Ayres spatula and endocervical broom or cytobrush when the transformation zone was not visible. The cells of this specimen were directly suspended in a vial containing 20 mL of PreservCyt Liquid Solution (ThinPrep, Cytec/Hologic) and sent to

a centralized laboratory (CytoMol, Frankfurt, Germany) routinely conducting LBC assessment.

Cytologic findings at baseline were based on the Munich II Nomenclature, which was used prior to Munich III, the current classification system in Germany (24). As up to 10% of moderate cervical intraepithelial neoplasia (CIN2) and 4% of severely dysplastic CIN3 are detected in equivocal cytology (25), all women with atypical squamous cells of undetermined significance or worse (ASC-US+) were referred to colposcopy. In German nomenclature, Pap IIw is an unofficial category widely used to denote equivocal results and is considered equivalent to ASC-US (24) from the International Bethesda Classification for Cytology (2014) (26). Pap IIID is equivalent to low-grade intraepithelial lesions, LSIL, and was also assessed for comparison.

HPV DNA testing

Remaining PreservCyt solution was directly used for HPV DNA detection by Hybrid Capture2 (HC2, Qiagen), detecting 13 high-risk HPV types (hrHPV: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68). Detection of hrHPV was set at the manufacturer recommended cut-off ratio of 1.0 relative light units (RLU). In addition to HC2, we analyzed the accuracy of another standard HPV comparator. All available PreservCyt solution samples were processed *post hoc* using GP5+/6+ PCR with enzyme immunoassay (EIA) probes targeting 14 hrHPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) and low-risk types [6, 11, 26, 30, 32, 34, 40, 42, 43, 44, 53, 54, 55, 57, 61, 64, 67, 69, 70, 71, 72, 73, 81, 82 (variants mm4 and is39), 83, 84, 85, 86, 89 (formerly cp6108), 90 (formerly jc9710)]. GP5+/6+ PCR-EIA-positive samples were typed and classified by reverse line dot blot hybridization performed at the Department of Pathology, Amsterdam UMC location VU Medical Center, the Netherlands. PCR results were not used to refer women to colposcopy as these were processed *post hoc*. HrHPV types were based on IARC 2012 classifications of probably carcinogenic and cervical carcinogens thus HPV 66 was not analyzed as high-risk (27).

Colposcopy and histology

Women were considered screen-positive if either cytology was ASC-US+ or HC2 was positive (hrHC2) and subsequently referred to colposcopy, conducted centrally by certified study colposcopists (Department of Obstetrics and Gynecology, Mainz University Hospital, Mainz, Germany). Screen-positive women who did not arrange a colposcopy appointment within 2 months were contacted and encouraged to attend. If unable or unwilling, participants were further interviewed on reasons for non-attendance. Screen-negative was defined as both cytology (negative for intraepithelial lesion or malignancy, NILM) and HC2 negative. PCR results were not considered for colposcopy referral, as this test was only conducted *post hoc*. A random sample (5%) of all screen-negative women was also invited to colposcopy (Fig. 1).

Colposcopic examinations were conducted in accordance to 2002 International Federation of Cervical Pathology and Colposcopy (IFCPC) guidelines (28) with 5% acetic acid application first followed by Lugol's iodine solution. Participants with macroscopically visible lesions (abnormal colposcopic findings, colposcopic features suggestive of invasive cancer) underwent punch biopsy with multiple samples obtained for multiple acetowhitened lesions. Endocervical curettage was conducted if the transformation zone was obscured. Colposcopists were additionally instructed to take two biopsies from participants without visible lesions at the 12 and 6 o'clock regions of the cervix. All biopsies were assessed centrally by an experienced histopathologist. To maintain quality assurance, all histopathologic samples were

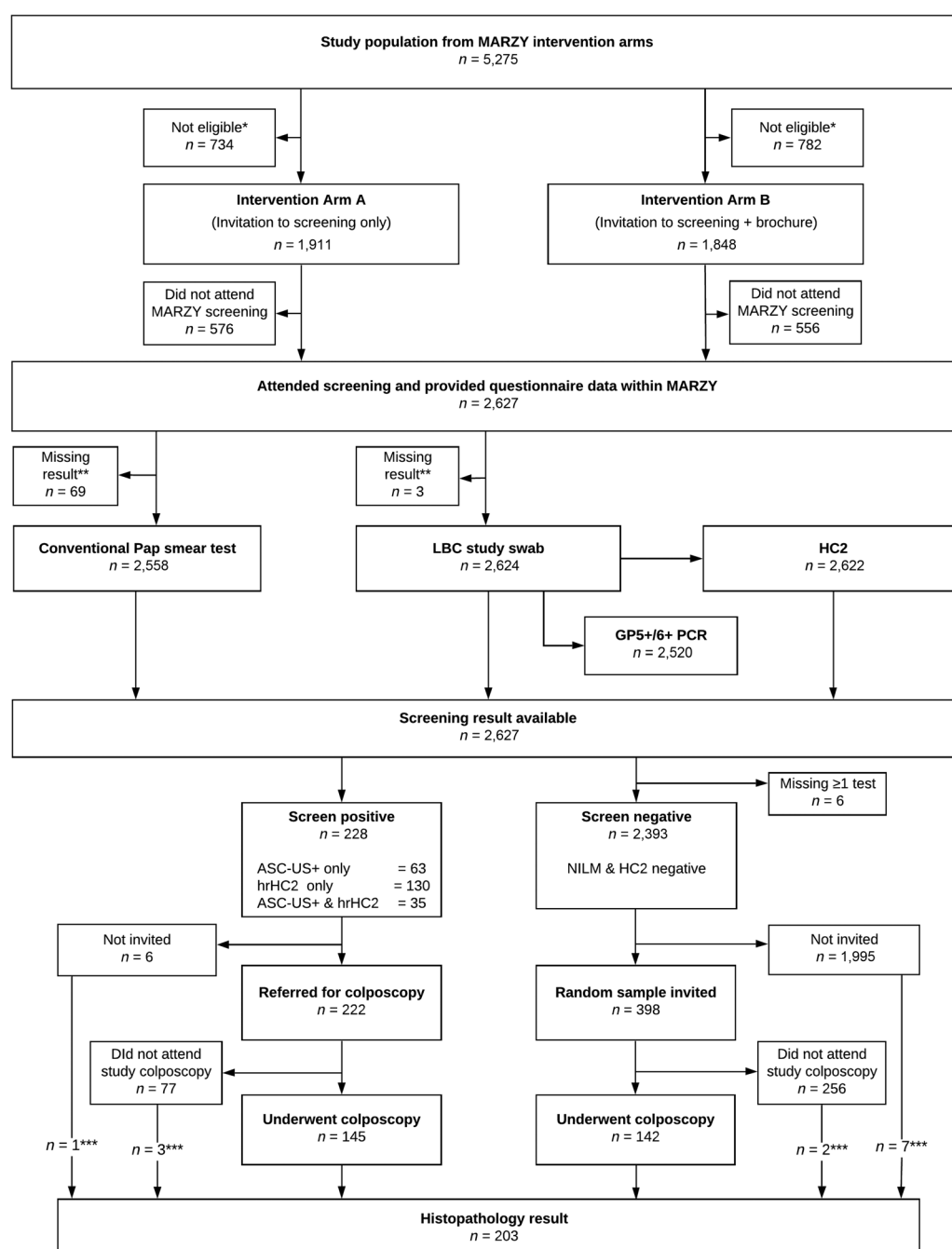


Figure 1.

