



OPEN

Comparative accuracy of cervical cancer screening strategies in healthy asymptomatic women: a systematic review and network meta-analysis

Teruhiko Terasawa^{1✉}, Satoyo Hosono², Seiju Sasaki³, Keika Hoshi⁴, Yuri Hamashima⁵, Takafumi Katayama⁶ & Chisato Hamashima⁷

To compare all available accuracy data on screening strategies for identifying cervical intraepithelial neoplasia grade ≥ 2 in healthy asymptomatic women, we performed a systematic review and network meta-analysis. MEDLINE and EMBASE were searched up to October 2020 for paired-design studies of cytology and testing for high-risk genotypes of human papillomavirus (hrHPV). The methods used included a duplicate assessment of eligibility, double extraction of quantitative data, validity assessment, random-effects network meta-analysis of test accuracy, and GRADE rating. Twenty-seven prospective studies (185,269 subjects) were included. The combination of cytology (atypical squamous cells of undetermined significance or higher grades) and hrHPV testing (excepting genotyping for HPV 16 or 18 [HPV16/18]) with the either-positive criterion (OR rule) was the most sensitive/least specific, whereas the same combination with the both-positive criterion (AND rule) was the most specific/least sensitive. Compared with standalone cytology, non-HPV16/18 hrHPV assays were more sensitive/less specific. Two algorithms proposed for primary cytological testing or primary hrHPV testing were ranked in the middle as more sensitive/less specific than standalone cytology and the AND rule combinations but more specific/less sensitive than standalone hrHPV testing and the OR rule combination. Further research is needed to assess these results in population-relevant outcomes at the program level.

Cervical cancer is the fourth most frequently diagnosed cancer and fourth most common cause of cancer-specific mortality in women, with a worldwide estimated prevalence of 570,000 cases and 311,000 associated deaths in 2018^{1,2}. Observational studies have clearly demonstrated a reduction in the invasive cancer incidence and mortality in well-organized screening programs using cervical cytological testing that have been implemented³. Moreover, randomized controlled trials (RCTs) of well-screened populations have shown that strategies incorporating testing for high-risk human papillomavirus (hrHPV) subtypes, which are the central etiological agents of cervical cancer pathogenesis⁴, were, in aggregate, associated with a reduction in the invasive cancer incidence relative to that shown by cytological screening alone⁵. Therefore, current guidelines recommend three primary screening options: cytological testing alone, standalone hrHPV testing, and cytological + hrHPV combination testing (co-testing)^{6–10}. However, subsequent management strategies for women with positive primary testing are complex. Although specific triage and/or follow-up testing algorithms for primary cytology and co-testing¹¹

¹Section of General Internal Medicine, Department of Emergency and General Internal Medicine, Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukakecho, Toyoake, Aichi 470-1192, Japan. ²Division of Cancer Screening Assessment and Management, Center for Public Health Science, National Cancer Center, Tokyo, Japan. ³Center for Preventive Medicine, St. Luke's International Hospital Affiliated Clinic, Tokyo, Japan. ⁴Center for Public Health Informatics, National Institute of Public Health, Wako, Japan. ⁵Department of Population Health Science, Bristol Medical School, University of Bristol, Bristol, UK. ⁶Department of Statistics and Computer Science, College of Nursing Art and Science, University of Hyogo, Hyogo, Japan. ⁷Department of Nursing, Faculty of Medical Technology, Teikyo University, Tokyo, Japan. ✉email: terasawa@fujita-hu.ac.jp

and for primary hrHPV testing⁹ have been proposed, the evidence base to improve patient-important outcomes with these algorithms is immature.

The comparative effectiveness of alternative screening strategies should be based on a comprehensive assessment of benefits and harms. Given the low incidence and mortality due to cervical cancer in high-income countries and the challenges associated with conducting de novo large and long-term RCTs, decision modeling is an alternative realistic option to better understand the theoretical utility of the screening options¹². In this regard, comprehensive synthesis of the screening accuracy, a key model parameter of cytological and hrHPV testing and their available combination algorithms reported in rigorously conducted paired-design studies, is a valuable intermediate step. However, recent meta-analyses have focused on either standalone cytological and/or hrHPV testing^{13–15} or a comparison of cytological testing with a specific combination algorithm not proposed in guidelines only¹⁶.

For those studies that assessed the diagnostic accuracy of selected and different pairs of tests of interest and their combination algorithms, network meta-analysis of diagnostic test accuracy studies is a useful approach that can compare all the assessed tests and combination algorithms in a single analysis¹⁷. The current study aimed to perform network meta-analysis to quantitatively compare and rank the cross-sectional accuracy of all reported screening algorithms based on cytological and hrHPV testing. We specifically focused on the comparative accuracy of guideline-proposed combination algorithms by examining data derived from primary studies of healthy asymptomatic women that addressed verification bias because such bias is commonly observed in cancer screening accuracy studies.

Methods

This extended systematic review is based on an update evidence review conducted for revision of the Japanese Guidelines for Cervical Cancer Screening^{18,19}. Although the complete evidence review was planned before analysis, no protocol was registered for this extended review. This report followed PRISMA guidelines for diagnostic test accuracy (PRISMA-DTA)²⁰ and did not require ethics review or patient consent.

Search strategy. We searched OVID MEDLINE and EMBASE for publications between January 1, 1992, and October 14, 2020, with no language restrictions. The search strategies are detailed in the Supplementary methods. Complementarily, the reference lists of eligible studies and relevant review articles were also screened for other appropriate studies.

Study eligibility. Three paired reviewers independently double screened the first 3000 abstracts in a calibration phase. The same reviewers single screened the remaining abstracts. Two reviewers independently determined the eligibility of potential full-text articles, with discrepancies adjudicated by a third reviewer. Only fully paired-design screening studies of cytology and hrHPV testing, either opportunistic or organized screening, aimed at detecting cervical intraepithelial neoplasia \geq grade 2 (CIN2+) in healthy asymptomatic women were eligible for inclusion. We included all studies that performed either routine colposcopy-directed biopsy or colposcopy and selective biopsy in all screened women to verify target lesions along with studies that performed either of the colposcopy methods among women with protocol-specified screening results and statistical corrections for data from unverified samples. In studies that analyzed both eligible and ineligible populations, only those with relevant and extractable data were included. In case of multiple publications, we included the publication with the largest sample size (see Supplementary methods for more details).

Data extraction. One reviewer extracted descriptive data, which were independently confirmed by another reviewer. Next, two reviewers independently extracted numerical data, with discrepancies resolved by consensus. We preferred cross-tabulated count data over reported accuracy estimates when both data types were extractable (see Supplementary methods for more details).

Operationalization. Cytology results were standardized according to the Bethesda system^{21,22} if other classification systems had been used. For studies that used both conventional and liquid-based cytology tests (CC and LBC, respectively), we favored LBC data over CC data; we jointly analyzed both smear preparation methods.

Operationally, hrHPV assays were categorized into four groups: hybridization with signal amplifications of DNA (e.g., Hybrid Capture 2 [HC2], Qiagen, Gaithersburg, MD), polymerase chain reaction (PCR) of DNA from ≥ 13 hrHPV genotypes, amplification of E6/E7 viral messenger RNA (mRNA), and assays identifying DNA or RNA of genotypes, either HPV16 or HPV18 or both (HPV16/18)²³. For mRNA-based genotyping, since the genotype HPV45 was additionally targeted with HPV16 and HPV18 (HPV16/18/45), we adopted these results. HC2 positivity was defined as ≥ 1.0 relative light units. We did not assess point-of-care testing platforms (e.g., careHPV, Qiagen, Gaithersburg, MD).

We operationally categorized combination tests as follows: (i) combination algorithms based on the OR rule (women with either test positive were categorized as screening positive while women with both tests negative as screening negative) or the AND rule (women with both tests positive were categorized as screening positive while women with at least 1 negative test as negative)²⁴; (ii) thresholds for cytological testing as, e.g., undetermined significance or worse grades (\geq ASCUS), or low- or high-grade squamous intraepithelial lesions or worse grades (\geq LSIL or \geq HSIL, respectively); and (iii) hrHPV assays (Table 1)^{6–9,25}. As cross-sectional representation of guideline-proposed algorithms, we assessed two specific strategies: “ \geq LSIL OR [hrHPV AND ASCUS]”, which classified only women with cytologic testing \geq LSIL, or both by cytologic testing ASCUS and hrHPV testing positive as screening positive; and “HPV16/18(/45) OR [hrHPV AND \geq ASCUS]”, which classified only women

