Performance of alternative strategies for primary cervical cancer screening in sub-Saharan Africa: systematic review and meta-analysis of diagnostic test accuracy studies

Joël Fokom-Domgue,[1, 2] Christophe Combescure,[3] Victoire Fokom-Defo,[4] Pierre Marie Tebeu,[1] Pierre Vassilakos,[5] André Pascal Kengne,[6] Patrick Petignat[2]

[1]Department of Gynecology and Obstetrics, Faculty of Medicine and Biomedical Sciences, University of Yaoundé, Yaoundé, Cameroon

[2]Division of Gynecology, Department of Gynecology and Obstetrics, Geneva University Hospitals, 1211 Geneva 14, Switzerland

[3]Division of Clinical Epidemiology, Geneva University Hospitals, Geneva, Switzerland

[4]Division of Infectious and Chronic Diseases, Department of Internal Medicine, Yaoundé Central Hospital, Yaoundé, Cameroon

[5]Geneva Foundation for Medical Education and Research, Geneva, Switzerland

[6]South African Medical Research Council and University of Cape Town, Cape Town, South Africa

Correspondence to:
J Fokom-Domgue
fokom.domgue@gmail.com

Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmj.h3084)

ABSTRACT

OBJECTIVE

To assess and compare the accuracy of visual inspection with acetic acid (VIA), visual inspection with Lugol's iodine (VILI), and human papillomavirus (HPV) testing as alternative standalone methods for primary cervical cancer screening in sub-Saharan Africa.

DESIGN

Systematic review and meta-analysis of diagnostic test accuracy studies.

DATA SOURCES

Systematic searches of multiple databases including Medline, Embase, and Scopus for studies published between January 1994 and June 2014.

REVIEW METHODS

Inclusion criteria for studies were: alternative methods to cytology used as a standalone test for primary screening; study population not at particular risk of cervical cancer (excluding studies focusing on HIV positive women or women with gynaecological symptoms); women screened by nurses; reference test (colposcopy and directed biopsies) performed at least in women with positive screening results. Two reviewers independently screened studies for eligibility and extracted data for inclusion, and evaluated study quality using the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) checklist. Primary outcomes were absolute accuracy measures (sensitivity and specificity) of screening tests to detect cervical intraepithelial neoplasia grade 2 or worse (CIN2+).

RESULTS

15 studies of moderate quality were included (n=61 381 for VIA, n=46 435 for VILI, n=11 322 for HPV testing). Prevalence of CIN2+ did not vary by screening test and ranged from 2.3% (95% confidence interval 1.5% to 3.3%) in VILI studies to 4.9% (2.7% to 7.8%) in HPV testing studies. Positivity rates of VILI, VIA, and HPV testing were 16.5% (9.8% to 24.7%), 16.8% (11.0% to 23.6%), and 25.8% (17.4% to 35.3%), respectively. Pooled sensitivity was higher for VILI (95.1%; 90.1% to 97.7%) than VIA (82.4%; 76.3% to 87.3%) in studies where the reference test was performed in all women (P<0.001). Pooled specificity of VILI and VIA were similar (87.2% (78.1% to 92.8%) *v* 87.4% (77.1% to 93.4%); P=0.85). Pooled sensitivity and specificity were similar for HPV testing versus VIA (both P≥0.23) and versus VILI (both P≥0.16). Accuracy of VIA and VILI increased with sample size and time period.

CONCLUSIONS

For primary screening of cervical cancer in sub-Saharan Africa, VILI is a simple and affordable alternative to cytology that demonstrates higher sensitivity than VIA. Implementation studies are needed to assess the effect of these screening strategies on the incidence and outcomes of cervical cancer in the region.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Many trials on cervical cancer prevention have been implemented in developing countries

In sub-Saharan Africa, primary data from these studies have been conflicting, with some reports concluding that visual inspection with acetic acid (VIA) is not suited for cervical cancer screening in the region

Previous reviews have looked at the accuracy of alternative methods to cytology for cervical cancer screening in resource poor countries, but none has focused on sub-Saharan Africa

WHAT THIS STUDY ADDS

Our findings provide strong evidence that cervical visual inspection with Lugol's iodine (VILI) performs better than VIA for primary cervical cancer screening in sub-Saharan Africa

As with VIA, VILI should be recommended as a suitable test for primary cervical cancer screening in sub-Saharan Africa

Introduction

Cervical cancer is the leading cause of cancer related deaths among women in limited resource settings (about 230 000 deaths per year).[1] In sub-Saharan Africa, age standardised annual incidence and mortality rates for invasive cervical cancer are 56 and 41 per 100 000 women, respectively, which are the highest rates in the world.[2] Implementation of cytology based screening strategies has improved the control of cervical cancer in developed countries. Replicating these strategies in resource poor countries has been hampered by cost constraints, lack of infrastructures, insufficient health workforce, and lag time between sample collection and availability of test results.[3]

The development of vaccines against human papillomavirus (HPV), a key cofactor of cervical cancer, was a major breakthrough for the prevention of cervical cancer. However, large scale implementation of HPV vaccination is still challenging, and even when effective, it will not override the need for cervical cancer screening.[4] Thus, attention has been directed at new tests for cervical cancer screening that are appropriate for limited resource settings. Alternative methods to cytology have been developed, such as visual inspection of the cervix after application of acetic acid (VIA) or Lugol's

iodine (VILI) and molecular testing for high risk types of HPV (HPV testing). These innovative methods can be combined with immediate or slightly delayed management of precursor lesions, and prevention strategies using these techniques are potentially viable and cost effective for reducing the incidence of invasive cervical cancer in resource poor countries.[5-7] However, concerns about the effectiveness of screen and treat strategies in sub-Saharan Africa have recently emerged.[8] Furthermore, the use of these novel tests for cervical cancer screening is not universally accepted.

Over the past few decades, many trials on cervical cancer prevention have been implemented in developing countries by the World Health Organization and partner organisations. In sub-Saharan Africa, most of these projects aimed to evaluate the feasibility and effectiveness of alternative methods for cervical cancer screening.[9] Primary data from these studies have been conflicting, with some reports concluding that VIA was not suited for cervical cancer screening in sub-Saharan African countries.[10 11] Previous reviews have looked at the accuracy of alternative screening methods for cervical cancer in resource poor countries, but none has focused on those in sub-Saharan Africa.[12-15] Furthermore, comprehensive assessment of the quality of accuracy studies on cervical cancer screening conducted in the region has yet to be performed. To evaluate the reliability of results obtained from concluded or ongoing projects and to facilitate their implementation, we conducted a systematic review and meta-analysis to assess the accuracy of VIA, VILI, and HPV testing as standalone tests for primary screening for cervical cancer in sub-Saharan Africa.