Flow chart of study design and end results. LBC, liquid-based cytology; HC2, Hybrid Capture2 HPV test; hrHC2, high-risk HC2 type; ASC-US+, atypical squamous cells of undetermined significance or worse; NILM, negative for intraepithelial lesion or malignancy; * excluded due to hysterectomy, pregnancy, or history of cervical cancer; ** no sample obtained; *** external histopathology results reported within 1 year of study swab.

independently reviewed by a second histopathologist. A third histopathologist was called upon to settle discrepancies. The final agreed upon result was used for evaluation. In addition, external information regarding colposcopy and histopathology conducted outside the study during the study period were traced. Results reported within 1 year of the study swab were included. This active tracing of information was necessary due to the lack of centralized data registration of precancerous

lesions in Germany and an opportunistic screening system. Women with suspected lesions at colposcopy or histopathologic lesions were managed as per local protocols for standard care.

Statistical analyses

The *a priori* sample size estimation for MARZY was based on the primary outcome assuming 5% increase in participation rate

between randomized arms described elsewhere (22). The endpoints of interest for screening purposes were CIN2 or worse (CIN2+) and CIN3 or worse (CIN3+). Absolute sensitivity, specificity, and positive predictive values (PPV) were calculated. We calculated the complement of the negative predictive value (cNPV) to show the risk of CIN2+ or CIN3+ among screen-negative women (1-NPV). Although a random 5% sample of screen-negatives were invited to colposcopy, partial verification bias could still lead to an overestimation of sensitivity and an underestimation of specificity. Therefore, we adjusted all test accuracy estimates based on the probability to be followed-up for verification via the following sampling fractions (formula previously described in ref. 29: negative (0.04), cytology positive only (ASC-US+; 0.44), and hrHC2 (0.46). The inverse of these probabilities was applied as a weight to participants in their assigned strata (negative: 24.39, ASC-US+: 2.27, hrHC2: 2.17). As PCR test results did not influence test status nor strata allocation (processed *post hoc*), verification adjustment is equally appropriate for *post hoc* test results. Confidence intervals (CI) were obtained using bootstrap resampling methods ($n = 1,000$) at the lower 2.5% and upper 97.5% quantiles (30). To avoid problems with proportion calculations, we added 0.5 to each 2×2 contingency cell for HC2 cotesting and all PCR-based strategies (Haldane correction; ref. 31).

Comparisons of stand-alone sensitivity and specificity were conducted using McNemar's paired sample test, stratified by verified disease status. Positive and negative likelihood ratios (PLR, NLR) were calculated to compare cotesting strategies with stand-alone components. Relative sensitivity and specificity were calculated to directly compare all strategies, defined as the ratios of sensitivity and specificity between tests (no Haldane correction). CIs for crude ratios were based on Wald for paired data and adjusted ratios were based on bootstrap resampling.

For potential harms, we calculated false positive and negative rates (FPR: $1 - \text{specificity}$, FNR: $1 - \text{sensitivity}$) and the number of women needed to undergo colposcopy to detect one CIN2+ or CIN3+ case (NNC: $1/\text{PPV}$) per test strategy. In sensitivity analyses, we calculated accuracy for women aged ≥ 35 years and of within-study collected biopsies (i.e., excluding external findings). HC2 test accuracy at higher viral load cutoffs at 2.0, 3.0, and 10.0 RLU were also conducted to determine specificity. All analyses were conducted using SAS 9.4 (SAS Institute). We complied with the STARD guidelines for reporting and followed Good Epidemiological Practice guidelines. The MARZY study was approved by the ethical committee of the state of Rhineland-Palatinate and the state government data protection office.

Ethics approval and consent to participate

All participants provided signed informed consent to the study. The MARZY study was approved by the ethical committee of the state of Rhineland-Palatinate [Landesärztekammer Rheinland-Pfalz: 837.438.03 (4100)] and the state government data protection office. All recruitment, data collection, and analyses were performed in accordance to Good Epidemiological Practice guidelines and the Declaration of Helsinki.

Consent for publication

Not applicable.

Data availability statement

Anonymized data that support the findings of this study may be available from the corresponding author upon reasonable request.

Results

Of the 5,275 women invited for screening within MARZY arms A and B, 2,627 (49.8%) were screened (Fig. 1). Mean age was 47.09 years (SD = 9.97; range 30–68 years). In women aged 30–39 years, 27% attended screening while only 15% of ≥ 60 -year-old women attended. Approximately 9% of all participants either reported to have never undergone screening or did not attend screening at the recommended interval nor within a 5-year period (Supplementary Table S1 shows characteristics).

Pap and LBC detected 69 (2.7%) and 47 (1.8%) equivocal or worse cytology (ASC-US+), respectively, while HC2 and PCR detected 165 (6.3%) and 165 (6.6%) hrHPV, respectively. Among the 2,627 screened (Fig. 1), 228 (8.7%) were screen-positive where 63 (2.4%) were ASC-US+ only, 130 (5.0%) were hrHC2 only, and 35 (1.3%) were both ASC-US+ and hrHC2. Six women were not referred to colposcopy for reasons including planned hysterectomy elsewhere. Of 222 remaining screen-positives, despite active callback, 145 (65.3%) underwent colposcopy at the study center. Of all 2,393 screen-negatives, 142 (5.9%) attended study colposcopy (attendance rate, $142/398 = 35.7\%$). Colposcopies were conducted on average 4.9 months after screening (SD = 4.9), 6.0 months among screen-positives (SD = 6.4) and 3.7 months among screen-negatives (SD = 1.9).

Of the 203 histopathologic results (190 from study colposcopy; range of biopsies taken 1–5; 13 from externally conducted colposcopies), 3 squamous cell carcinomas (SCC; 1.5%), 7 high-grade (CIN3; 3.5%), 9 moderate-grade (CIN2; 4.4%), and 7 mild lesions (CIN1; 3.5%) were reported (Fig. 1). No adenocarcinomas or glandular lesions were detected (Supplementary Table S2). All CIN2+ were HPV positive (Supplementary Table S3).

Absolute test accuracy

Estimates adjusted for verification bias for ASC-US+ are presented in Table 1 (crude estimates: Supplementary Tables S4 and S5) and are based on 41 CIN2+ and 22 CIN3+ hypothetical lesions after adjustment. HC2 presented the highest sensitivities (cotesting 98.82%, stand-alone 94.56%) with HC2 stand-alone significantly more sensitive than either cytology (Pap and LBC both 47.47%; $P < 0.0001$). Specificity of HC2 stand-alone (95.12%) was significantly lower than cytology (Pap 97.48%; LBC 98.64%; $P < 0.0001$). Contrasting to stand-alone, cotesting specificity was reduced (Pap/HC2 93.09%; LBC/HC2 94.58%). For CIN3+, sensitivity of both Pap and LBC stand-alone was 70.11% and 89.67% for HC2 stand-alone. Specificities were similar to CIN2+.

With PCR, high sensitivities were also observed for CIN2+ (both cotests 84.24%, stand-alone 78.99%) and stand-alone was significantly higher than cytology ($P < 0.01$). PCR cotesting conferred the lowest specificities (Pap/PCR 92.21%; LBC/PCR 93.73%) increasing to 94.25% stand-alone, but significantly lower than cytology ($P < 0.0001$). For CIN3+, PCR presented the highest sensitivity (97.81%) but specificities were lower than cytology.