	Methods	FDA-approved systems
Standalone or component tests		
Cytology		
Conventional or liquid-based cytology	A specified cytological grade, ASCUS, ASCH, LSIL, or HSIL (and higher grades if indicated) defined by the Bethesda system is used as the positive criterion	SurePath (TriPath Imaging, Inc., Burlington, NC) and ThinPrep (Cytec Corp, Marlborough, MA)
hrHPV assays		
Signal amplifications	Nucleic acid hybridization with signal amplification or cleavage-based signal amplification of ≥ 13 hrHPV DNA	Hybrid Capture 2 (Qiagen, Gaithersburg, MD); Cervista HPV HR (Hologic, Madison, WI)
PCR-based tests	PCR of DNA from ≥ 13 hrHPV genotypes	Cobas HPV test (Roche Molecular Systems, Pleasanton, CA)
mRNA-based tests	Amplification of E6/E7 viral mRNA	Aptima (Hologic, Bedford, MA)
HPV 16/18/(45)	Genotyping assays identifying DNA or RNA of HPV genotypes either 16 or 18 or both	Cobas HPV test (Roche Molecular Systems, Pleasanton, CA); Aptima HPV 16 and HPV18/45 (Hologic, Bedford, MA); Cervista HPV 16/18 (Hologic, Madison, WI)
Specific algorithms	Test operation system	Definition of screening strategy positive
Combination tests		
AND rule (e.g., hrHPV AND \geq ASCUS)	Co-testing or serial testing	Women with both positive tests as screening positive and women with ≥ 1 negative test as screening negative
OR rule (e.g., hrHPV OR \geq ASCUS)	Co-testing only	Women with ≥ 1 positive test as screening positive and women with both negative tests as screening negative
\geq LSIL OR [hrHPV AND ASCUS]	Co-testing or primary cytologic testing with reflex to hrHPV testing by a joint guideline published in 2012 among ACS, ASCCP, and ASCP	Cytologic testing \geq LSIL, or both by cytologic testing ASCUS and hrHPV testing positive
HPV16/18/(45) OR [hrHPV AND \geq ASCUS]	Co-testing or primary hrHPV testing with HPV genotyping with reflex to cytologic testing proposed by the SGO 2015	HPV genotypes 16 or 18 positive, or both cytologic testing \geq ASCUS and hrHPV testing positive for non-16/18 hrHPV genotypes

Table 1. Operational categorizations of cytological testing, assays for hrHPV testing, and their combination algorithms. ACS American Cancer Society, ASCCP American Society for Colposcopy and Cervical Pathology, ASCH atypical squamous cells, cannot exclude HSIL, ASCP American Society for Clinical Pathology, ASCUS atypical squamous cells of undetermined significance, BPR both-positive rule, DNA deoxyribonucleic acid, EPR either-positive rule, hrHPV high-risk human papillomavirus, HPV human papillomavirus, HR high risk, HSIL high-grade squamous intraepithelial lesion, LSIL low-grade squamous intraepithelial lesion, mRNA messenger ribonucleic acid, SGO Society for Gynecologic Oncology.

with HPV genotypes 16 or 18 (or 45) positive, or both cytologic testing \geq ASCUS and hrHPV testing positive for non-16/18/(45) hrHPV genotypes as screening positive (Table 1).

Quality assessment. Paired independent reviewers double rated the validity of a study using a risk of bias tool for comparative diagnostic accuracy studies (QUADAS-C)²⁶, an extension to the existing Quality Assessment of Diagnostic Accuracy Studies 2 tool²⁷. Discrepancies were resolved via consensus. Operationally, a study was defined to have low risk of verification bias only when all screened samples had been histologically verified.

Data synthesis and statistical analysis. The primary outcome was sensitivity and specificity for detecting CIN2+. We used their relative risk values for and absolute differences in (Δ) sensitivity and specificity for any paired alternative screening algorithms (e.g., a standalone test vs. a combination algorithm) as measures of comparative accuracy.

Between-study heterogeneity was assessed visually by using crosshair plots of sensitivity and specificity estimates in the receiver operating characteristic (ROC) space²⁸. We calculated the average sensitivity and specificity estimates and their derived relative and Δ sensitivity and specificity values with their corresponding 95% credible intervals (CrIs) by using an arm-based, two-stage hierarchical, Bayesian bivariate random-effects network meta-analysis model²⁹. Credible regions for the average estimates were constructed by using the standard method³⁰. For comparison, we also calculated average sensitivity and specificity estimates separately by using the standard bivariate meta-analysis model for diagnostic accuracy³¹. Hierarchical summary ROC (HSROC) curves were derived on the basis of the estimated parameters³².

We performed study-level univariable meta-regression for the following prespecified binary predictors when ≥ 10 studies were available: study location (countries ranked as “very high human development” by the Human Development Index 2017³³ vs. those that were not), study design (histology-based vs. colposcopy-based verification), and type of sample collectors (physicians vs. nonphysicians). Scarce data on young individuals (<30 years old) precluded meta-regression based on age. Complete details of the methodology, model fitting, choice of prior distributions for parameters assessed, and operational definitions used in sensitivity analyses are provided in the Supplementary methods.

We used the Grading of Recommendation Assessment, Development, and Evaluation (GRADE) tool³⁴ to assess the certainty of evidence and focused on the comparisons among cytological testing (\geq ASCUS) alone, standalone hrHPV assays, and the guideline-proposed combination algorithms. For calculating false negatives

First author and publication year	Country	Enrollment year	N	Target age	Cytologic test	HPV assay
Histology-based studies						
Belinson (2001) ^{36–38}	China	1999	1997	35–45	LBC	HC2
Cárdenas-Turanzas (2008) ³⁹	USA; Canada	1998–2005	957	≥ 30	CC	HC2
Hovland (2010) ⁴⁰	Congo	2003	313	25–60	CC; LBC	PCR
Histology-based correction studies						
Schneider (2000) ⁴¹	Germany	1996–1998	4761	18–70	CC	PCR
Kulasingam (2002) ^{42,43}	USA	1997–2000	4075	18–50	LBC	HC2
Bigras (2005) ⁴⁴	Switzerland	2002–2004	13,842	17–93	LBC	HC2
Mayrand (2007) ^{45,46}	Canada	2002–2004	10,154	30–69	CC	HC2
Li (2009) ⁴⁷	China	2004–2005	2562	15–59	LBC	HC2
Castle (2011) ^{48,49}	USA	2008–2009	41,026	25–93	LBC	HPV16/18
Mahmud (2012) ⁵⁰	Congo	2003–2004	1528	≥ 30	CC	HC2
Sangrajrang (2017) ^{51,52}	Thailand	2014–2015	5046	30–60	CC	PCR; mRNA; HPV16/18
Kurokawa (2018) ⁵³	Japan	2015–2016	7585	25–69	LBC	PCR; HPV16/18
Colposcopy-based studies						
Blumenthal (2001) ^{54,55}	Zimbabwe	1995–1997	2073	25–55	CC	HC2
Coste (2003) ^{56,57}	France	1999–2000	1324	18–	CC; LBC	HC2
Sankaranarayanan (2004) ⁵⁸	India	1999–2003	18,085	25–65	CC	HC2
Qiao (2008) ⁵⁹	China	2007	2388	30–54	LBC	HC2
McAdam (2010) ⁶⁰	Vanuatu	2006	494	30–50	CC	HC2
Quincy (2012) ⁶¹	Nicaragua	ND	245	25–60	LBC	HC2
Colposcopy-based correction studies						
Cuzick (2003) ⁶²	UK	1998–2001	10,358	30–60	CC	HC2
Petry (2003) ⁶³	Germany	1998–2000	7908	≥ 30	CC	HC2
Gravitt (2010) ⁶⁴	India	2005–2007	2331	≥ 25	CC	HC2
Moy (2010) ⁶⁵	China	2003–2006	9057	30–54	LBC	HC2
Monsonogo (2011) ⁶⁶	France	2008–2009	4429	20–65	LBC	HC2; mRNA
Ferreccio (2013) ⁶⁷	Chile	2009–2010	8265	25–64	CC	HC2
Agorastos (2015) ⁶⁸	Greece	2011–2013	3993	25–55	LBC	PCR
Iftner (2015) ⁶⁹	Germany	ND	9451	30–60	LBC	HC2; mRNA; HPV16/18
Wu (2017) ⁷⁰	China	2015	11,064	21–65	LBC	PCR

Table 2. Study, participant, and screening test characteristics. *CC* conventional cytology, *HC2* Hybrid Capture 2, *HPV* human papillomavirus, *LBC* liquid-based cytology, *mRNA* messenger ribonucleic acid, *PCR* polymerase chain reaction.

(FNs) and false positives (FPs), we assumed a healthy screening population of 1,000 women in which 20 are CIN2+ (i.e., a prevalence of 2%)¹³.

We did not evaluate funnel-plot asymmetry because the required tests did not permit valid assessment of the extent and impact of missing studies²⁰. All analyses were performed by using WinBUGS 1.4.3 (MRC Biostatistics Unit, Cambridge, UK) and Stata/SE 16.1 (Stata Corp, College Station, TX)³⁵. We estimated the probability that the true value (i.e., posterior distribution) of relative sensitivity or specificity was ≥ 1 (or ≤ 1) as a measure of superiority of a test over a comparator test. A conventional, frequentist, two-tailed *P*-value of 0.05 corresponds to a Bayesian posterior probability of 0.025, which we considered to be the threshold of statistical significance.

Results

Study selection. Our literature search identified 15,488 citations, of which 27 prospective studies reported in 35 publications corresponding to 185,269 women were included for the meta-analysis (Supplementary Fig. S1)^{36–70}. Supplementary material provides a list of excluded studies.