With this systematic review and meta-analysis, we aimed to answer three questions in relation to VIA, VILI, and HPV testing:

- What is the validity (risk of bias and applicability) of accuracy studies on the three techniques when applied independently for primary cervical cancer screening in sub-Saharan Africa?
- What is the prevalence of cervical intraepithelial neoplasia grade 2 or worse (CIN2+) and positivity rate of these screening methods in sub-Saharan Africa?
- What is the overall performance (sensitivity, specificity, positive predictive value, and negative predictive value) of the screening methods to detect high grade cervical lesions, and how do they compare with each other in sub-Saharan Africa?

Methods
Data sources and studies selection
We did a computerised literature search for studies and technical reports published between January 1994 and June 2014, on Medline (via PubMed), Embase, Google Scholar, Scopus, the Cochrane Library, OCLC (Paper First and Proceedings First), PAIS International Database (EBSCO), WHO Global Health Library, and POPLINE. The concept of "alternative methods" for cervical cancer screening was introduced in 1994. Web appendix 1 shows the PICOS components (population,

intervention, comparison, outcomes, and studies) of this review. We used a combination of medical subject heading terms and free text terms relating to "cervical cancer," "screening," and "sub-Saharan Africa," limiting to studies in human beings (web appendix 2). We also searched the databases of Web of Knowledge and scanned the references lists of relevant reviews. We made no restriction by publication language or study design. The last search date was 11 July 2014.

Studies identified were sequentially screened by title, abstract, and full text to retain those in the final review. Included studies met the following criteria:

- Alternative methods of cervical cancer screening (VIA, VILI, or HPV testing) were used as standalone tests for primary screening
- The study population was not at particular risk of cervical cancer (studies focusing on HIV positive women or on women presenting with gynecological symptoms were excluded)
- The screening procedure was performed by paramedical healthcare providers (excluding physicians)
- The reference test was either random biopsies or colposcopy guided biopsies, loop excision, or endocervical curettage performed at least in screen positive women.

For HPV testing, studies were eligible if sampling was done by the provider (that is, excluding self sampling) and if the technique for HPV detection was either polymerase chain reaction (PCR) or second generation hybrid capture because the first generation assay is no longer available. Authors of relevant studies were contacted for missing data. Web appendix 3 lists the excluded studies after review of the abstract or methods of full texts and the reasons for exclusion.

Definition of positive screening tests
Visual methods (VIA and VILI) involved examination of the cervix with the naked eye (without magnification) using a focus lamp or a torch light. In the most recent studies, VIA and VILI were categorised as positive or negative according to the criteria of the International Agency for Research on Cancer (IARC).[16] But in earlier studies, the mere presence of an acetowhite lesion or growth on the cervix was considered positive for VIA. A positive result to VIA from IARC criteria was defined as the presence of well defined acetowhite areas in the transformation zone (near the squamocolumnar junction or external os) or the presence of an acetowhite growth, about one minute after direct application of a 3-5% diluted solution of acetic acid using a cotton swab or a spray. A positive result to VILI was based on the appearance of a definite yellow area in the transformation zone, close to the squamocolumnar junction or the external os or on a cervical growth, after application of a 5% solution of Lugol's iodine.

For HPV testing, most studies carried out in sub-Saharan Africa used the second generation hybrid capture system, or a spin-off system known as careHPV. Results from either system were scored as positive if the ratio of specimen relative light units (RLUs) was equal to or

greater than the reference RLU (1 pg/mL). The other method for HPV detection included in this review was the PCR technique for which we kept the definition of positive results as defined by the tests manufacturers.

Definition of disease and gold standard

Disease was defined as the presence of histologically confirmed CIN2+ or equivalent according to the WHO classification.[17] The reference standard test (or gold standard) that we initially selected for this review was either a combination of colposcopy and histology, or histology without colposcopy. However, no study using histology without colposcopy as a gold standard fulfilled our inclusion criteria. In most studies, biopsy or loop excision was performed if a lesion was suspected at colposcopy, and endocervical curettage, random biopsies, or no histological sampling was done if there was no visible lesion or squamocolumnar junction.

Management of screened women

The management of women across selected studies when reported was based on the results of screening and reference tests. When colposcopic findings or biopsy results showed the presence of high grade cervical dysplasia, women generally received cryotherapy (in recent studies), cauterisation (in earlier studies), or conisation (loop electrosurgical excision procedure or cold knife conisation) as indicated. If the histological interpretation was suggestive of invasive cervical cancer, patients were referred to a specialised centre for disease assessment and appropriate treatment (hysterectomy, radiotherapy, or chemotherapy when necessary). Women who screened negative and those presenting minor grade cervical lesions on colposcopy or biopsy were advised to repeat screening within the next three years.

Patient involvement

No patients were involved in setting the research question or the outcome measures, nor were they involved in the design and implementation of the study. There are no plans to involve patients in dissemination.

Data abstraction and quality assessment

Our protocol followed the PRISMA guidelines[18] (web appendix 4). Two reviewers independently screened studies for eligibility and extracted data for inclusion in this systematic review. We extracted the following data: first author's name, year of publication, study period, country, geographical area (middle Africa, eastern Africa, western Africa, and southern Africa; web figure 1, web appendix 5), study design, study population, sample size, age range, screening test(s) assessed, qualification of screeners, place of screening (field clinic, primary and tertiary levels), definition of positive screening results, type of screening test (for HPV testing), gold standard or reference test, disease thresholds, and recipients of the reference test.

If many study samples from different countries were included in a single report and data were provided separately, we treated each sample as an individual study. For each screening test per study, we extracted data for the number of true positives, false negatives, false positives, and true negatives as defined by the threshold of CIN2+. Studies were classified into two groups, depending on whether the gold standard was performed in all women (GSA) or only in a portion of the study population that included all women with positive screening results (GSP). For the second group, we assumed that women who screened negative and did not undergo reference standard were disease free.[19] Two investigators separately evaluated the quality of eligible studies with the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) checklist.[20] Disagreements were solved by consensus.