PPVs also indicated higher probability of disease by cytology, particularly with LBC, than HPV-based screening. However, for CIN2+ lesions, HC2-based strategies revealed similar PPVs to Pap. cNPVs revealed greater safety against CIN2+ among screen-negatives with HPV-based strategies, particularly HC2 cotesting ($< 0.1\%$). Safety against CIN2+ was lowest with cytology only ($\sim 0.86\%$).

For LSIL+, sensitivities of cytology were lower (Table 2). LBC and HC2 cotesting conferred lower sensitivity than Pap and HC2 cotesting, but the former showed identical sensitivity as HC2 stand-alone. LBC and HC2 cotesting performed similarly to Pap and HC2 cotesting in terms of specificity and PPV. Sensitivity for stand-alone PCR for

Table 1. Sensitivity, specificity, NPV, PPV, and likelihood ratios adjusted for verification bias for ASC-US+ equivalent.

Cut-off	Endpoint	Test	Sensitivity % (95% CI) ^a	Specificity % (95% CI) ^a	PPV % (95% CI) ^a	cNPV % (95% CI) ^a	PLR (95% CI)	NLR (95% CI)	
ASC-US+	CIN2+	Pap	47.47 (25.00-68.26) ^{b,c}	97.48 (96.44-98.41) ^{b,c}	23.25 (10.16-37.23)	0.86 (0.37-1.38)	18.86 (12.63-28.16)	0.54 (0.40-0.72)	
		LBC	47.47 (25.00-70.11) ^{b,c}	98.64 (97.89-99.26) ^{b,c,d}	35.78 (16.62-54.93)	0.85 (0.36-1.38)	34.79 (21.99-55.04)	0.53 (0.40-0.71)	
	HC2	Pap	94.56 (82.09-100.0)	95.12 (93.55-96.42)	23.68 (14.57-33.14)	0.09 (0.0-0.32)	19.38 (16.10-23.33)	0.06 (0.02-0.20)	
		LBC	78.99 (58.40-94.74)	94.25 (90.08-97.52)	18.19 (8.33-38.26)	0.36 (0.08-0.75)	13.73 (10.99-17.15)	0.22 (0.12-0.40)	
	Pap/HC2	Pap	98.82 (96.54-100.0)	93.09 (91.09-95.07)	19.01 (11.73-26.68)	0.02 (0.02-0.02)	14.31 (12.37-16.55)	0.01 (0.00-0.20)	
		LBC	98.82 (96.54-100.0)	94.58 (92.94-95.97)	23.01 (13.97-31.82)	0.02 (0.02-0.02)	18.23 (15.47-21.49)	0.01 (0.00-0.20)	
	Pap/PCR	Pap	84.24 (66.67-100.0)	92.21 (88.09-95.95)	14.92 (7.52-26.46)	0.28 (0.0-0.63)	10.81 (8.96-13.05)	0.17 (0.08-0.35)	
		LBC	84.24 (65.10-100.0)	93.73 (89.49-97.03)	17.85 (8.68-34.63)	0.27 (0.0-0.62)	13.43 (11.00-16.40)	0.17 (0.08-0.34)	
	ASC-US+	CIN3+	Pap	70.11 (36.47-100.0)	97.34 (96.24-98.27) ^{b,c}	18.10 (6.78-30.54)	0.26 (0.0-0.60)	26.31 (18.36-37.69)	0.31 (0.16-0.58)
			LBC	70.11 (38.12-100.0)	98.48 (97.64-99.11) ^{b,c,d}	27.86 (11.29-45.73)	0.25 (0.0-0.57)	46.09 (30.49-69.68)	0.30 (0.16-0.58)
HC2		Pap	89.67 (65.87-100.0)	94.41 (92.66-95.76)	11.84 (4.66-19.62)	0.09 (0.0-0.32)	16.03 (12.96-19.83)	0.11 (0.03-0.38)	
		LBC	97.81 (95.71-100.0)	93.85 (89.70-97.15)	12.35 (4.81-26.80)	0.02 (0.02-0.02)	15.91 (13.52-18.72)	0.02 (0.0-0.36)	
Pap/HC2		Pap	97.81 (93.69-100.0)	92.39 (90.11-94.15)	10.13 (4.18-15.95)	0.02 (0.02-0.02)	12.86 (11.09-14.90)	0.02 (0.0-0.37)	
		LBC	97.81 (93.69-100.0)	93.87 (92.13-95.32)	12.26 (5.02-19.38)	0.02 (0.02-0.02)	15.95 (13.56-18.77)	0.02 (0.0-0.36)	
Pap/PCR		Pap	97.81 (93.69-100.0)	91.75 (87.65-95.26)	9.51 (3.91-17.76)	0.02 (0.02-0.02)	11.85 (10.27-13.67)	0.02 (0.0-0.37)	
		LBC	97.81 (93.69-100.0)	93.25 (89.08-96.62)	11.37 (4.64-23.26)	0.02 (0.02-0.02)	14.49 (12.40-16.94)	0.02 (0.0-0.36)	

Note: cNPV = 1-NPV; complement of the NPV.

ASC-US+ = Atypical squamous cells of undetermined significance or worse.

CIN2+ = Moderate cervical intraepithelial neoplasia or worse.

CIN3+ = Severe cervical intraepithelial neoplasia (incl. carcinoma in situ) or worse.

Abbreviations: HC2, Hybrid Capture2 HPV test; LBC, liquid-based cytology; NLR, negative likelihood ratio; Pap, conventional Pap smear; PCR, GP5+/6+ HPV PCR test; PLR, positive likelihood ratio; PPV, positive predictive value.

^a95% CI based on bootstrap resampling ($n = 1,000$ resamples).

^bSignificant McNemar's test comparing Pap or LBC to HC2, $P < 0.05$.

^cSignificant McNemar's test comparing Pap or LBC to PCR, $P < 0.05$.

^dSignificant McNemar's test comparing LBC to Pap, $P < 0.05$.

Table 2. Sensitivity, specificity, NPV, PPV, and likelihood ratios adjusted for verification bias for LSIL+ equivalent.