Characteristics of included studies. All included studies had a prospective design, and 14 studies (52%) were from high-income countries (Table 2). The average age of study participants ranged from 25 to 47 years. Data on type of sample collectors was available for 20 studies (74%), with physician collectors in 14 studies and nonphysician providers, typically trained nurses or midwives, in 6 studies. Thirteen studies had used only CC, and 12 had adopted only LBC, whereas two other studies had used both CC and LBC (Table 2). Of the four available hrHPV testing subgroups, HC2 was the most commonly reported hrHPV assay (assessed in 20 studies), whereas six studies assessed PCR-based tests, four genotyped for HPV16/18, and three used mRNA-based tests, of which also genotyped for HPV16/18/45. Data on one or more combination algorithm(s) were available in 19 studies (reported in 20 publications; 70%). The most commonly assessed combinations were

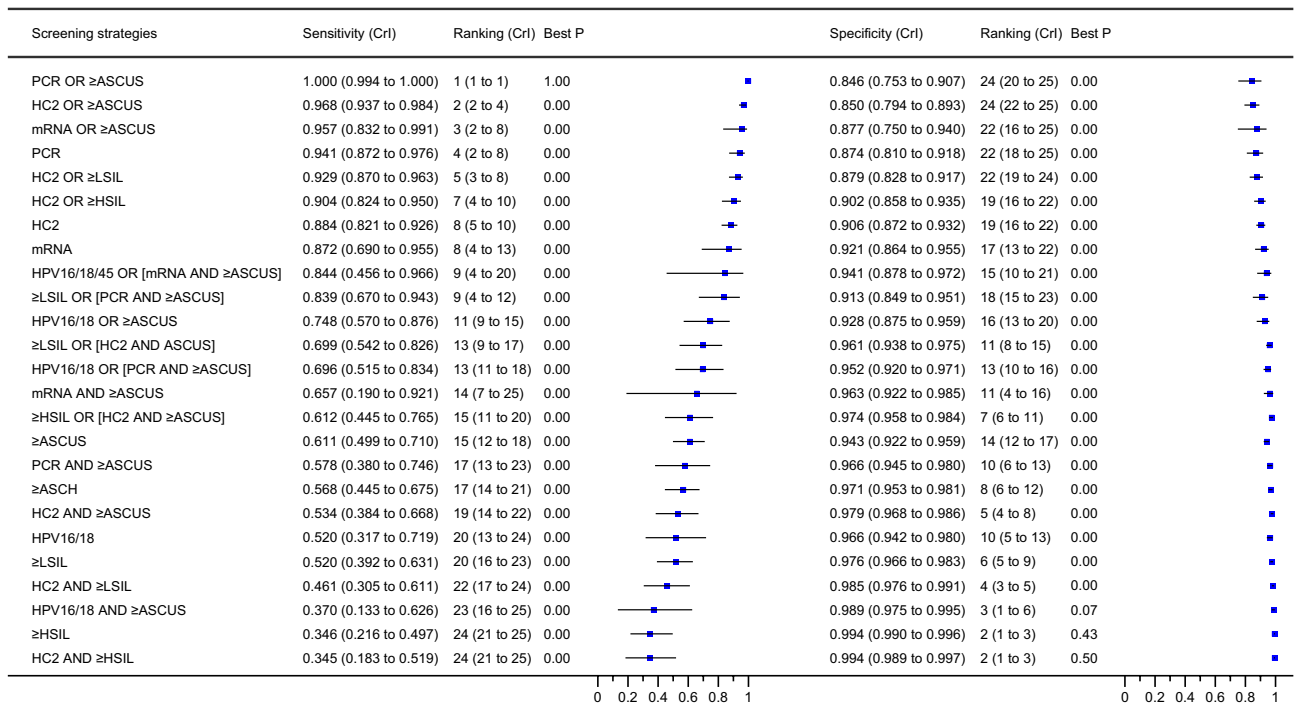


Figure 2. Average sensitivity and specificity and ranking of standalone tests and combination algorithms for cervical cancer screening for detecting CIN2+. Point estimates (blue squares) and CrIs (extending lines) are presented (ordered by the ranking of each test/combination's sensitivity). See Table 1 for the definition of each strategy. *ASCH* atypical squamous cells cannot exclude high-grade lesion, *ASCUS* atypical squamous cells of undetermined significance, *CrI* 95% credible interval, *HC2* Hybrid Capture 2, *HPV16/18/45* genotyping for HPV types 16 or 18 (or 45), *HSIL* high-grade squamous intraepithelial lesion, *LBC* liquid-based cytology, *LSIL* low-grade squamous intraepithelial lesion, *mRNA* messenger ribonucleic acid, *PCR* polymerase chain reaction.

Figure 2 provides the average accuracy estimates and ranking estimated through the network meta-analysis. Overall, the combinations with the OR rule of hrHPV and cytological testing were most sensitive and least specific, whereas combinations with the AND rule of hrHPV and cytological testing were most specific and least sensitive. The rankings estimated in the network meta-analysis reflected the trade-off between sensitivity and specificity by altering the thresholds; lowering the thresholds of cytological testing (e.g., from ≥HSIL to ≥ASCUS) led to higher sensitivity but at the cost of reduced specificity, and tightening the thresholds increased specificity at the cost of reduced sensitivity. This behavior resulted in average estimates and rankings for tests or combination algorithms relying on few studies (e.g., HPV16/18- and mRNA-based combinations assessed in only one study each), which were inconsistent with the standard meta-analysis.

In the network meta-analysis, PCR OR ≥ASCUS was most sensitive (1.0; CrI: 0.994–1.0; probability of best sensitivity: 1.0) but was one of the two least specific screening algorithms (0.846; CrI: 0.753–0.907). In contrast, standalone ≥HSIL and HC2 AND ≥HSIL were the two most specific (respectively, 0.994 [CrI: 0.990–0.996; probability of best specificity: 0.43] and 0.994 [CrI: 0.989–0.997]; probability of best specificity: 0.50) but were the two least sensitive (respectively, 0.346 [95% CrI: 0.216–0.497] and 0.345 [CrI: 0.183–0.519]) algorithms.

Comparative accuracy. Supplementary Figure S6, Supplementary Tables S5 and S6, respectively, summarize the average relative sensitivity and specificity and ΔFNs and ΔFPs estimated based on a population of 1000 healthy women, with a 2% prevalence of CIN2+, across all possible paired comparisons of available standalone tests and combination algorithms.

Comparative accuracy of standalone tests. For cytological testing, the average relative estimates of screening accuracy reflected the effect of altering the thresholds (Fig. 3a, Supplementary Table S7). For example, ≥ASCUS was more sensitive than ≥LSIL (relative sensitivity: 0.86 [CrI: 0.69–0.97; Bayesian $P(\geq 1) < 0.001$]) but less specific than ≥LSIL (relative specificity: 1.03 [CrI: 1.05–1.02; Bayesian $P(\leq 1) < 0.001$]). Two studies that directly compared the alternative smear preparation methods showed identical sensitivity and specificity for CC and LBC for each threshold (Supplementary Fig. S7).

HPV16/18 was more specific but less sensitive than the other hrHPV assays (Fig. 3a, Supplementary Table S7). For example, for comparing HPV16/18 with HC2, the relative specificity was 1.06 [CrI: 1.10–1.04; Bayesian $P(\leq 1) < 0.001$] and relative sensitivity was 0.59 [CrI: 0.36–0.81; Bayesian $P(\geq 1) < 0.001$]. Among HC2, PCR-based tests, and mRNA-based tests, data were limited as to whether a specific hrHPV assay was more sensitive or specific than any other. For example, although the PCR-based tests appeared more sensitive but less specific than HC2, the CrIs for the relative accuracy crossed 1, the null value (i.e., the relative sensitivity of PCR vs. HC2