Data synthesis

We pooled prevalence of disease (CIN2+) by screening test (VIA, VILI, or HPV testing), geographical region, study group (GSA or GSP), and positivity rate (proportion of screen positive women in the study population) of each screening test by study group using random effects meta-analytical methods for proportions with arcsine transformation.[21] Heterogeneity across studies was assessed and quantified with the Cochran Q test and I^2 statistic.[22] Absolute accuracy measures (sensitivity, specificity) were pooled by screening group in a bivariate random effects model, allowing for inter-setting variability.[23] We jointly illustrated the absolute pooled sensitivity and specificity for CIN2+ using hierarchical summary receiver operating characteristic (HSROC) regression curves.[24 25] In the GSA group, we used meta-regression approach within the bivariate model[26] to explore the following studies characteristics as potential sources of heterogeneity in test performance: year of publication, sample size, geographical area, and place of screening. Taking into account the variation of disease prevalence across studies and their pooled values within each geographical area, we computed negative and positive predictive values for each screening test from the pooled estimates of sensitivity and specificity using Bayes' theorem.

The pooled relative sensitivities and specificities of either screening test versus another were obtained from the pooled absolute accuracy measures assessed by a bivariate model with the method of moments,[27] allowing for inter-setting variability. This analysis was based on indirect comparisons, because the relative accuracy measures were obtained from aggregate estimates across all studies that contributed any sensitivity or specificity for either test. Because the within study correlation was unknown, we defined three scenarios in which correlation between sensitivities assessed for different screening tests in the same study was low (0), medium (0.5), and high (0.8). The value of the correlation had little effect on pooled relative sensitivities; therefore, we only presented results assuming a correlation of 0, which are slightly conservative.

To assess the robustness of our results, a leave one out sensitivity analysis of the pooled estimates of relative sensitivity and specificity, disease prevalence, and
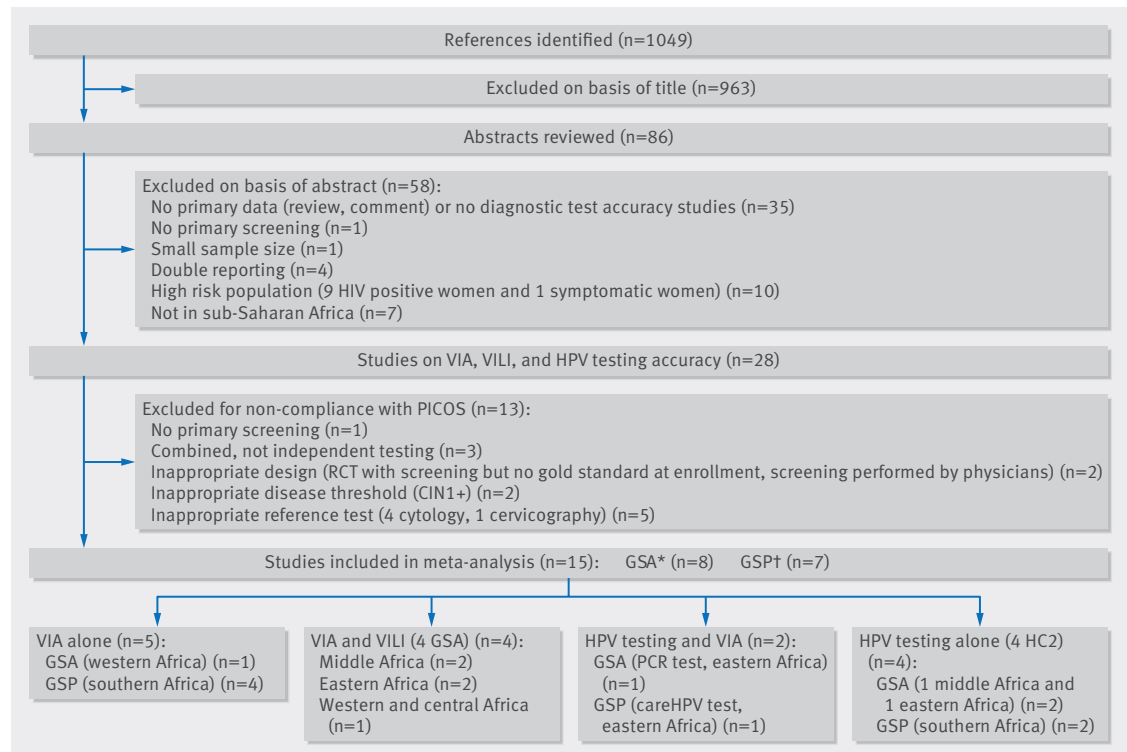
Fig 1 | Study flowchart of papers in selection process. *Gold standard performed in all women in the study population. †Gold standard performed in all women with a positive screening result and only a portion of women with a negative screening result. HC2=second generation hybrid capture assay; PICOS=population, intervention (screening tests), comparison (gold standard), outcomes (absolute and relative sensitivity and specificity, prevalence of CIN2+, and positivity rate of tests), study design; PCR=polymerase chain reaction; RCT=randomised controlled trial

tests' positivity rate were pooled with the double arcsine transformation of Freeman-Tukey.[28] For the relative accuracy estimates, we also made direct comparisons by restricting analyses to studies that assessed pairs of tests. Publication bias was assessed by funnel plots of the diagnostic odds ratios, as well as the trim and fill method.[29] We used Egger's test to detect a potential association between the precision and the value of the estimates.

We defined statistical significance as P<0.05 for two sided tests. All statistical analyses were performed with the R packages metafor version 1.9.3 for meta-analysis of prevalence and positivity rates,[30] mada version 0.5.5 for bivariate meta-analysis of sensitivity and specificity,[31] and mvmeta version 0.4.4 for assessing pooled relative sensitivities and specificities.[32]

Results

Study characteristics
Of 1049 entries identified via searches, 15 papers (eight in the GSA group, seven in the GSP group) fulfilled the inclusion criteria[19 33-46] (fig 1). Two thirds of included papers were from southern (six papers) or eastern (four) Africa. One article reported data from five countries in western and middle Africa;[40] we considered them separately. Therefore, the meta-analyses included 15 reports (10 in the GSA group; five in the GSP group) of 61 381 women screened by VIA, eight reports (all GSA) of 46 435 women screened by VILI, and six reports (three GSA; three GSP) of 11 322 women screened by HPV

testing (four on second generation hybrid capture assays, one on careHPV, and one on PCR).