Cut-off	Endpoint	Test	Sensitivity % (95% CI) ^a	Specificity % (95% CI) ^a	PPV % (95% CI) ^a	cNPV % (95% CI) ^a	PLR (95% CI)	NLR (95% CI)	
LSIL+	CIN2+	Pap	42.22 (18.93-64.42) ^{b,c}	99.40 (98.88-99.83) ^{b,c}	53.06 (26.53-79.71)	0.92 (0.46-1.49)	70.41 (38.18-129.85)	0.58 (0.45-0.75)	
		LBC	36.77 (16.62-58.16) ^{b,c}	99.07 (98.43-99.59) ^{b,c}	38.73 (16.57-61.37)	1.01 (0.49-1.60)	39.48 (22.46-69.39)	0.64 (0.51-0.81)	
	HC2	Pap	94.56 (82.09-100.0)	95.12 (93.55-96.42)	23.68 (14.57-33.14)	0.09 (0.0-0.32)	19.38 (16.10-23.33)	0.06 (0.02-0.20)	
		LBC	78.99 (58.40-94.74)	94.25 (90.08-97.52)	18.19 (8.33-38.26)	0.36 (0.08-0.75)	13.73 (10.99-17.15)	0.22 (0.12-0.40)	
	Pap/HC2	Pap	98.82 (96.54-100.0)	94.84 (93.11-96.29)	23.90 (14.46-33.12)	0.02 (0.02-0.02)	19.14 (16.17-22.66)	0.01 (0.00-0.20)	
		LBC	94.56 (82.09-100.0)	94.95 (93.37-96.27)	23.06 (14.25-32.39)	0.09 (0.0-0.32)	18.71 (15.59-22.46)	0.06 (0.02-0.20)	
	Pap/PCR	Pap	84.24 (64.92-100.0)	93.89 (89.68-97.15)	18.27 (9.35-35.21)	0.27 (0.0-0.62)	13.78 (11.27-16.86)	0.17 (0.08-0.34)	
		LBC	84.24 (65.97-100.0)	94.07 (89.94-97.39)	18.70 (9.48-35.80)	0.27 (0.0-0.62)	14.21 (11.60-17.41)	0.17 (0.08-0.34)	
	LSIL+	CIN3+	Pap	60.14 (25.68-90.98) ^c	99.24 (98.67-99.70) ^{b,c}	39.86 (15.19-66.48)	0.34 (0.08-0.71)	78.88 (45.22-137.60)	0.40 (0.24-0.67)
			LBC	59.78 (27.63-88.89) ^{b,c}	98.99 (98.34-99.50) ^{b,c}	33.20 (12.39-53.20)	0.34 (0.08-0.69)	59.31 (35.49-99.11)	0.41 (0.24-0.68)
HC2		Pap	89.67 (65.87-100.0)	94.41 (92.66-95.76)	11.84 (4.66-19.62)	0.09 (0.0-0.32)	16.03 (12.96-19.83)	0.11 (0.03-0.38)	
		LBC	97.81 (95.71-100.0)	93.85 (89.70-97.15)	12.35 (4.81-26.80)	0.02 (0.02-0.02)	15.91 (13.52-18.72)	0.02 (0.0-0.36)	
Pap/HC2		Pap	97.81 (93.69-100.0)	94.12 (92.23-95.54)	12.74 (5.28-19.87)	0.02 (0.02-0.02)	16.65 (14.10-19.65)	0.02 (0.0-0.36)	
		LBC	89.67 (65.87-100.0)	94.23 (92.47-95.61)	11.53 (4.60-18.93)	0.09 (0.0-0.32)	15.55 (12.59-19.20)	0.11 (0.03-0.38)	
Pap/PCR		Pap	97.81 (93.69-100.0)	93.41 (89.35-96.72)	11.64 (4.60-23.77)	0.02 (0.02-0.02)	14.84 (12.67-17.38)	0.02 (0.0-0.36)	
		LBC	97.81 (93.69-100.0)	93.59 (89.52-97.01)	11.91 (5.14-24.76)	0.02 (0.02-0.02)	15.26 (13.01-17.91)	0.02 (0.0-0.36)	

Note: cNPV = 1-NPV, complement of the NPV.
 LSIL+ = Low-grade squamous intraepithelial lesion or worse.
 CIN2+ = Moderate cervical intraepithelial neoplasia or worse.
 CIN3+ = Severe cervical intraepithelial neoplasia (incl. carcinoma *in situ*) or worse.
 Abbreviations: HC2, Hybrid Capture2 HPV test; LBC, liquid-based cytology; NLR, negative likelihood ratio; Pap, conventional Pap smear; PCR, GP5+/6+ HPV PCR test; PLR, positive likelihood ratio; PPV, positive predictive value.
^a95% CI based on bootstrap resampling (*n* = 1,000 resamples).
^bSignificant McNemar's test comparing Pap or LBC to HC2, *P* < 0.05.
^cSignificant McNemar's test comparing Pap or LBC to PCR, *P* < 0.05.
^dSignificant McNemar's test comparing LBC to Pap, *P* < 0.05.

CIN2+ was lower than PCR cotesting sensitivities, but for CIN3+ no differences were observed.

Relative test accuracy

In Fig. 2, the relative sensitivity and specificity for CIN2+ conferred similar estimates for crude and verification bias-adjusted calculations, but specificities appear under or overestimated (from unity) when potential verification bias is not accounted for. When compared with either cytology (Fig. 2A), HC2 stand-alone [1.99, 95% CI, 1.30–4.00] and both respective cotesting strategies detected twice as many CIN2+ lesions (Pap/HC2 2.11, 95% CI, 1.43–4.04; LBC/HC2 2.11, 95% CI, 1.39–4.01). Cotesting did not detect more CIN2+ compared with HC2 stand-alone (Pap and LBC 1.06, 95% CI, 1.00–1.21). Similar results were also observed among PCR strategies (Fig. 2C), however sensitivity estimates were reduced (PCR stand-alone 1.66; PCR cotesting 1.77).

Specificity of HC2 stand-alone (Fig. 2B) was significantly lower than cytology (Pap 0.98, 95% CI, 0.97–0.98; LBC 0.96, 95% CI, 0.96–0.97) and similar findings were observed for PCR stand-alone versus cytology (Fig. 2D). Pap cotesting was significantly less specific than HPV stand-alone while LBC cotesting presented no significant difference in detection compared with either HPV test stand-alone. For CIN3+, relative sensitivities were not statistically significant due to the low number of CIN3+ ($n = 10$). These relative specificities appeared similar to the CIN2+ cutoff (Supplementary Fig. S1).

Potential harms

For CIN2+, the highest FPRs were observed with HPV testing (Table 3), particularly cotesting strategies (6.27%–7.79%) with the exception of HC2 cotesting (5.42%). HC2 and PCR stand-alone demonstrated moderate FPRs (4.88%–5.75%), followed by