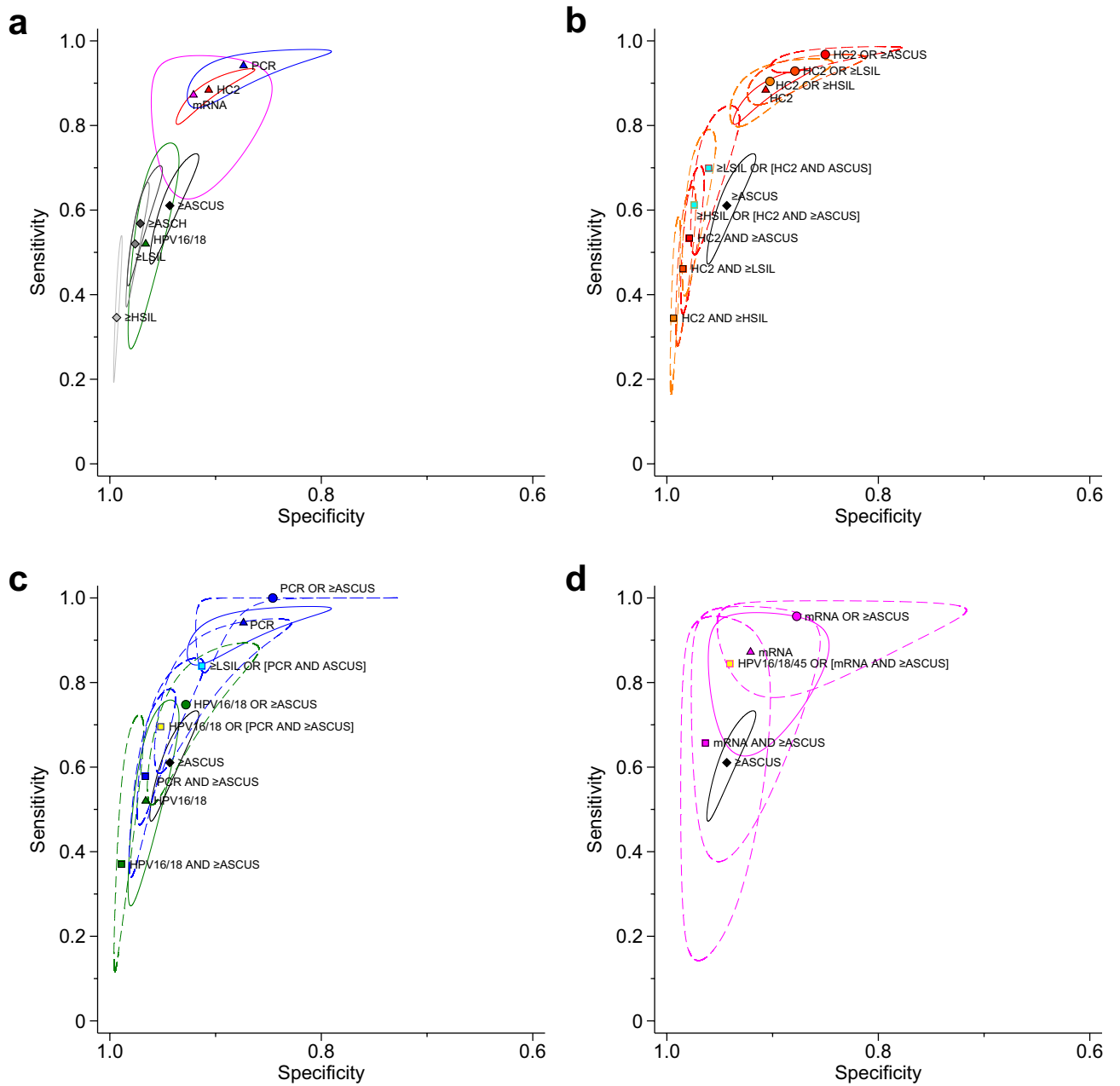


Figure 3. Network meta-analysis of standalone tests and combination algorithms for cervical cancer screening for detecting CIN2+. Average sensitivity and specificity and their 95% credible regions for (a) standalone cytology or hrHPV testing, (b) HC2-based combination algorithms, (c) PCR-based combination algorithms (including PCR-based genotyping for HPV16/18), and (d) mRNA-based combination algorithms (including mRNA-based genotyping for HPV16/18/45). Graded colors (black, dark gray, gray, and light gray) indicate cytology with a specific threshold, red indicates HC2, blue indicates PCR-based tests, green indicates HPV16/18, and magenta indicates mRNA-based tests. Triangles and diamonds represent standalone hrHPV testing and cytology, respectively. Circles and squares represent combinations based on the OR-rule and the AND-rule, respectively. For combination algorithms (b–d), standalone component hrHPV testing and cytology (\geq ASCUS) are also presented as reference. See Table 1 for the definition of each strategy. ASCUS atypical squamous cells cannot exclude high-grade lesion, ASCUS atypical squamous cells of undetermined significance, HC2 Hybrid Capture 2, HPV16/18(/45) genotyping for HPV types 16 or 18 (or 45), HSIL high-grade squamous intraepithelial lesion, LBC liquid-based cytology, LSIL low-grade squamous intraepithelial lesion, mRNA messenger ribonucleic acid, PCR polymerase chain reaction.

Index and comparator tests or combination algorithms	Index (for specificity) and comparator (for sensitivity) tests or combination algorithms							
	PCR	HC2	mRNA	HPV16/18/45 OR [mRNA AND ≥ ASCUS]	≥ LSIL OR [PCR AND ASCUS]	≥ LSIL OR [HC2 AND ASCUS]	HPV16/18 OR [PCR AND ≥ ASCUS]	≥ ASCUS
Index (for sensitivity) and comparator (for specificity) tests or combination algorithms								
PCR	–	0.94 (0.87 to 1.02) [0.06]	0.93 (0.73 to 1.04) [0.10]	0.90 (0.49 to 1.05) [0.12]	0.89 (0.72 to 1.01) [0.04]	0.75 (0.58 to 0.89) [<0.001]	0.74 (0.56 to 0.89) [<0.001]	0.65 (0.54 to 0.75) [<0.001]
HC2	1.04 (0.99 to 1.11) [0.08]	–	0.99 (0.79 to 1.10) [0.42]	0.96 (0.52 to 1.11) [0.35]	0.95 (0.76 to 1.08) [0.26]	0.79 (0.63 to 0.93) [0.001]	0.79 (0.59 to 0.95) [0.003]	0.69 (0.58 to 0.79) [<0.001]
mRNA	1.05 (0.99 to 1.13) [0.05]	1.02 (0.96 to 1.06) [0.26]	–	0.97 (0.53 to 1.24) [0.41]	0.97 (0.77 to 1.24) [0.37]	0.81 (0.62 to 1.05) [0.049]	0.81 (0.59 to 1.05) [0.049]	0.71 (0.58 to 0.89) [0.005]
HPV16/18/45 OR [mRNA AND ≥ ASCUS]	1.07 (1.00 to 1.16) [0.02]	1.04 (0.97 to 1.08) [0.10]	1.02 (0.96 to 1.08) [0.23]	–	1.00 (0.77 to 1.82) [0.49]	0.84 (0.62 to 1.54) [0.19]	0.84 (0.60 to 1.50) [0.19]	0.73 (0.58 to 1.33) [0.09]
≥ LSIL OR [PCR AND ASCUS]	1.04 (0.97 to 1.12) [0.10]	1.01 (0.94 to 1.05) [0.39]	0.99 (0.92 to 1.06) [0.40]	0.97 (0.91 to 1.04) [0.18]	–	0.84 (0.64 to 1.08) [0.08]	0.84 (0.65 to 0.97) [<0.001]	0.73 (0.59 to 0.92) [0.004]
≥ LSIL OR [HC2 AND ASCUS]	1.10 (1.05 to 1.18) [<0.001]	1.06 (1.03 to 1.09) [<0.001]	1.04 (1.00 to 1.11) [0.02]	1.02 (0.98 to 1.09) [0.16]	1.05 (1.01 to 1.13) [0.007]	–	1.00 (0.72 to 1.33) [0.49]	0.88 (0.71 to 1.10) [0.11]
HPV16/18 OR [PCR AND ≥ ASCUS]	1.09 (1.04 to 1.16) [<0.001]	1.05 (1.02 to 1.09) [0.004]	1.03 (0.99 to 1.09) [0.06]	1.01 (0.97 to 1.08) [0.29]	1.04 (1.01 to 1.11) [<0.001]	0.99 (0.96 to 1.02) [0.25]	–	0.88 (0.71 to 1.16) [0.16]
≥ ASCUS	1.08 (1.03 to 1.15) [<0.001]	1.04 (1.02 to 1.07) [<0.001]	1.02 (0.99 to 1.08) [0.10]	1.00 (0.97 to 1.07) [0.45]	1.03 (0.99 to 1.10) [0.06]	0.98 (0.96 to 1.00) [0.04]	0.99 (0.97 to 1.02) [0.23]	–

Table 3. Comparative accuracy of guideline-proposed combination algorithms. Above the diagonal line (formed by cells with an en dash) represents relative sensitivity (95% CrI) [probability that relative sensitivity is ≥ 1] and below the diagonal line represents relative specificity (95% CrI) [probability that relative specificity is ≤ 1]. For relative sensitivity, the rows and columns, respectively, represent the index (the test of interest) and comparator (the test in comparison) tests or combination algorithms. For relative specificity, the columns and rows, respectively, represent the index and comparator tests or combination algorithms. *ASCUS* atypical squamous cells of undetermined significance, *CrI* credible interval, *HC2* Hybrid Capture 2, *HPV16/18(/45)* genotyping for HPV types 16 or 18 (or 45), *HSIL* high-grade squamous intraepithelial lesion, *LBC* liquid-based cytology, *LSIL* low-grade squamous intraepithelial lesion, *mRNA* messenger ribonucleic acid, *PCR* polymerase chain reaction.

was 1.06 [CrI: 0.98–1.15; Bayesian $P(\leq 1) = 0.06$] and relative specificity of HC2 vs. PCR was 1.04 [CrI: 0.99–1.11; Bayesian $P(\leq 1) = 0.08$].

Compared with standalone cytological testing irrespective of the thresholds, all standalone hrHPV assays other than HPV16/18 were more sensitive but less specific in general (Fig. 3a, Supplementary Table S7). In contrast, the accuracy of HPV16/18 was comparable to cytological testing. For example, the relative specificity for comparing \geq LSIL with HPV16/18 was 1.0 [CrI: 0.68–1.62; Bayesian $P(\geq 1) = 0.50$] and relative specificity was 1.01 (CrI: 1.00–1.03; Bayesian $P(\leq 1) = 0.10$).

Comparative accuracy among combination algorithms based on specific hrHPV assays. The ROC plots of the average accuracy estimates and their credible regions reflected the effect of altering the thresholds in combined cytological testing (i.e., lower thresholds with increased sensitivity and decreased specificity, and higher thresholds with increased specificity and decreased sensitivity) and the effect of combination methods (i.e., the OR rule with increased sensitivity and decreased specificity, and the AND rule with increased specificity and decreased sensitivity) across the subgroups based on alternative hrHPV assays (Fig. 3b–d). Among 45 pairwise comparisons based on cytology, HC2, and their combinations, most (40 [89%] for sensitivity and 42 [93%] for specificity) showed a significant difference, reflecting the effect of the thresholds and combination methods (Fig. 3b, Supplementary Table S8). Similarly, among 36 pairwise comparisons based on cytology, PCR-based tests, and their combinations, 28 (78%) for sensitivity and 27 (75%) for specificity showed a significant difference (Fig. 3c, Supplementary Table S9). In contrast, 10 pairwise comparisons based on mRNA-based combinations (Fig. 3d, Supplementary Table S10), only five (50%) and four (40%) contrasts for sensitivity and specificity, respectively, were significantly different.