Included women were more likely to be rural dwellers, previously unscreened, and asymptomatic. Women attending primary health centres for non-gynecological symptoms were recruited in two studies;[34 37] in another study, participants were reported to be healthy women attending an urban family planning clinic (84%), and a small proportion of women (16%) referred for gynecological complaints.[38] Age of participants across studies ranged from 15 to 83 years. Screeners were more often nurses, whereas colposcopic examination, punch biopsy, loop excision, or endocervical curettage when indicated were done by trained physicians. Histological interpretation of biopsies was provided by skilled pathologists. Table presents a summary of study characteristics.

Assessment of studies quality
Overall, the 15 selected papers were methodologically of moderate quality. Poor grades for the 13 QUADAS items ranged from 15% (two items) to 54% (seven), equivocal grades from 8% (one) to 39% (five), and good grades from 23% (three) to 77% (10; web table 2, web appendix 5). Concerning participants' selection, risk of bias was low in 40% (six papers) of the studies and unclear in 60% (nine); concerns on applicability were small in 93% (14) and major in 7% (one) of the studies. In the conduct and interpretation of screening tests, overall risk of bias was low in 14 (93%) studies and

Table 1 | Characteristics of included studies

Study	Publication year	Screening test(s)	Country	No of women	Screener	Place of screening	Gold standard category	HPV testing technique
Megevand[33]	1996	VIA	South Africa	2426	Nurse	Field clinic	GSP	—
University of Zimbabwe[34]	1999	VIA	Zimbabwe	2130	Nurse	Primary	GSA	—
Denny[35]	2000	VIA	South Africa	2944	Nurse	Field clinic	GSP	—
Wright[36]	2000	HPV testing	South Africa	1365	Nurse	Primary	GSP	HC2
Kuhn[38]	2000	HPV testing	South Africa	2861	Nurse	Primary	GSP	HC2
Womack[37]	2000	HPV testing	Zimbabwe	2145	Nurse	Primary	GSA	HC2
Denny[39]	2002	VIA	South Africa	2698	Nurse	Field clinic	GSP	—
Sankaranarayan[40]	2004	VIA/VILI	Burkina	2051/2051	Nurse	Field clinic	GSA	—
Sankaranarayan[40]	2004	VIA/VILI	Congo	6935/6935	Nurse	Field clinic	GSA	—
Sankaranarayan[40]	2004	VIA/VILI	Guinea	7462/7462	Nurse	Field clinic	GSA	—
Sankaranarayan[40]	2004	VIA/VILI	Mali	5552/5552	Nurse	Field clinic	GSA	—
Sankaranarayan[40]	2004	VIA/VILI	Niger	1827/1827	Nurse	Field clinic	GSA	—
H De Vuyst[41]	2005	VIA/HPV testing	Kenya	653/453	Nurse	Tertiary	GSA	PCR
Sangwa-Lugoma[42]	2006	VIA/VILI	Congo	1528/1528	Nurse	Primary	GSA	—
Muwonge[43]	2010	VIA/VILI	Angola	8849/8842	Nurse	Primary	GSA	—
Ngoma[44]	2010	VIA/VILI	Tanzania	10 374/10 367	Nurse	Primary	GSA	—
Mahmud[46]	2012	HPV testing	Congo	1352	Nurse	Primary	GSA	HC2
Ahmed Ibrahim[45]	2012	VIA	Sudan	934	Nurse	Primary	GSP	—
Jeronimo[19]	2014	VIA/HPV testing	Uganda	3146/3146	Nurse	Tertiary	GSP	careHPV

HC2=second generation hybrid capture assay; Gold standard=colposcopy followed by colposcopy directed biopsies; GSA=gold standard performed in all women of the study population; GSP=gold standard performed in screen positive women and only a portion of screen negative women; PCR=polymerase chain reaction.

Table 2 | Prevalence of disease (CIN2+) and positivity rate of VIA, VILI, and HPV testing in sub-Saharan Africa

Test and subgroup	No of studies	Prevalence of CIN2+		Positivity rate	
		Pooled estimate (%; 95% CI)	I^2, P*	Pooled estimate (%; 95% CI)	I^2 (%), P*
VIA					
GSA	10	3.3 (2.1 to 4.7)	98.4, P<0.001	17.4 (10.4 to 25.6)	99.8, P<0.001
GSP	5	3.5 (2.1 to 5.4)	96.0, P<0.001	15.8 (6.0 to 29.0)	99.7, P<0.001
P†	—	0.82	—	0.83	—
All studies	15	3.4 (2.4 to 4.4)	97.9, P<0.001	16.8 (11.0 to 23.6)	99.8, P<0.001
VILI					
(all GSA)	8	2.3 (1.5 to 3.3)	97.6, P<0.001	16.5 (9.8 to 24.7)	99.8, P<0.001
HPV testing					
GSA	3	6.9 (1.7 to 15.2)	98.4, P<0.001	31.4 (11.5 to 55.8)	99.5, P<0.001
GSP	3	3.3 (2.6 to 4.0)	62.2, P=0.0708	20.7 (19.1 to 22.4)	66.3, P=0.0516
P†	—	0.23	—	0.34	—
All studies	6	4.9 (2.7 to 7.8)	97.4, P<0.001	25.8 (17.4 to 35.3)	99.2, P<0.001

I^2=heterogeneity index; GSA=gold standard (colposcopy and directed biopsies) performed in all women of the study population; GSP=gold standard performed in a portion of the study population, including all women who screened positive.
*P value from the Cochran's Q test for detecting heterogeneity.
†P value for comparison between subgroups.

unclear in one (7%); suitability for our review was estimated to be good in nine (60%) studies, unclear in two (13%), and poor in four (27%).

The imperfect gold standard bias was present in all studies, because the reference test was a combination of colposcopy and histology. Although the histological results were interpreted without knowledge of the index test's results in 60% of studies, this information was not available in 40% of studies. The risk of incorporation bias was greater in studies evaluating the accuracy of visual methods than in studies evaluating the accuracy of HPV testing. Applying the reference test to our research question was slightly problematic in all but one study.[41] The time interval between completion of the index test and confirmation with the reference test was shorter than one month in 12 (80%) of the 15 studies, longer than one month in two studies (13%), and unreported in one study (7%). The risk of ascertainment bias was lower in studies of the GSA group than in those of the GSP group (53% v 47%). Reasons for excluding participants were reported in nine studies (60%) and not given in six studies (40%). Most reports did not provide enough information about uninterpretable results for the index test (13 articles, 87%), and up to half of studies did not clearly mention uninterpretable results of the reference test (seven papers, 47%).