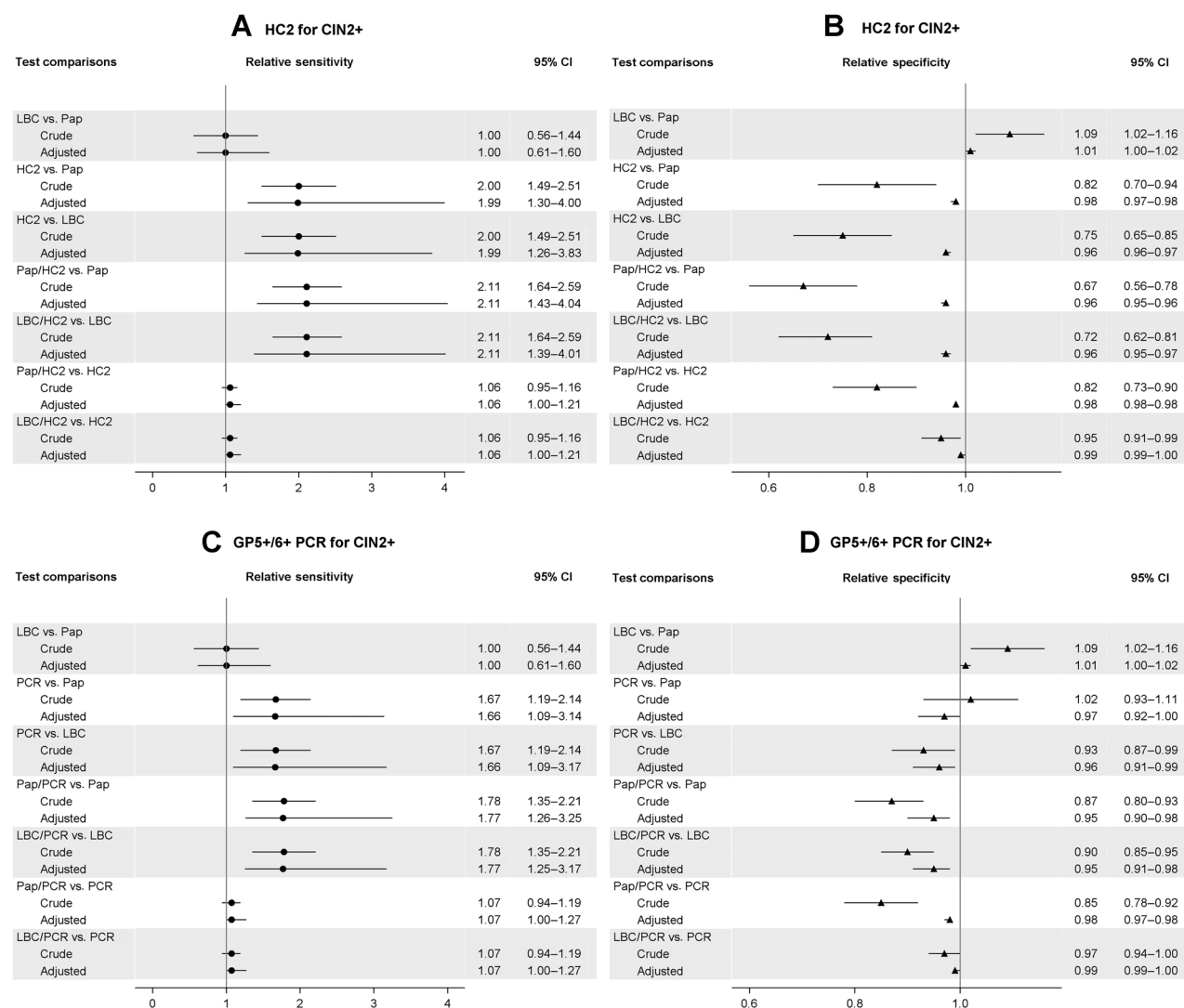


Figure 2. Relative sensitivity and specificity of tests comparing both crude and adjusted estimates for CIN2 or worse as ASC-US+. **A**, Relative sensitivity for HC2. **B**, Relative specificity for HC2. **C**, Relative sensitivity for PCR. **D**, Relative specificity for PCR. Crude CIs based on Wald for paired data and adjusted CIs based on bootstrap resampling ($n = 1,000$); CIN2+, moderate cervical intraepithelial neoplasia or worse; Pap, conventional Pap smear; LBC, liquid-based cytology; HC2, Hybrid Capture2 HPV test; PCR, GP5+/6+ HPV PCR test.

Table 3. False positives, false negatives, and number needed to colposcopy at all cytology and precancerous lesion cutoffs.

Endpoint	Test	ASC-US+		LSIL+		NNC 1/PPV (95% CI) ^a
		False positive rate % (95% CI) ^a	False negative rate % (95% CI) ^a	False positive rate % (95% CI) ^a	False negative rate % (95% CI) ^a	
CIN2+	Pap	2.52 (1.59–3.56)	52.53 (31.74–75.00)	0.60 (0.18–1.12)	57.78 (35.58–81.07)	4.30 (2.80–8.81)
	LBC	1.36 (0.74–2.11)	52.53 (29.89–75.00)	0.93 (0.41–1.57)	63.23 (41.84–83.38)	2.79 (1.82–5.63)
	HC2	4.88 (3.58–6.45)	5.44 (0.0–17.35)	4.88 (3.58–6.45)	5.44 (0.0–17.35)	4.22 (3.00–7.10)
	PCR	5.75 (2.43–9.70)	21.01 (5.56–44.22)	5.75 (2.43–9.70)	21.01 (4.77–42.60)	5.50 (2.78–10.81)
	Pap/HC2	6.91 (5.23–9.22)	1.18 (0.84–1.89)	5.16 (3.85–7.05)	1.18 (0.84–1.89)	5.30 (3.72–8.62)
	LBC/HC2	5.42 (4.03–7.06)	1.18 (0.84–1.89)	5.05 (3.73–6.63)	5.44 (0.0–17.35)	4.37 (3.07–7.10)
	Pap/PCR	7.79 (4.23–11.60)	15.76 (0.0–35.71)	6.11 (2.58–9.94)	15.76 (0.0–35.71)	6.70 (3.85–12.82)
	LBC/PCR	6.27 (2.99–10.25)	15.76 (0.0–35.71)	5.93 (2.65–9.80)	15.76 (0.0–35.71)	5.60 (2.97–11.04)
CIN3+	Pap	2.66 (1.73–3.76)	29.89 (0.0–63.04)	0.76 (0.30–1.33)	39.86 (12.44–83.33)	5.52 (3.42–14.85)
	LBC	1.52 (0.89–2.36)	29.89 (0.0–62.08)	1.01 (0.50–1.66)	40.22 (13.91–77.87)	3.59 (2.21–8.69)
	HC2	5.59 (4.24–7.34)	10.33 (0.0–35.15)	5.59 (4.24–7.34)	10.33 (0.0–35.15)	8.44 (5.20–20.00)
	PCR	6.15 (2.88–10.05)	2.19 (1.38–4.43)	6.15 (2.88–10.05)	2.19 (1.38–4.43)	8.24 (3.91–19.59)
	Pap/HC2	7.61 (5.85–9.89)	2.19 (1.38–4.43)	5.88 (4.46–7.77)	2.19 (1.38–4.43)	10.05 (6.26–21.42)
	LBC/HC2	6.13 (4.68–7.87)	2.19 (1.38–4.43)	5.77 (4.39–7.53)	10.33 (0.0–35.15)	8.30 (5.19–17.39)
	Pap/PCR	8.25 (4.72–12.28)	2.19 (1.38–4.43)	6.59 (3.05–10.49)	2.19 (1.38–4.43)	10.71 (5.76–24.67)
	LBC/PCR	6.75 (3.47–10.69)	2.19 (1.38–4.43)	6.41 (3.10–10.32)	2.19 (1.38–4.43)	8.95 (4.51–20.79)

Note: ASC-US+ = Atypical squamous cells of undetermined significance or worse.

LSIL+ = Low-grade squamous intraepithelial lesion or worse.

False positive rate = Proportion of index test positives among biopsy verified normal results (1-specificity).

False negative rate = Proportion of index test negatives among biopsy verified abnormal results i.e., CIN present (1-sensitivity).

NNC = Number of women needed to undergo colposcopy to detect 1 precancerous lesion with ASC-US+.

Abbreviations: HC2, Hybrid Capture2 HPV test; LBC, liquid-based cytology; Pap, conventional Pap smear; PCR, GP5+/6+ HPV PCR test.

^a95% CI based on bootstrap resampling ($n = 1,000$ resamples).