Comparative accuracy and GRADE assessment of guideline-proposed combination algorithms. Data on the guideline-proposed algorithms are available for HC2 and PCR-based tests on “ \geq LSIL OR [hrHPV AND ASCUS]” and for mRNA-based tests and PCR-based tests on “HPV16/18(/45) OR [hrHPV AND \geq ASCUS]”. Table 3 summarizes the comparative accuracy, and Supplementary Table S11 and Table 4 show the GRADE summary of findings on specific tests or combination algorithms and their comparisons, respectively.

In general, the proposed algorithms were less sensitive but more specific than the standalone component hrHPV assays. However, only HC2-based “ \geq LSIL OR [hrHPV AND ASCUS]” and PCR-based “HPV16/18 OR [hrHPV AND \geq ASCUS]” were significantly less sensitive (the average relative sensitivity ranged from 0.74 to 0.79; Bayesian $P(\geq 1)$ ranged from <0.001 to 0.003) and more specific (the average relative specificity ranged from

Index and comparator tests or combination algorithms		Index (for FN) and comparator (for FP) tests or combination algorithms							
		PCR	HC2	mRNA	HPV16/18/45 OR [mRNA AND ≥ASCUS]	≥LSIL OR [PCR AND ASCUS]	≥LSIL OR [HC2 AND ASCUS]	HPV16/18 OR [PCR AND ≥ASCUS]	≥ASCUS
Index (for FP) and comparator (for FN) tests or combination algorithms	PCR	—	+1 (−0 to +2) Very low ⊕⊖⊖⊖⊖	+1 (−1 to +5) Very low ⊕⊖⊖⊖⊖	+2 (−1 to +10) Very low ⊕⊖⊖⊖⊖	+2 (−0 to +5) Very low ⊕⊖⊖⊖⊖	+5 (+2 to +8) Very low ⊕⊖⊖⊖⊖	+5 (+2 to +8) Low ⊕⊖⊖⊖⊖	+7 (+5 to +9) Moderate ⊕⊖⊖⊖⊖
	HC2	−32 (−88 to +11) Very low ⊕⊖⊖⊖⊖	—	+0 (−2 to +4) Very low ⊕⊖⊖⊖⊖	+1 (−2 to +8) Very low ⊕⊖⊖⊖⊖	+1 (−1 to +4) Very low ⊕⊖⊖⊖⊖	+4 (+1 to +7) Low ⊕⊖⊖⊖⊖	+4 (+1 to +7) Very low ⊕⊖⊖⊖⊖	+5 (+4 to +7) Moderate ⊕⊖⊖⊖⊖
	mRNA	−45 (−105 to +11) Very low ⊕⊖⊖⊖⊖	−14 (−51 to +35) Very low ⊕⊖⊖⊖⊖	—	+1 (−4 to +8) Very low ⊕⊖⊖⊖⊖	+1 (−3 to +4) Very low ⊕⊖⊖⊖⊖	+3 (−1 to +7) Very low ⊕⊖⊖⊖⊖	+3 (−1 to +7) Very low ⊕⊖⊖⊖⊖	+5 (+2 to +8) Moderate ⊕⊖⊖⊖⊖
	HPV16/18/45 OR [mRNA AND ≥ASCUS]	−64 (−126 to −3) Very low ⊕⊖⊖⊖⊖	−33 (−73 to +23) Very low ⊕⊖⊖⊖⊖	−19 (−73 to +39) Very low ⊕⊖⊖⊖⊖	—	+0 (−8 to +4) Very low ⊕⊖⊖⊖⊖	+3 (−5 to +7) Very low ⊕⊖⊖⊖⊖	+3 (−5 to +7) Very low ⊕⊖⊖⊖⊖	+5 (−3 to +8) Very low ⊕⊖⊖⊖⊖
	≥LSIL OR [PCR AND ASCUS]	−38 (−98 to +22) Very low ⊕⊖⊖⊖⊖	−7 (−48 to +51) Very low ⊕⊖⊖⊖⊖	+7 (−49 to +69) Very low ⊕⊖⊖⊖⊖	+26 (−35 to +88) Very low ⊕⊖⊖⊖⊖	—	+3 (−1 to +6) Very low ⊕⊖⊖⊖⊖	+3 (+1 to +6) Very low ⊕⊖⊖⊖⊖	+4 (+1 to +7) Very low ⊕⊖⊖⊖⊖
	≥LSIL OR [HC2 AND ASCUS]	−84 (−144 to −41) Very low ⊕⊖⊖⊖⊖	−52 (−80 to −30) Very low ⊕⊖⊖⊖⊖	−38 (−91 to −4) Very low ⊕⊖⊖⊖⊖	−19 (−78 to +16) Very low ⊕⊖⊖⊖⊖	−46 (−105 to −7) Very low ⊕⊖⊖⊖⊖	—	+0 (−4 to +4) Very low ⊕⊖⊖⊖⊖	+2 (−1 to +5) Very low ⊕⊖⊖⊖⊖
	HPV16/18 OR [PCR AND ≥ASCUS]	−76 (−131 to −35) Low ⊕⊖⊖⊖⊖	−44 (−75 to −14) Very low ⊕⊖⊖⊖⊖	−30 (−81 to +6) Very low ⊕⊖⊖⊖⊖	−11 (−67 to +25) Very low ⊕⊖⊖⊖⊖	−37 (−92 to −6) Very low ⊕⊖⊖⊖⊖	+8 (−16 to +38) Very low ⊕⊖⊖⊖⊖	—	+2 (−2 to +5) Very low ⊕⊖⊖⊖⊖
≥ASCUS	−68 (−123 to −30) Moderate ⊕⊖⊖⊖⊖	−36 (−59 to −18) Moderate ⊕⊖⊖⊖⊖	−22 (−71 to +9) Low ⊕⊖⊖⊖⊖	−3 (−59 to +29) Very low ⊕⊖⊖⊖⊖	−29 (−86 to +6) Very low ⊕⊖⊖⊖⊖	+17 (−2 to +34) Very low ⊕⊖⊖⊖⊖	+8 (−17 to +28) Very low ⊕⊖⊖⊖⊖	—	

Table 4. The GRADE summary of findings table for comparative evidence. Above the diagonal line (formed by cells with an em dash) represents the number of the difference in (Δ) FNs (95% CrI) and below the diagonal line represents Δ FPs (95% CrI). For Δ FPs, the rows and columns, respectively, represent the index (the test of interest) and comparator (the test in comparison) tests or combination algorithms. For Δ FNs, the columns and rows, respectively, represent the index and comparator tests or combination algorithms. Results are based on a healthy screening population of 1000 women in which 20 are CIN2+ (2%). ASCUS atypical squamous cells of undetermined significance, CIN2+, CrI cervical intraepithelial neoplasia grade 2 or higher grades; 95% credible interval, FN false negative, FP false positive, GRADE Grading of Recommendations Assessment, Development and Evaluation, HC2 Hybrid Capture 2, HPV16/18(45) genotyping for HPV types 16 or 18 (or 45), HSIL high-grade squamous intraepithelial lesion, LBC liquid-based cytology, LSIL low-grade squamous intraepithelial lesion, mRNA messenger ribonucleic acid, PCR polymerase chain reaction, TN true negative, TP true positive.

1.04 to 1.10; Bayesian $P(\leq 1)$ ranged from <0.001 to 0.004). These results suggested that the proposed algorithms, compared with their standalone component hrHPV tests, decreased by an average of 44 to 88 FPs but increased 4 to 5 more FNs (very low to low certainty of evidence).

In contrast, the proposed algorithms were in general equally specific but more sensitive than standalone \geq ASCUS. However, only PCR-based “LSIL OR [hrHPV AND ASCUS]” was significantly less sensitive than \geq ASCUS alone (the relative sensitivity = 0.73 [CrI: 0.59–0.92; Bayesian $P(\geq 1) = 0.004$]; four more FNs [CrI: 1–7]; very low certainty of evidence), but evidence as to whether this combination was more specific or less specific than \geq ASCUS alone was insufficient (relative sensitivity = 0.98 [CrI: 0.96–1.00; Bayesian $P(\geq 1) = 0.04$]).

Comparative evidence across alternative guideline-proposed algorithms was generally limited. PCR-based “LSIL OR [hrHPV AND ASCUS]” was significantly more specific and less specific than “HPV16/18 OR [hrHPV AND \geq ASCUS]” (relative sensitivity; 1.04 [CrI: 1.01–1.11]; Bayesian $P(\leq 1) < 0.001$); 37 fewer FPs [CrI: 6–92] and relative specificity: 0.84 [CrI: 0.65–0.97]; Bayesian $P(\geq 1) < 0.001$; three more FNs [CrI: 1–6]; very low certainty of evidence). Although only HC2-based “LSIL OR [hrHPV AND ASCUS]” was more specific than PCR-based “LSIL OR [hrHPV AND ASCUS]” (relative specificity: 1.05 [CrI: 1.01–1.13]; Bayesian $P(\leq 1) = 0.007$; 46 fewer FPs [CrI: 7–105]; very low certainty of evidence) across-hrHPV assays, comparative data on the guideline-proposed algorithms were insufficient.

Meta-regression and sensitivity analyses. Due to data paucity, meta-regression was undertaken for only HC2, cytological testing, and their OR combination separately. Although high-income countries (vs. non-high-income countries) for sensitivity of HC2 and sample collection by physicians (vs. nonphysician collectors) for sensitivity and specificity of \geq ASCUS were associated with higher estimates, these covariates were no longer associated with higher (or lower) sensitivity or specificity in their combination, HC2 OR \geq ASCUS (Supplementary Fig. S8).