Quality assurance of the reference test was carried out and reported in six studies (all conducted under the auspices of the IARC); quality control was conducted internally (regular onsite assessment and retraining of colposcopists and pathologists throughout the project) and externally (review of slides in a reference laboratory). In eight studies, histological samples were shipped to and analysed by skilled pathologists in the United States, Canada, and Belgium, but quality assurance was not mentioned except for in one report where histological interpretation was performed independently by two pathologists. The remaining six studies did not provide information on quality assurance of the reference test.

Prevalence of CIN2+ and positivity of screening tests

Table 2 summarises the pooled estimates of disease prevalence and positivity rate of screening tests by study group; web appendix 6 shows the forest plots. Four studies (three GSA, one GSP) reported prevalence higher than 8%, whereas the other studies reported it below 6%. The pooled prevalence varied by geographical region (P=0.03; 2.0% (95% confidence interval 1.2% to 2.9%) in western Africa, 2.5% (0.8% to 5.1%) in middle Africa, 3.7% (2.4% to 5.3%) in southern Africa, and 6.2% (2.9% to 10.5%) in eastern Africa). In the GSA

group, there was no significant difference in the pooled prevalence of CIN2+ between HPV testing and VIA studies (6.9% v 3.3%, P=0.24) and between HPV testing and VILI studies (6.9% v 2.3%, P=0.10; table 2).

The pooled positivity rate of VIA was 16.8% (95% confidence interval 11.0% to 23.6%) and ranged between 3.1% and 39.9%. Likewise, the pooled positivity rate for VILI was 16.5% (9.8% to 24.7%), with the highest value in Angola (32.0%) and the lowest value in Tanzania (4.8%). The positivity rate for HPV testing varied from 12.5% in the Democratic Republic of Congo (DRC) to 42.8% in Zimbabwe, with a pooled estimate of 25.8% (17.4% to 35.3%). Overall, included studies were highly heterogeneous with respect to disease prevalence and positivity rate of the screening tests (table 2).

Accuracy of VIA, VILI, and HPV testing
Figure 2 shows the variation of absolute sensitivity and specificity of VIA, VILI, and HPV testing in detecting CIN2+ in the GSA group. The pooled sensitivity for VIA in this group was 82.4% (95% confidence interval 76.3% to 87.3%) and ranged between 65.0% (40.8% to 84.6%) in Mali and 94.4% (90.6% to 97.0%) in Tanzania. The pooled specificity of VIA was 87.4% (77.1% to 93.4%), with the lowest value observed in Zimbabwe (64.1%, 61.9% to 66.2%) and the highest in Tanzania (98.2%, 98.0% to 98.5%).

For VILI, absolute sensitivity varied between 72.4% (confidence interval 52.8% to 87.3%) in DRC and 99.1% (96.9% to 99.9%) in Tanzania, whereas specificity ranged from 73.1% (71.1% to 75.0%) in Burkina Faso to 97.3% (97.0% to 97.6%) in Tanzania. The pooled sensitivity of VILI (95.1%, 90.1% to 97.7%) was significantly higher than that for VIA in the GSA group. However, the pooled specificity of VILI (87.2%, 78.1% to 92.8%) was not different from that for VIA.

Pooled absolute sensitivity of HPV testing was 88.3% (confidence interval 73.1% to 95.5%), with the lowest estimate (80.2%, 74.1% to 85.4%) observed in Zimbabwe and the highest value (96.2%, 87.0% to 99.5%) found in the only study that used PCR for HPV detection, in Kenya.[41] Absolute specificity of HPV testing varied between 61.2% (59.0% to 63.4%) in Zimbabwe and 88.9% (87.0% to 90.5%) in DRC, with a pooled value of 73.9% (50.7% to 88.7%). The pooled sensitivity and specificity of HPV testing were not different from those for VIA (both P≥0.23) and VILI (both P≥0.16).

Table 3 shows the pooled relative sensitivity and specificity of each test versus others using indirect comparisons. This ratio confirms that pooled absolute sensitivity for VIA was significantly lower than for VILI, because the relative sensitivity of VIA versus VILI (0.86, 95% confidence interval 0.82 to 0.90, P<0.001) was significantly lower than 1. However, relative sensitivity