Pap (2.52%) and LBC (1.36%). For CIN3+ lesions a similar pattern was observed. Conversely, FNRs were lowest among HC2 strategies but for CIN3+, PCR-based strategies and HC2 cotesting were identical. The number of women needed to undergo colposcopy to detect one CIN2+ was highest under Pap and PCR cotesting (6.70) followed by other cotesting strategies and HPV stand-alone (HC2 4.22, PCR 5.50; **Table 3**). For CIN3+ a larger difference between Pap and LBC cotesting was observed, and had greater colposcopy referrals than HPV stand-alone and cytology.

Sensitivity analyses

For women ≥ 35 years, test accuracy increased for CIN2+ (Supplementary Table S6), namely sensitivity of cytology stand-alone (up to 56.35% for Pap and LBC with ASC-US+ and 50.11% for Pap, 43.65% for LBC with LSIL+). Accuracy based on the 190 within-study histopathology results yielded similar estimates (Supplementary Table S7). After increasing the RLU cutoff of HC2 testing to 2.0, 3.0, and 10.0, further gains in specificity and PPV were observed (Supplementary Table S8). However, sensitivity was further reduced. These patterns were similar for both HC2 cotesting strategies. At all RLU cutoffs, NPV remained very similar, decreasing slightly with increasing RLU. Screening women ≥ 30 and ≥ 35 years of age revealed similar adjusted FPRs (Supplementary Figs. S2 and S3). All HPV-based strategies incurred more false positives; however, this was more pronounced among cotesting strategies.

We observed 94 discordant HPV results with genotyping information. 82 (87.2%) were HC2 negative but high-risk PCR positive and the most common detected types were HPV 16 (53.7%), 56 (12.2%), 45 (9.8%), and 18 (7.3%). All 12 PCR high-risk negative but hrHC2 positive were low-risk HPV types.

Discussion

Few studies have compared stand-alone HPV test accuracy to cotesting strategies (6, 15–17) and to our knowledge none have directly compared the two most common cytology methods and standard HPV comparators using these strategies. On the basis of a large population-based sample of women above 30 years of age within an opportunistic screening setting and notably poor quality in cytology (3), our results demonstrated similar accuracy of stand-alone HPV testing and LBC cotesting. In particular, sensitivity of any cotesting strategy was equivalent to stand-alone HPV, and specificity of Pap cotesting was significantly lower than stand-alone HPV. Between cotesting strategies, LBC cotesting indicated some advantage over Pap cotesting where specificity was equivalent to HPV stand-alone. Furthermore, false positive test results and colposcopy referrals were highest with cotesting, particularly Pap cotesting. These results are relevant for countries that offer cotesting like Germany (32) and the United States (10), and for many other countries globally that are yet to decide on HPV-based screening.

We found neither cotesting strategies outperformed stand-alone HC2 or PCR. Between cotests, LBC cotesting was more favorable over Pap cotesting in terms of specificity and PPV. These findings correspond to meta-analysis results of five large randomized trials, although Pap and LBC-based cotesting were not assessed separately (6). In a meta-analysis of observational studies, cotesting demonstrated marginally but significantly higher sensitivity and reduced specificity over HPV testing for CIN2+; however, this was predominantly based on Pap cotesting (15). Furthermore, the higher sensitivity of cotesting could be due to the inconsistent use of the gold standard by some individual studies leading to misclassification bias (15, 33). Although these two studies indirectly compared test accuracy, that is, across study populations or varying trial arms and are thus prone to biases,

our results support the argument that cotesting, regardless of cytology method, does not outperform stand-alone HPV screening in detection.

Current arguments for cotesting are based on retrospective results from the United States, which have demonstrated marginally lower cumulative incidence of CIN3+ under triennial cotesting compared with HPV stand-alone (18). However, the translation of this marginally lower risk by cotesting into real screening practice may not be realized until many tens of thousands of women are screened (13), particularly with opportunistic screening. Cotesting arguments are also further undermined because this strategy leads to greater costs and number of lifetime tests (34, 35). Up to an additional 400 colposcopy referrals per 1,000 women could be expected when cotesting at triennial intervals (34). This evidence highlights screening algorithm complexities, greater costs, and potential harm for apparent minimal gains in detection with cotesting.

On the other hand, positivity to HPV without adequate triage may lead to an increase in colposcopies (36), which could result in overtreatment (7). In our study, colposcopies needed to detect one precancer were greatest under cotesting strategies (17). Between cotests, Pap cotesting incurred a greater degree of harms than LBC cotesting. The latter indicated similar but elevated potential harms compared with stand-alone HPV testing. It is conceivable that screening with other HPV tests detecting mRNA for example can mitigate these costs and harms (37), but these technologies may not be widely available and are not yet approved for stand-alone screening. As we observed, increasing the cutoff of viral load for HPV DNA detection might mitigate false positives, especially if using HC2 (38). In addition, compared with cotesting with triage, fewer colposcopies were needed when screening with HPV 16/18 genotyping and triage, further highlighting the benefit of stand-alone HPV testing (16).

Although observational studies with opportunistic screening (19, 29, 39) do not directly compare cotesting strategies to HPV stand-alone (40–43), our study confirms observations that HPV testing is superior to cytology in detection of precancerous lesions. We observed low accuracy of cytology, particularly for ASC-US+. However sensitivity was higher than previous reports in Germany possibly due to biopsies of nonvisible lesions, but is still low compared with other high-resource countries (3, 39, 44). This might explain why our results were higher than relative sensitivity and specificity from previous studies (3, 43). Possible reasons for poorer accuracy of Pap include the continued use of dry cotton-tipped swabs in screening and lack of standardized quality assurance with opportunistic screening (9). Fewer inadequate samples and from-the-vial testing advantages of LBC may also explain why LBC cotesting performed similarly to stand-alone HPV testing (45). Furthermore, in the same screening context, accuracy of LBC has been reported to be higher than Pap, likely due to the poor quality of the latter (46).

Our results conferred lower HC2 sensitivity than previously reported in Germany (39, 44), possibly because we recruited a random population-based sample via population registries rather than women already attending routine screening. In addition, our sample represents older women. The reduced sensitivity of HC2 for CIN3+ compared with CIN2+ is likely due to the low number of CIN3+ detected. In addition, in our study, all CIN3+ were correctly identified by HC2 cotesting and PCR-based strategies, while one woman with invasive cancer tested stand-alone HC2 negative (Supplementary Table S3). HPV test results may differ possibly due to insufficient viral load, differences in targeted regions of the HPV DNA or cross-reactivity to IARC classified group 2b types (47). Nonetheless, discordance can be avoided by stringent quality assurance and control (9). This is espe-

cially important to note as Germany rolls out cotesting of women ≥ 35 years within an organized screening program, but specific details on approved tests are yet to be defined (21), despite existing criteria and recommendations (48).

Limitations

We report cross-sectional results. Longitudinal outcomes such as cumulative risk incidence among screen-negative women are needed to determine the interval of protection. Nevertheless, we were able to make direct comparisons of distinct cytologic and HPV test strategies within the same study population, which have previously not been reported. Second, despite active reminders for colposcopy, attendance was less than optimal among screen-positives (65.3%) and negatives (35.7%). Historically, follow-up colposcopies in Germany were rather uncommon and the lack of a centralized screening register complicates disease verification. There is still a need for more novel tactics to improve compliance with follow-up of positive screening results and with the roll-out of the new organized program, the latter issue of incomplete data might improve. Accordingly, we adjusted the analyses to account for verification bias and although there may be residual bias due to low sampling fractions of screen-negatives (49), our estimates aligned with previous observations (19, 29, 39, 44). Third, no masking to screening results of the colposcopist and first histopathologist was possible as we attempted to maintain real-world screening. This was addressed by independent second and third histopathology reviews. The number of severe precancerous lesions CIN3+ and cervical carcinomas was also low in our study and we included HPV-unvaccinated women.