The sensitivity analysis using the model with a common correlation parameter across tests yielded results comparable to those of the main analysis based on the model with test-specific correlation parameters (Supplementary Table S12). Relaxing threshold constraints yielded results not compliant with the expected threshold effects in two specific thresholds for cytological testing (\geq LSIL and \geq ASCH) and unstable results with wide CrIs

for sensitivity in four combination algorithms (i.e., mRNA AND \geq ASCUS, HPV16/18 AND \geq ASCUS, HPV16/18 OR \geq ASCUS, “ \geq LSIL OR [PCR AND ASCUS]”, and “ \geq HSIL OR [HC2 AND \geq ASCUS]”) regardless of whether correlation parameters were separately assumed or not; all of these tests, except for \geq LSIL, depended on only a few primary studies. With lower deviance information criterion estimates, the models with threshold constraints were deemed to be better-fitting than the models without threshold constraints; however, the differences were < 5 , suggesting no definitively preferred model.

Discussion

To the best of our knowledge, this is the first network meta-analysis that has comprehensively compared and ranked the cross-sectional screening accuracy of standalone cytology or hrHPV testing with combination algorithms for detecting CIN2+. Importantly, this analysis is based on published accuracy estimates from fully paired-design comparative accuracy studies that addressed verification bias. First, our network meta-analysis confirmed and quantified the theoretically expected gain in and trade-off of screening performance when combining two tests³⁴, that is, the combinations with the OR rule (i.e., either test positive) of hrHPV and cytological testing were most sensitive and least specific, whereas combinations with the AND rule (i.e., both test positive) of hrHPV and cytological testing were most specific and least sensitive. Second, our network meta-analysis confirmed that the guideline-proposed combination algorithms, HC2-based “ \geq LSIL OR [hrHPV AND ASCUS]” and PCR-based “HPV16/18 OR [hrHPV AND \geq ASCUS]” appeared to compensate the shortcomings of the two component tests if used as standalone, which, though expected theoretically, had never been quantitatively synthesized. Specifically, these proposed algorithms were not as sensitive but more specific than the component standalone hrHPV testing. Similarly, these proposed algorithms appeared equally specific but more sensitive than standalone \geq ASCUS, though definitive conclusions could not be made due to limited comparative data. Third, sparse, insufficient comparative evidence precluded reliable assessment of the comparative accuracy across these alternative guideline-proposed algorithms.

Effectiveness of screening should be assessed as a whole program consisting of a set of activities⁷¹. Since the ultimate goal is to maximize participant-relevant benefits and simultaneously minimize harms, accuracy of testing is, though an important measure, only an intermediate parameter. As already elucidated in the previous meta-analyses^{13,14}, which is congruent with our results, standalone testing for hrHPV using an assay other than HPV 16/18 genotyping, if all screen-positive women underwent colposcopy, would identify more women with CIN2+ than cytological testing alone but at the cost of more healthy women misclassified as CIN2+. The OR rule combinations, the most sensitive group of strategies found in our meta-analysis, if used for primary co-testing (i.e., performing both tests concurrently), would further increase the number of healthy women misclassified as CIN2+ while identifying only a few more women with CIN2+. The consequences of such FP results include unnecessary colposcopy, triage, or repeat testing with cytology, hrHPV, or other tests. Although infections with hrHPV, and HPV16/18 in particular, carry a higher risk of progression than positive cytology^{72–75}, immediate incremental costs and psychological burden incurred due to increased false-positive results may not be justified in low risk screening settings as only a fraction of the identified CIN2+ lesions detected through standalone hrHPV testing or its combinations progress to invasive cancer; the others actually carry a moderate chance of regression⁷⁶. The AND rule combinations, the most specific group of strategies identified in our meta-analysis, may substantially minimize FPs and their negative consequences. However, sensitivity is lower than cytology alone (\geq ASCUS), potentially leading to unignorable large numbers of FNs depending on the prevalence of CIN2+ in a screened population.

As interim recommendations, several protocols for triage and/or repeat testing followed by colposcopy for screen-positive women have been proposed by professional societies. “ \geq LSIL OR [hrHPV AND ASCUS]” and “HPV16/18 OR [hrHPV AND \geq ASCUS]” were cross-sectional representations for two such protocols, respectively, proposed for positive primary cytological testing¹¹ and primary hrHPV testing⁹. Our meta-analysis found that the accuracy of these combination algorithms were generally ranked in the middle, being more sensitive and less specific than standalone cytology (\geq ASCUS) and the AND rule combinations but more specific and less sensitive than standalone hrHPV testing and the OR rule combination. We also quantified how each combination algorithm increased or decreased the number of FNs and FPs relative to those of another specific standalone test or combination, which is a strength of our study results. However, any benefits and harms associated with specific screening tests or combinations should be formally assessed at the whole program level along with its necessary resources and costs⁷¹.

We focused on cross-sectional accuracy of initial screening tests or combinations and their immediate consequences. Our accuracy-based arguments necessarily lack long-term outcomes. Given the chance of regression⁷⁶, the results based on our cross-sectional approach may be only relevant in populations with a low participation rate of follow-up testing. Additionally, the positive criteria we adopted for the estimation of accuracy do not necessarily represent the optimal indications of colposcopy in real-life practice; rather the criteria included the joint indications of any additional intervention; i.e., triage and/or repeat testing, colposcopy, and immediate direct treatments jointly. In this regard, a recent expert consensus statement proposed individualized risk-based management decisions based on the combinations of the available screening results⁷⁷.

Colposcopy-directed biopsy is an imperfect test even for routine biopsies on normal-appearing sites⁷⁸ and more so for colposcopy and selective biopsy⁷⁹. Despite the theoretical superiority of verification bias-corrected accuracy estimates over naïvely calculated estimates, these corrections are not error-free. Given the complex mechanisms of missing verification⁸⁰ and limitations in inverse probability weighting⁸¹, bias may not necessarily have been corrected in the right direction. In addition, the effect of the excluded observations due to unsatisfactory or missing test results, even though the reported proportions were not substantial, could be unpredictably large. Furthermore, our meta-analysis was based on aggregate data and thus only accounted for the dependence

of the two tests at the aggregate data level⁸²; however, a more sophisticated approach to address these limitations would require individual-level data.

Our GRADE assessment used a typical population-based screening context in high-income countries as adopted in a previous review¹³; however, the large spread of the credible and predictive accuracy values in our study suggests wide-ranging real-life variations, implying that specific scenarios with different risks might yield divergent conclusions. Finally, we did not assess combinations involving newer screening modalities, such as p16/Ki-67 dual-stain-based cytology⁸³, as this was beyond the scope of our meta-analysis.

Conclusions

Limited evidence suggests that specific test combinations might complement the weaknesses of standalone cytological or hrHPV screening and help reduce FN and/or FP results. However, the strategies that provide more benefits than harms at reasonable cost in a population need to be assessed at the program level. As comparative evidence on alternative hrHPV assays is sparse, further research is needed to acquire relevant data. Additionally, future research should elucidate long-term outcomes of specific algorithms and acquire data from HPV-vaccinated populations.

Data availability

The data and statistical codes that supports the findings of this study will be shared on reasonable request to the corresponding author.