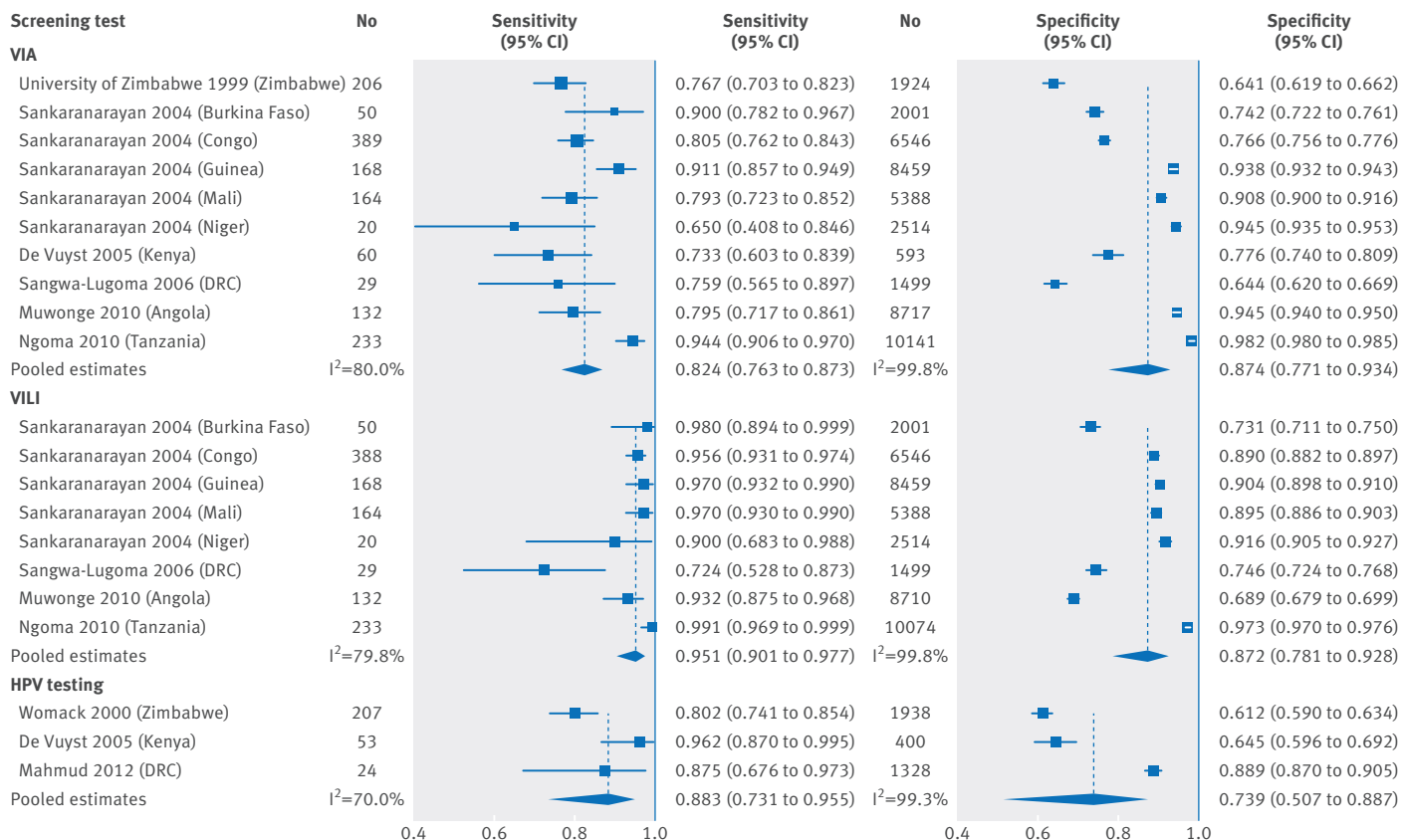| Screening test	No	Sensitivity (95% CI)	No	Specificity (95% CI)
VIA				
University of Zimbabwe 1999 (Zimbabwe)	206	0.767 (0.703 to 0.823)	1924	0.641 (0.619 to 0.662)
Sankaranarayan 2004 (Burkina Faso)	50	0.900 (0.782 to 0.967)	2001	0.742 (0.722 to 0.761)
Sankaranarayan 2004 (Congo)	389	0.805 (0.762 to 0.843)	6546	0.766 (0.756 to 0.776)
Sankaranarayan 2004 (Guinea)	168	0.911 (0.857 to 0.949)	8459	0.938 (0.932 to 0.943)
Sankaranarayan 2004 (Mali)	164	0.793 (0.723 to 0.852)	5388	0.908 (0.900 to 0.916)
Sankaranarayan 2004 (Niger)	20	0.650 (0.408 to 0.846)	2514	0.945 (0.935 to 0.953)
De Vuyst 2005 (Kenya)	60	0.733 (0.603 to 0.839)	593	0.776 (0.740 to 0.809)
Sangwa-Lugoma 2006 (DRC)	29	0.759 (0.565 to 0.897)	1499	0.644 (0.620 to 0.669)
Muwonge 2010 (Angola)	132	0.795 (0.717 to 0.861)	8717	0.945 (0.940 to 0.950)
Ngoma 2010 (Tanzania)	233	0.944 (0.906 to 0.970)	10141	0.982 (0.980 to 0.985)
Pooled estimates	I²=80.0%	0.824 (0.763 to 0.873)	I²=99.8%	0.874 (0.771 to 0.934)
VILI				
Sankaranarayan 2004 (Burkina Faso)	50	0.980 (0.894 to 0.999)	2001	0.731 (0.711 to 0.750)
Sankaranarayan 2004 (Congo)	388	0.956 (0.931 to 0.974)	6546	0.890 (0.882 to 0.897)
Sankaranarayan 2004 (Guinea)	168	0.970 (0.932 to 0.990)	8459	0.904 (0.898 to 0.910)
Sankaranarayan 2004 (Mali)	164	0.970 (0.930 to 0.990)	5388	0.895 (0.886 to 0.903)
Sankaranarayan 2004 (Niger)	20	0.900 (0.683 to 0.988)	2514	0.916 (0.905 to 0.927)
Sangwa-Lugoma 2006 (DRC)	29	0.724 (0.528 to 0.873)	1499	0.746 (0.724 to 0.768)
Muwonge 2010 (Angola)	132	0.932 (0.875 to 0.968)	8710	0.689 (0.679 to 0.699)
Ngoma 2010 (Tanzania)	233	0.991 (0.969 to 0.999)	10074	0.973 (0.970 to 0.976)
Pooled estimates	I²=79.8%	0.951 (0.901 to 0.977)	I²=99.8%	0.872 (0.781 to 0.928)
HPV testing				
Womack 2000 (Zimbabwe)	207	0.802 (0.741 to 0.854)	1938	0.612 (0.590 to 0.634)
De Vuyst 2005 (Kenya)	53	0.962 (0.870 to 0.995)	400	0.645 (0.596 to 0.692)
Mahmud 2012 (DRC)	24	0.875 (0.676 to 0.973)	1328	0.889 (0.870 to 0.905)
Pooled estimates	I²=70.0%	0.883 (0.731 to 0.955)	I²=99.3%	0.739 (0.507 to 0.887)

Fig 2 | Absolute sensitivity and specificity of VIA, VILI, and HPV testing for CIN2+ detection in sub-Saharan Africa. Diamond=95% confidence interval of the pooled measure computed with a bivariate random effects model. This figure includes only studies where the gold standard (colposcopy and colposcopy directed biopsies) was performed in all women of the study population (that is, GSA group). Heterogeneity across studies was assessed with Cochran's Q test for each screening test (all P<0.05). DRC=Democratic Republic of Congo

Table 3 | Pooled relative sensitivity and specificity of VIA, VILI, and HPV testing for CIN2+ detection in sub-Saharan Africa based on indirect comparisons