Conclusions

We found similar accuracy of stand-alone HPV testing and LBC cotesting, and superior accuracy of stand-alone HPV compared with Pap-based cotesting. However, adding cytology to HPV as a cotest offers nearly no benefit in detection at the cost of more false positive results and colposcopy referrals. For settings optimizing cervical cancer screening such as Germany coming from opportunistic and annual cytology-based screening, triennial cotesting in women 35 years and older is a positive first step toward HPV-based screening. Ultimately, consideration of stand-alone HPV screening once the organized program has been adequately implemented with high quality is warranted. Screening women aged ≥ 30 years with sole HPV-based testing should also be considered in the future to maximize early detection and to further reduce the incidence of cervical cancer toward elimination.

Authors' Disclosures

H. Ikenberg reports co-ownership of a laboratory for cytology and molecular diagnostics. C.J.L.M. Meijer reports personal fees and other from Self-Screen, personal fees from Qiagen, other from MDxHealth, personal fees and other from SPMSD/Merck, and personal fees from GSK outside the submitted work; in addition, C.J.L.M. Meijer has a patent for HPV assay issued and licensed to self-screen and a patent for methylation markers issued and licensed to self-screen. S.J. Klug reports grants from German Cancer Aid (Deutsche Krebshilfe) during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

L.A. Liang: Formal analysis, visualization, methodology, writing—original draft, writing—review and editing. **T. Einzmann:** Investigation. **A. Franzen:** Investigation. **K. Schwarzer:** Investigation. **G. Schaubeger:** Formal analysis, validation, methodology. **D. Schriefer:** Data curation, validation. **K. Radde:** Data curation, project administration. **S.R. Zeissig:** Project administration. **H. Ikenberg:** Investigation. **C.J.L.M. Meijer:** Investigation. **C.J. Kirkpatrick:** Investigation.

H. Kölbl: Investigation. **M. Blettner:** Conceptualization, funding acquisition. **S.J. Klug:** Conceptualization, supervision, writing–review, funding acquisition.

Acknowledgments

The authors would like to thank the following people for their contributions to the MARZY study: Natalja Dik, Sabine Tensing, Martina Wankmüller, Dr. Meike Rensing, Sebastian Czech, Tanja Heinemann, Larissa Tarasenko, Dagmar Lautz, Veronika Weyer, Dr. Gabriele von Wahlert, Dr. Tanja Neunhöffer, Dr. Jean Baptist du Prel, and all colleagues at IMBEI who supported the study. The authors extend special thanks to the late Prof. Peter J.F. Snijders (Amsterdam UMC, location VUMC), who provided the HPV data obtained by GP5+/6+ consensus PCR to the MARZY study, and to all participating office-based gynecologists, general practitioners, pathologists, physicians, and other cooperation partners for their support in

the MARZY study. The MARZY study was funded by the German Cancer Aid [Deutsche Krebshilfe (DKH), No. 105827, 106619, 107247, 108047, and 107159]. Research reported in this publication was supported by these DKH grants. All DKH grants were directed to both recipients S.J. Klug and M. Blettner.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 9, 2020; revised September 17, 2020; accepted November 2, 2020; published first November 13, 2020.

References

- Vaccarella S, Lortet-Tieulent J, Plummer M, Franceschi S, Bray F. Worldwide trends in cervical cancer incidence: impact of screening against changes in disease risk factors. *Eur J Cancer* 2013;49:3262–73.
- Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet Global Health* 2020;8:e191–203.
- Koliopoulos G, Nyaga VN, Santesso N, Bryant A, Martin-Hirsch PP, Mustafa RA, et al. Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database Syst Rev* 2017;8:CD008587.
- de Sanjosé S, Quint WG, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, et al. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol* 2010;11:1048–56.
- Drolet M, Bénard É, Pérez N, Brisson M, Ali H, Boily M-C, et al. Population-level impact and herd effects following the introduction of human papillomavirus vaccination programmes: updated systematic review and meta-analysis. *Lancet* 2019;394:497–509.
- Arbyn M, Ronco G, Anttila A, Meijer CJ, Poljak M, Ogilvie G, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. *Vaccine* 2012;30:F88–99.
- Melnikow J, Henderson JT, Burda BU, Senger CA, Durbin S, Weyrich MS. Screening for cervical cancer with high-risk human papillomavirus testing: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA* 2018;320:687–705.
- Wright TC, Stoler MH, Behrens CM, Sharma A, Zhang G, Wright TL. Primary cervical cancer screening with human papillomavirus: end of study results from the ATHENA study using HPV as the first-line screening test. *Gynecol Oncol* 2015;136:189–97.
- von Karsa L, Arbyn M, De Vuyst H, Dillner J, Dillner L, Franceschi S, et al. European guidelines for quality assurance in cervical cancer screening. Summary of the supplements on HPV screening and vaccination. *Papillomavirus Res* 2015; 1:22–31.
- Fontham ETH, Wolf AMD, Church TR, Etzioni R, Flowers CR, Herzog A, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. *CA Cancer J Clin* 2020;70:321–46.
- Polman NJ, Snijders PJF, Kenter GG, Berkhof J, Meijer CJLM. HPV-based cervical screening: rationale, expectations and future perspectives of the new Dutch screening programme. *Prev Med* 2019;119:108–17.
- Blatt AJ, Kennedy R, Luff RD, Austin RM, Rabin DS. Comparison of cervical cancer screening results among 256,648 women in multiple clinical practices. *Cancer Cytopathol* 2015;123:282–8.
- Stoler MH, Austin RM, Zhao C. Point-counterpoint: cervical cancer screening should be done by primary human papillomavirus testing with genotyping and reflex cytology for women over the age of 25 years. *J Clin Microbiol* 2015;53: 2798–804.
- Wentzensen N, Arbyn M. HPV-based cervical cancer screening- facts, fiction, and misperceptions. *Prev Med* 2017;98:33–5.
- Koliopoulos G, Arbyn M, Martin-Hirsch P, Kyrgiou M, Prendiville W, Paraskevaidis E. Diagnostic accuracy of human papillomavirus testing in primary cervical screening: a systematic review and meta-analysis of non-randomized studies. *Gynecol Oncol* 2007;104:232–46.
- Cox JT, Castle PE, Behrens CM, Sharma A, Wright TC Jr, Cuzick J. Comparison of cervical cancer screening strategies incorporating different combinations of cytology, HPV testing, and genotyping for HPV 16/18: results from the ATHENA HPV study. *Am J Obstet Gynecol* 2013;208:184.
- Schiffman M, Kinney WK, Cheung LC, Gage JC, Fetterman B, Poitras NE, et al. Relative performance of HPV and cytology components of cotesting in cervical screening. *J Natl Cancer Inst* 2018;110:501–8.
- Demarco M, Lorey TS, Fetterman B, Cheung LC, Guido RS, Wentzensen N, et al. Risks of CIN 2+, CIN 3+, and cancer by cytology and human papillomavirus status: The foundation of risk-based cervical screening guidelines. *J Low Genit Tract Dis* 2017;21:261–7.
- Baseman JG, Kulasingam SL, Harris TG, Hughes JP, Kiviat NB, Mao C, et al. Evaluation of primary cervical cancer screening with an oncogenic human papillomavirus DNA test and cervical cytologic findings among women who attended family planning clinics in the United States. *Am J Obstet Gynecol* 2008; 199:26.
- Ronco G, Segnan N, Giorgi-Rossi P, Zappa M, Casadei GP, Carozzi F, et al. Human papillomavirus testing and liquid-based cytology: results at recruitment from the new technologies for cervical cancer randomized controlled trial. *J Natl Cancer Inst* 2006;98:765–74.
- Gemeinsamer Bundesausschuss (G-BA). Richtlinie des Gemeinsamen Bundesausschusses für organisierte Krebsfrüherkennungsprogramme: oKFE-Richtlinie/oKFE-RL: Gemeinsamer Bundesausschuss (G-BA); 2018 [updated 1 Jan 2020]. Available from: <https://www.g-ba.de/downloads/62-492-1685/oKFE-RL-2018-08-02-iK-2018-10-19.pdf>.
- Radde K, Gottschalk A, Bussas U, Schüle S, Schriefer D, Seifert U, et al. Invitation to cervical cancer screening does increase participation in Germany: results from the MARZY study. *Int J Cancer* 2016;139:1018–30.
- Zeissig SR, Radde K, Kaiser M, Blettner M, Klug SJ. Quality assurance in an epidemiological cohort study: on-site monitoring in gynaecological practices. *Z Evid Fortbild Qual Gesundheitswes* 2014;108:517–27.
- Cirkel C, Barop C, Beyer DA. Method comparison between Munich II and III nomenclature for Pap smear samples. *J Turk Ger Gynecol Assoc* 2015;16: 203–7.
- Arbyn M, Sasieni P, Meijer CJ, Clavel C, Koliopoulos G, Dillner J. Chapter 9: clinical applications of HPV testing: a summary of meta-analyses. *Vaccine* 2006; 24:S3/78–89.
- Herbert A, Bergeron C, Wiener H, Schenk U, Klinkhamer P, Bulten J, et al. European guidelines for quality assurance in cervical cancer screening: recommendations for cervical cytology terminology. *Cytopathology* 2007;18:213–9.
- International Agency for Research on Cancer. A review of human carcinogens. Part B: biological agents. Lyon, France: IARC; 2012.
- Walker P, Dexeus S, De Palo G, Barrasso R, Campion M, Girardi F, et al. International terminology of colposcopy: an updated report from the International Federation for Cervical Pathology and Colposcopy. *Obstet Gynecol* 2003; 101:175–7.
- Kulasingam SL, Hughes JP, Kiviat NB, Mao C, Weiss NS, Kuypers JM, et al. Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: comparison of sensitivity, specificity, and frequency of referral. *JAMA* 2002;288:1749–57.
- Efron B, Tibshirani RJ. An introduction to the bootstrap. New York, NY: Chapman & Hall; 1993.
- Haldane J. The estimation and significance of the logarithm of a ratio of frequencies. *Ann Hum Genet* 1956;20:309–11.
- Basu P, Ponti A, Anttila A, Ronco G, Senore C, Vale DB, et al. Status of implementation and organization of cancer screening in The European Union Member States—Summary results from the second European screening report. *Int J Cancer* 2018;142:44–56.