Received: 21 July 2021; Accepted: 17 December 2021

Published online: 07 January 2022

References

- Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424. <https://doi.org/10.3322/caac.21492> (2018).
- WHO. *Human Papillomavirus (HPV) and Cervical Cancer* (WHO, 2019).
- Peirson, L., Fitzpatrick-Lewis, D., Ciliska, D. & Warren, R. Screening for cervical cancer: A systematic review and meta-analysis. *Syst. Rev.* **2**, 35. <https://doi.org/10.1186/2046-4053-2-35> (2013).
- Crosbie, E. J., Einstein, M. H., Franceschi, S. & Kitchener, H. C. Human papillomavirus and cervical cancer. *Lancet* **382**, 889–899. [https://doi.org/10.1016/s0140-6736\(13\)60022-7](https://doi.org/10.1016/s0140-6736(13)60022-7) (2013).
- Ronco, G. *et al.* Efficacy of HPV-based screening for prevention of invasive cervical cancer: Follow-up of four European randomised controlled trials. *Lancet* **383**, 524–532. [https://doi.org/10.1016/s0140-6736\(13\)62218-7](https://doi.org/10.1016/s0140-6736(13)62218-7) (2014).
- Curry, S. J. *et al.* Screening for cervical cancer: US Preventive Services Task Force Recommendation Statement. *JAMA* **320**, 674–686. <https://doi.org/10.1001/jama.2018.10897> (2018).
- Jeronimo, J., Castle, P. E., Temin, S. & Shastri, S. S. Secondary Prevention of Cervical Cancer: American Society of Clinical Oncology Resource-stratified clinical practice guideline summary. *J. Oncol. Pract.* **13**, 129–133. <https://doi.org/10.1200/jop.2016.017889> (2017).
- Sawaya, G. F., Kulasingam, S., Denberg, T. D. & Qaseem, A. Cervical cancer screening in average-risk women: Best practice advice from the Clinical Guidelines Committee of the American College of Physicians. *Ann. Intern. Med.* **162**, 851–859. <https://doi.org/10.7326/m14-2426> (2015).
- Huh, W. K. *et al.* Use of primary high-risk human papillomavirus testing for cervical cancer screening: Interim clinical guidance. *Obstet. Gynecol.* **125**, 330–337. <https://doi.org/10.1097/aog.0000000000000669> (2015).
- Fontham, E. T. H. *et al.* Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. *CA Cancer J. Clin.* **70**, 321–346. <https://doi.org/10.3322/caac.21628> (2020).
- Saslow, D. *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.* **62**, 147–172. <https://doi.org/10.3322/caac.21139> (2012).
- Dahabreh, I. J., Trikalinos, T. A., Balk, E. M. & Wong, J. B. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (Agency for Healthcare Research and Quality (US), 2016).
- Koliopoulos, G. *et al.* Cytology versus HPV testing for cervical cancer screening in the general population. *Cochrane Database. Syst. Rev.* **8**, CD008587. <https://doi.org/10.1002/14651858.CD008587.pub2> (2017).
- Mustafa, R. A. *et al.* Systematic reviews and meta-analyses of the accuracy of HPV tests, visual inspection with acetic acid, cytology, and colposcopy. *Int. J. Gynaecol. Obstet.* **132**, 259–265. <https://doi.org/10.1016/j.ijgo.2015.07.024> (2016).
- Fokom-Domgue, J. *et al.* Performance of alternative strategies for primary cervical cancer screening in sub-Saharan Africa: Systematic review and meta-analysis of diagnostic test accuracy studies. *BMJ* **351**, h3084. <https://doi.org/10.1136/bmj.h3084> (2015).
- Li, T. *et al.* Diagnostic value of combination of HPV testing and cytology as compared to isolated cytology in screening cervical cancer: A meta-analysis. *J. Cancer Res. Ther.* **12**, 283–289. <https://doi.org/10.4103/0973-1482.154032> (2016).
- Biondi-Zoccai, G. (ed.) *Diagnostic Meta-analysis: A Useful Tool for Clinical Decision-Making* 183–197 (Springer, 2018).
- Hamashima, C. *et al.* The Japanese guideline for cervical cancer screening. *Jpn. J. Clin. Oncol.* **40**, 485–502. <https://doi.org/10.1093/jjco/hyq036> (2010).
- The Japanese Research Group for Systematic Review and Guideline Development for Cancer Screening. *An Evidence Report for the Japanese Guideline for Cervical Cancer Screening 2019*, 1–258 (2020). <http://canscreen.ncc.go.jp/guideline/shikyukeireport2019.pdf>. Accessed 20 Dec 2021.
- McInnes, M. D. F. *et al.* Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *JAMA* **319**, 388–396. <https://doi.org/10.1001/jama.2017.19163> (2018).
- Denton, K. J. *et al.* The revised BSCC terminology for abnormal cervical cytology. *Cytopathology* **19**, 137–157. <https://doi.org/10.1111/j.1365-2303.2008.00585.x> (2008).
- Cirkel, C., Barop, C. & Beyer, D. A. Method comparison between Munich II and III nomenclature for Pap smear samples. *J. Turk. Ger. Gynecol. Assoc.* **16**, 203–207. <https://doi.org/10.5152/jtgga.2015.0147> (2015).
- Schiffman, M. *et al.* Human papillomavirus testing in the prevention of cervical cancer. *J. Natl. Cancer Inst.* **103**, 368–383. <https://doi.org/10.1093/jnci/djq562> (2011).
- Macaskill, P., Walter, S. D., Irwig, L. & Franco, E. L. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Stat. Med.* **21**, 2527–2546. <https://doi.org/10.1002/sim.1227> (2002).
- Dickinson, J. *et al.* Recommendations on screening for cervical cancer. *CMAJ* **185**, 35–45. <https://doi.org/10.1503/cmaj.121505> (2013).

26. Yang, B. et al. Development of QUADAS-C, a risk of bias tool for comparative diagnostic accuracy studies. <https://doi.org/10.17605/OSF.IO/HQ8MF> (2021).
27. Whiting, P. F. et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann. Intern. Med.* **155**, 529–536. <https://doi.org/10.7326/0003-4819-155-8-201110180-00009> (2011).
28. Phillips, B., Stewart, L. A. & Sutton, A. J. “Cross hairs” plots for diagnostic meta-analysis. *Res. Synth. Methods* **1**, 308–315. <https://doi.org/10.1002/jrsm.26> (2010).
29. Owen, R. K., Cooper, N. J., Quinn, T. J., Lees, R. & Sutton, A. J. Network meta-analysis of diagnostic test accuracy studies identifies and ranks the optimal diagnostic tests and thresholds for health care policy and decision-making. *J. Clin. Epidemiol.* **99**, 64–74. <https://doi.org/10.1016/j.jclinepi.2018.03.005> (2018).
30. Harbord, R. M., Deeks, J. J., Egger, M., Whiting, P. & Sterne, J. A. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* **8**, 239–251. <https://doi.org/10.1093/biostatistics/kxl004> (2007).
31. Chu, H., Nie, L., Cole, S. R. & Poole, C. Meta-analysis of diagnostic accuracy studies accounting for disease prevalence: Alternative parameterizations and model selection. *Stat. Med.* **28**, 2384–2399. <https://doi.org/10.1002/sim.3627> (2009).
32. Arends, L. R. et al. Bivariate random effects meta-analysis of ROC curves. *Med. Decis. Making* **28**, 621–638. <https://doi.org/10.1177/0272989x08319957> (2008).
33. Human Development Indicators and Indices: 2018 Statistical Update Team. In *United Nations Development Programme; 2018* (United Nations Development Programme, 2018).
34. Schunemann, H. J. et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336**, 1106–1110. <https://doi.org/10.1136/bmj.39500.677199.AE> (2008).
35. Thompson, J. *Bayesian Analysis with Stata* (Stata Press, 2014).
36. Belinson, J. et al. Shanxi Province cervical cancer screening study: A cross-sectional comparative trial of multiple techniques to detect cervical neoplasia. *Gynecol. Oncol.* **83**, 439–444. <https://doi.org/10.1006/gyno.2001.6370> (2001).
37. Pan, Q. et al. A thin-layer, liquid-based pap test for mass screening in an area of China with a high incidence of cervical carcinoma. A cross-sectional, comparative study. *Acta. Cytol.* **47**, 45–50. <https://doi.org/10.1159/000326474> (2003).
38. Zhao, F. H. et al. A study of cervical cancer screening algorithms. *Zhonghua Zhong Liu Za Zhi* **32**, 420–424 (2010).
39. Cardenas-Turanzas, M. et al. The performance of human papillomavirus high-risk DNA testing in the screening and diagnostic settings. *Cancer Epidemiol. Biomark. Prev.* **17**, 2865–2871. <https://doi.org/10.1158/1055-9965.epi-08-0137> (2008).
40. Hovland, S. et al. A comprehensive evaluation of the accuracy of cervical pre-cancer detection methods in a high-risk area in East Congo. *Br. J. Cancer* **102**, 957–965. <https://doi.org/10.1038/sj.bjc.6605594> (2010).
41. Schneider, A. et al. Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. *Int. J. Cancer* **89**, 529–534 (2000).
42. Kulasingam, S. L. et al. Evaluation of human papillomavirus testing in primary screening for cervical abnormalities: Comparison of sensitivity, specificity, and frequency of referral. *JAMA* **288**, 1749–1757 (2002).
43. Balasubramanian, A. et al. Accuracy and cost-effectiveness of cervical cancer screening by high-risk human papillomavirus DNA testing of self-collected vaginal samples. *J. Low. Genit. Tract. Dis.* **14**, 185–195. <https://doi.org/10.1097/LGT.0b013e3181cd6d36> (2010).
44. Bigras, G. & de Marval, F. The probability for a Pap test to be abnormal is directly proportional to HPV viral load: Results from a Swiss study comparing HPV testing and liquid-based cytology to detect cervical cancer precursors in 13,842 women. *Br. J. Cancer* **93**, 575–581. <https://doi.org/10.1038/sj.bjc.6602728> (2005).
45. Mayrand, M. H. et al. Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N. Engl. J. Med.* **357**, 1579–1588. <https://doi.org/10.1056/NEJMoa071430> (2007).
46. Mayrand, M. H. et al. Randomized controlled trial of human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: Design, methods and preliminary accrual results of the Canadian cervical cancer screening trial (CCCaST). *Int. J. Cancer* **119**, 615–623. <https://doi.org/10.1002/ijc.21897> (2006).
47. Li, N. et al. Different cervical cancer screening approaches in a Chinese multicentre study. *Br. J. Cancer* **100**, 532–537. <https://doi.org/10.1038/sj.bjc.6604840> (2009).
48. Castle, P. E. et al. Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: A subanalysis of the ATHENA study. *Lancet Oncol.* **12**, 880–890. [https://doi.org/10.1016/s1470-2045\(11\)70188-7](https://doi.org/10.1016/s1470-2045(11)70188-7) (2011).
49. Stoler, M. H. et al. High-risk human papillomavirus testing in women with ASC-US cytology: Results from the ATHENA HPV study. *Am. J. Clin. Pathol.* **135**, 468–475. <https://doi.org/10.1309/ajcpz5jy6fcvnmot> (2011).
50. Mahmud, S. M. et al. Comparison of human papillomavirus testing and cytology for cervical cancer screening in a primary health care setting in the Democratic Republic of the Congo. *Gynecol. Oncol.* **124**, 286–291. <https://doi.org/10.1016/j.ygyno.2011.10.031> (2012).
51. Sangrajrang, S. et al. Comparative accuracy of Pap smear and HPV screening in Ubon Ratchathani in Thailand. *Papillomavirus Res.* **3**, 30–35. <https://doi.org/10.1016/j.pvr.2016.12.004> (2017).
52. Sangrajrang, S. et al. Human papillomavirus (HPV) DNA and mRNA primary cervical cancer screening: Evaluation and triaging options for HPV-positive women. *J. Med. Screen.* **26**, 212–218. <https://doi.org/10.1177/0969141319865922> (2019).
53. Kurokawa, T. et al. The ideal strategy for cervical cancer screening in Japan: Result from the Fukui Cervical cancer screening study. *Cytopathology* **29**, 361–367. <https://doi.org/10.1111/cyt.12576> (2018).
54. Blumenthal, P. D. et al. Adjunctive testing for cervical cancer in low resource settings with visual inspection, HPV, and the Pap smear. *Int. J. Gynaecol. Obstet.* **72**, 47–53 (2001).
55. Visual inspection with acetic acid for cervical-cancer screening: Test qualities in a primary-care setting. University of Zimbabwe/JHPIEGO Cervical Cancer Project. *Lancet* **353**, 869–873 (1999).
56. Coste, J. et al. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. *BMJ* **326**, 733. <https://doi.org/10.1136/bmj.326.7392.733> (2003).
57. de Cremoux, P. et al. Efficiency of the hybrid capture 2 HPV DNA test in cervical cancer screening. A study by the French Society of Clinical Cytology. *Am. J. Clin. Pathol.* **120**, 492–499. <https://doi.org/10.1309/xfuc-pp6m-5xua-94b8> (2003).
58. Sankaranarayanan, R. et al. Accuracy of human papillomavirus testing in primary screening of cervical neoplasia: Results from a multicenter study in India. *Int. J. Cancer* **112**, 341–347. <https://doi.org/10.1002/ijc.20396> (2004).
59. Qiao, Y. L. et al. A new HPV-DNA test for cervical-cancer screening in developing regions: A cross-sectional study of clinical accuracy in rural China. *Lancet Oncol.* **9**, 929–936. [https://doi.org/10.1016/s1470-2045\(08\)70210-9](https://doi.org/10.1016/s1470-2045(08)70210-9) (2008).
60. McAdam, M., Sakita, J., Tarivonda, L., Pang, J. & Frazer, I. H. Evaluation of a cervical cancer screening program based on HPV testing and LLETZ excision in a low resource setting. *PLoS ONE* **5**, e13266. <https://doi.org/10.1371/journal.pone.0013266> (2010).
61. Quincy, B. L., Turbow, D. J., Dabinett, L. N., Dillingham, R. & Monroe, S. Diagnostic accuracy of self-collected human papillomavirus specimens as a primary screen for cervical cancer. *J. Obstet. Gynaecol.* **32**, 795–799. <https://doi.org/10.3109/01443615.2012.717989> (2012).
62. Cuzick, J. et al. Management of women who test positive for high-risk types of human papillomavirus: The HART study. *Lancet* **362**, 1871–1876. [https://doi.org/10.1016/s0140-6736\(03\)14955-0](https://doi.org/10.1016/s0140-6736(03)14955-0) (2003).
63. Petry, K. U. 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* **88**, 1570–1577. <https://doi.org/10.1038/sj.bjc.6600918> (2003).