	No of studies	Relative sensitivity (95% CI)*	Relative specificity (95% CI)*
VIA v VILI	10 v 8	0.86 (0.82 to 0.90)	1.01 (0.90 to 1.13)
VIA v HPV testing	10 v 3	0.94 (0.82 to 1.16)	1.17 (0.95 to 1.69)
VILI v HPV testing	8 v 3	1.08 (0.98 to 1.30)	1.18 (0.96 to 1.71)

*Ratios higher than 1 indicate greater sensitivity or specificity of the first test, whereas ratios lower than 1 indicate lower sensitivity or specificity of the first test. When 95% confidence intervals do not include the unity, the difference in sensitivity or specificity between tests is statistically significant.

between VIA and HPV (0.94, 0.82 to 1.16) and between VILI and HPV (1.08, 0.98 to 1.30) was not significantly different from 1 (both P≥0.14). Pooled specificity was similar for VIA and VILI: relative specificity was 1.01

(0.90 to 1.13, P=0.85). Moreover, pooled specificity for HPV testing was lower than for VIA and VILI but not significantly, with relative values of 1.17 (0.95 to 1.69, P=0.15) and 1.18 (0.96 to 1.71, P=0.15), respectively.

Figure 3 shows the joint variation of observed and pooled sensitivity and specificity by screening test in the GSA group. The wide area of the confidence ellipse for HPV testing reflects the small number of accuracy studies conducted on HPV testing in sub-Saharan Africa. Curves for VILI are closer to the upper left corner of the rectangle than those for VIA, without overlapping ellipses between the two tests (fig 3). These results underpin the better performance of VILI compared with VIA for primary screening of cervical cancer in sub-Saharan Africa.
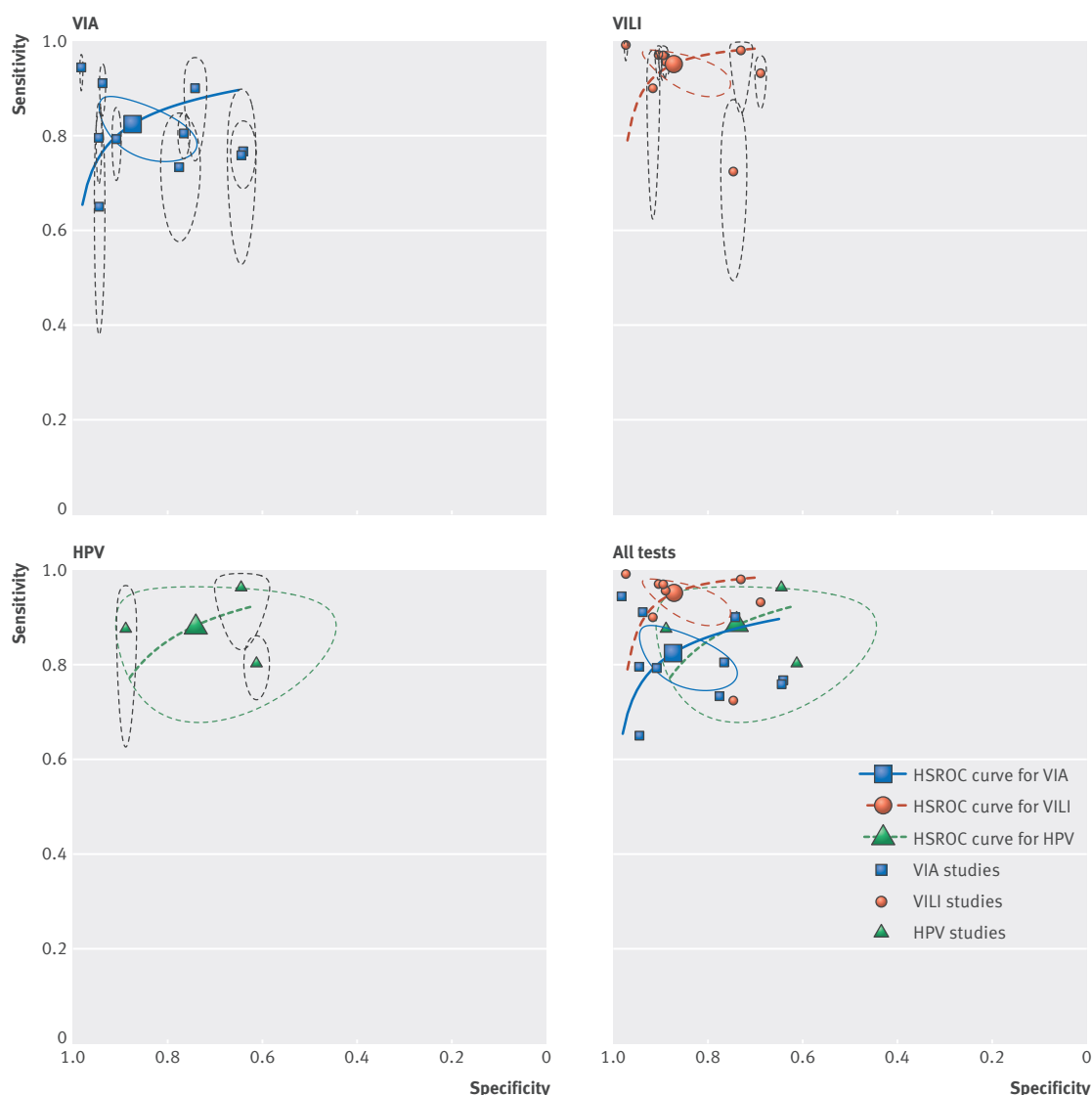
Fig 3 | Absolute accuracy of VIA, VILI and HPV testing for primary cervical cancer screening in sub-Saharan Africa. HSROC=hierarchical summary receiver operating characteristic regression curves. HSROC curves depict the fitted sensitivity as a function of specificity for VIA, VILI, HPV testing, and all three tests combined to detect CIN2+. These curves include only studies where gold standard (colposcopy and colposcopy directed biopsies) was applied to all women (that is, GSA group). Small blue squares, red circles, and green triangles and grey line circles=individual studies and 95% confidence ellipses; large blue squares, red circles, and green triangles and circles=pooled sensitivity and specificity and 95% confidence ellipses

Table 4 | Analysis of heterogeneity factors in assessing accuracy of VIA and VILI in sub-Saharan Africa