33. Arbyn M, Sankaranarayanan R, Muwonge R, Keita N, Dolo A, Mbalawa CG, et al. Pooled analysis of the accuracy of five cervical cancer screening tests assessed in eleven studies in Africa and India. *Int J Cancer* 2008;123:153–60.
34. Kim JJ, Burger EA, Regan C, Sy S. Screening for cervical cancer in primary care: a decision analysis for the US Preventive Services Task Force. *JAMA* 2018;320:706–14.
35. Petry KU, Barth C, Wasem J, Neumann A. A model to evaluate the costs and clinical effectiveness of human papilloma virus screening compared with annual papanicolaou cytology in Germany. *Eur J Obstet Gynecol Reprod Biol* 2017;212:132–9.
36. Rebolj M, Rimmer J, Denton K, Tidy J, Mathews C, Ellis K, et al. Primary cervical screening with high risk human papillomavirus testing: observational study. *BMJ* 2019;364:l240.
37. Felix JC, Lacey MJ, Miller JD, Lenhart GM, Spitzer M, Kulkarni R. The clinical and economic benefits of co-testing versus primary HPV testing for cervical cancer screening: a modeling analysis. *J Womens Health* 2016;25:606–16.
38. Rijkaart DC, Coupe VMH, van Kemenade FJ, Heideman DAM, Hesselink AT, Verweij W, et al. Comparison of hybrid capture 2 testing at different thresholds with cytology as primary cervical screening test. *Br J Cancer* 2010;103:939–46.
39. Iftner T, Becker S, Neis K-J, Castanon A, Iftner A, Holz B, et al. Head-to-head comparison of the RNA-based aptima human papillomavirus (HPV) assay and the DNA-based hybrid capture 2 HPV test in a routine screening population of women aged 30 to 60 years in Germany. *J Clin Microbiol* 2015;53:2509–16.
40. Kitchener HC, Almonte M, Thomson C, Wheeler P, Sargent A, Stoykova B, et al. HPV testing in combination with liquid-based cytology in primary cervical screening (ARTISTIC): a randomised controlled trial. *Lancet Oncol* 2009;10:672–82.
41. Naucler P, Ryd W, Tornberg S, Strand A, Wadell G, Elfgrén K, et al. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. *N Engl J Med* 2007;357:1589–97.
42. Rijkaart DC, Berkhof J, Rozendaal L, van Kemenade FJ, Bulkman NW, Heideman DA, 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.
43. Ronco G, Dillner J, Elfstrom KM, Tunesi S, Snijders PJ, Arbyn M, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet* 2014;383:524–32.
44. Petry KU, Menton S, Menton M, van Loenen-Frosch F, de Carvalho Gomes H, Holz B, et al. Inclusion of HPV testing in routine cervical cancer screening for women above 29 years in Germany: results for 8466 patients. *Br J Cancer* 2003;88:1570–7.
45. Ronco G, Cuzick J, Pierotti P, Cariaggi MP, Palma PD, Naldoni C, et al. Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial. *BMJ* 2007;335:28.
46. Klug SJ, Neis KJ, Harlfinger W, Malter A, König J, Spieth S, et al. A randomized trial comparing conventional cytology to liquid-based cytology and computer assistance. *Int J Cancer* 2013;132:2849–57.
47. de Thurah L, Bonde J, Lam JUH, Rebolj M. Concordant testing results between various human papillomavirus assays in primary cervical cancer screening: systematic review. *Clin Microbiol Infect* 2018;24:29–36.
48. Arbyn M, Snijders PJ, Meijer CJ, Berkhof J, Cuschieri K, Kocjan BJ, et al. Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening? *Clin Microbiol Infect* 2015;21:817–26.
49. Arbyn M, Ronco G, Cuzick J, Wentzensen N, Castle PE. How to evaluate emerging technologies in cervical cancer screening? *Int J Cancer* 2009;125:2489–96.