64. Gravitt, P. E. *et al.* Effectiveness of VIA, Pap, and HPV DNA testing in a cervical cancer screening program in a peri-urban community in Andhra Pradesh, India. *PLoS ONE* **5**, e13711. <https://doi.org/10.1371/journal.pone.0013711> (2010).
65. Moy, L. M. *et al.* Human papillomavirus testing and cervical cytology in primary screening for cervical cancer among women in rural China: Comparison of sensitivity, specificity, and frequency of referral. *Int. J. Cancer* **127**, 646–656. <https://doi.org/10.1002/ijc.25071> (2010).
66. Monsonog, J. *et al.* Evaluation of oncogenic human papillomavirus RNA and DNA tests with liquid-based cytology in primary cervical cancer screening: The FASE study. *Int. J. Cancer* **129**, 691–701. <https://doi.org/10.1002/ijc.25726> (2011).
67. Ferreccio, C. *et al.* Screening trial of human papillomavirus for early detection of cervical cancer in Santiago, Chile. *Int. J. Cancer* **132**, 916–923. <https://doi.org/10.1002/ijc.27662> (2013).
68. Agorastos, T. *et al.* Primary screening for cervical cancer based on high-risk human papillomavirus (HPV) detection and HPV 16 and HPV 18 genotyping, in comparison to cytology. *PLoS ONE* **10**, e0119755. <https://doi.org/10.1371/journal.pone.0119755> (2015).
69. Iftner, T. *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.* **53**, 2509–2516. <https://doi.org/10.1128/jcm.01013-15> (2015).
70. Wu, Q. *et al.* A cross-sectional study on HPV testing with type 16/18 genotyping for cervical cancer screening in 11,064 Chinese women. *Cancer Med.* **6**, 1091–1101. <https://doi.org/10.1002/cam4.1060> (2017).
71. Gray, J. A., Patnick, J. & Blanks, R. G. Maximising benefit and minimising harm of screening. *BMJ* **336**, 480–483. <https://doi.org/10.1136/bmj.39470.643218.94> (2008).
72. Dillner, J. *et al.* Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: Joint European cohort study. *BMJ* **337**, a1754. <https://doi.org/10.1136/bmj.a1754> (2008).
73. Katki, H. A. *et al.* Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: A population-based study in routine clinical practice. *Lancet Oncol.* **12**, 663–672. [https://doi.org/10.1016/s1470-2045\(11\)70145-0](https://doi.org/10.1016/s1470-2045(11)70145-0) (2011).
74. Kitchener, H. C. *et al.* A comparison of HPV DNA testing and liquid based cytology over three rounds of primary cervical screening: Extended follow up in the ARTISTIC trial. *Eur. J. Cancer* **47**, 864–871. <https://doi.org/10.1016/j.ejca.2011.01.008> (2011).
75. Luyten, A. *et al.* Early detection of CIN3 and cervical cancer during long-term follow-up using HPV/Pap smear co-testing and risk-adapted follow-up in a locally organised screening programme. *Int. J. Cancer* **135**, 1408–1416. <https://doi.org/10.1002/ijc.28783> (2014).
76. Tainio, K. *et al.* Clinical course of untreated cervical intraepithelial neoplasia grade 2 under active surveillance: Systematic review and meta-analysis. *BMJ* **360**, k499. <https://doi.org/10.1136/bmj.k499> (2018).
77. Perkins, R. B. *et al.* 2019 ASCCP risk-based management consensus guidelines for abnormal cervical cancer screening tests and cancer precursors. *J. Low Genit. Tract. Dis.* **24**, 102–131. <https://doi.org/10.1097/lgt.0000000000000525> (2020).
78. Wentzensen, N. *et al.* Multiple biopsies and detection of cervical cancer precursors at colposcopy. *J. Clin. Oncol.* **33**, 83–89. <https://doi.org/10.1200/jco.2014.55.9948> (2015).
79. Brown, B. H. & Tidy, J. A. The diagnostic accuracy of colposcopy—A review of research methodology and impact on the outcomes of quality assurance. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **240**, 182–186. <https://doi.org/10.1016/j.ejogrb.2019.07.003> (2019).
80. Naaktgeboren, C. A. *et al.* Anticipating missing reference standard data when planning diagnostic accuracy studies. *BMJ.* <https://doi.org/10.1136/bmj.i402> (2016).
81. Cronin, A. M. & Vickers, A. J. Statistical methods to correct for verification bias in diagnostic studies are inadequate when there are few false negatives: A simulation study. *BMC Med. Res. Methodol.* **8**, 75–75. <https://doi.org/10.1186/1471-2288-8-75> (2008).
82. Menten, J. & Lesaffre, E. A general framework for comparative Bayesian meta-analysis of diagnostic studies. *BMC Med. Res. Methodol.* **15**, 70. <https://doi.org/10.1186/s12874-015-0061-7> (2015).
83. Wentzensen, N. *et al.* Clinical evaluation of human papillomavirus screening with p16/Ki-67 dual stain triage in a large organized cervical cancer screening program. *JAMA Intern. Med.* **179**, 881–888. <https://doi.org/10.1001/jamainternmed.2019.0306> (2019).

Acknowledgements

We thank Dr. Alejandra Castanon (on behalf of Professor Thomas Iftner and Professor Peter Sasieni), Dr. Joel Coste, and Dr. Tetsuji Kurokawa for the provision of the additional information on their original work; and MARUZEN-YUSHODO Co., Ltd. (<https://kw.maruzen.co.jp/kousei-honyaku/>) for the English language editing.

Author contributions

T.T.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing—original draft, writing—review & editing. S.H.: conceptualization, investigation, methodology, validation, writing—review & editing. S.S.: conceptualization, investigation, validation, writing—review & editing. K.H.: conceptualization, investigation, validation, visualization, writing—review & editing. Y.H.: conceptualization, investigation, validation, writing—review & editing. T.K.: conceptualization, investigation, validation, writing—review & editing. C.H.: conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, writing—review & editing.

Funding

This work was supported by the National Cancer Center Research and Development Fund from the National Cancer Center, Tokyo, Japan (Grant Numbers 26-A-30, 29-A-16); and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (Grant Number 26460755 to TT and CH).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-04201-y>.

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022