Factors (no of studies)	Sensitivity (%; 95% CI)	Adjusted odds ratio (95% CI)	P	Specificity (%; 95% CI)	Adjusted odds ratio (95% CI)	P
VIA						
Geographical area						
Middle Africa (n=3)	80.0 (76.4 to 83.1)	1 (reference)	—	82.4 (55.5 to 94.6)	1 (reference)	—
Western Africa (n=4)	84.0 (70.8 to 91.9)	1.35 (0.49 to 3.92)	0.53	90.2 (80.6 to 95.4)	1.97 (0.29 to 13.30)	0.49
Eastern Africa (n=3)	85.1 (66.8 to 94.2)	1.45 (0.49 to 4.26)	0.50	87.5 (46.8 to 98.2)	1.49 (0.19 to 11.51)	0.70
Place of screening*						
Field clinic (n=5)	83.0 (73.7 to 89.5)	1 (reference)	—	88.2 (78.2 to 94.0)	1 (reference)	—
Primary (n=4)	83.3 (69.4 to 91.7)	1.03 (0.42 to 2.50)	0.95	88.2 (77.3 to 94.3)	1.01 (0.18 to 5.47)	0.99
Sample size† (n=10)	—	1.13 (1.03 to 1.23)	0.01	—	1.28 (1.10 to 1.48)	0.001
Publication year‡ (n=10)	—	1.07 (0.95 to 1.21)	0.24	—	1.26 (1.04 to 1.53)	0.02
VILI						
Geographical area§						
Middle Africa (n=3)	90.4 (73.6 to 97.0)	1 (reference)	—	78.9 (66.4 to 89.0)	1 (reference)	—
Western Africa (n=4)	96.9 (93.9 to 98.5)	2.62 (0.68 to 10.09)	0.16	87.5 (92.9 to 78.9)	1.86 (0.70 to 4.97)	0.21
Place of screening						
Field clinic (n=5)	96.4 (94.2 to 97.8)	1 (reference)	—	87.8 (81.5 to 92.1)	1 (reference)	—
Primary (n=3)	94.2 (64.8 to 99.3)	0.66 (0.13 to 3.41)	0.62	86.1 (51.9 to 97.3)	1.16 (0.28 to 4.86)	0.84
Sample size† (n=8)	—	1.32 (1.12 to 1.57)	0.001	—	1.14 (0.94 to 1.38)	0.18
Publication year‡ (n=8)	—	1.09 (0.79 to 1.51)	0.61	—	1.03 (0.79;1.36)	0.81

*The place of screening was tertiary for one study, which was omitted in the analysis.
†Adjusted odds ratio per 1000 additional trial participants.
‡Adjusted odds ratio per additional year.
§Only one VILI study was located in eastern Africa, which was omitted in the analysis.

Heterogeneity factors of screening tests accuracy

Although accuracy measures showed large inconsistencies among studies for all screening tests (most P<0.05 for the Cochran Q tests), the I² heterogeneity index was generally lower for HPV testing. Potential sources of variability were analysed in the GSA group for VIA and VILI (table 4). VIA and VILI accuracy did not vary by geographical area or by place of screening. However, the specificity for VIA was higher in recent studies than in older ones when definition of VIA was less precise. Moreover, sensitivity of VIA and VILI and specificity of VIA to detect CIN2+ increased significantly with the sample size, which we considered as a proxy for screening experience. The few number of studies in the GSA group (n=3) precluded reliable heterogeneity analyses for HPV testing.

Sensitivity analyses and publication bias assessment

In a leave one out sensitivity analysis, relative specificity of VIA versus HPV testing increased to 1.41 (95% confidence interval 1.26 to 1.54, P<0.001) when the study by Mahmud and colleagues[46] was removed. Specificity of HPV testing in this study was largely higher (88.9%) than in the other two studies (61.2% and 64.5%), and its exclusion resulted in a decrease in the pooled absolute specificity from 73.9% (95% confidence interval 50.7% to 88.7%) to 62.1% (59.0% to 64.9%). Furthermore, the relative sensitivity of VILI versus HPV testing was significantly different from 1 (1.18, 1.09 to 1.27, P<0.001) when the only study that used PCR for HPV detection[41] was removed, with the pooled absolute sensitivity of HPV testing decreasing from 88.3% (73.1% to 95.5%) to 80.8% (75.2% to 85.5%). Statistical significance was also achieved (1.09, 1.00 to 1.31, P=0.04) when the study by

Sangwa-Lugoma and colleagues[42] was removed, with the absolute sensitivity of VILI moving from 95.1% (90.1% to 97.7%) to 96.2% (94.0% to 97.6%).

Changes in other estimates were, in general, non-significant and clinically trivial. When we restricted analyses of the relative accuracy of screening tests to studies based on paired tests (direct comparisons), the results were consistent with those obtained by indirect comparisons except for the relative sensitivity between VIA and HPV testing (0.77 (95% confidence interval 0.63 to 0.89); P=0.004) that was based on a single study. No evidence of publication bias was found across VIA and VILI studies. Moreover, no asymmetry in the funnel plot of diagnostic odds ratios for VIA and VILI was detected by Egger's test (both P≥0.35). With the trim and fill approach, no study was missing for the funnel plot to be symmetric. Publication bias was not investigated for HPV testing, owing to the small number of studies. Finally, implementing variance stabilisation via double arcsine transformations had no meaningful effect on pooled prevalence and positivity rate estimates.

GSP studies

In the GSP group, four studies evaluated VIA, three HPV testing, and one both VIA and HPV testing. Pooled prevalence (P=0.82 for VIA and P=0.23 for HPV testing) and positivity rate (P=0.83 for VIA and P=0.34 for HPV testing) did not differ significantly between GSA and GSP groups (table 2). The pooled sensitivity of VIA in detecting CIN2+ was lower in GSP studies than in GSA studies (68.6% v 82.4%, P=0.009), while the specificity was similar (87.4% v 89.8%, P=0.73). No significant difference was found for sensitivity and specificity of HPV testing (P=0.56 and P=0.39, respectively) among the two groups. Web appendices 6 and 7 show variation of

CIN2+ prevalence, positivity rate, sensitivity and specificity across studies of the GPS group.

Discussion

Main findings

Results from this meta-analysis suggest that in the context of primary cervical cancer screening in sub-Saharan Africa, VILI seems to be the most accurate alternative to cytology. When performed by nurses, VILI was about 13% more sensitive than and as specific as VIA in this region. VILI was about 7% more sensitive and 13% more specific than HPV testing, although the difference was not significant.

In resource limited settings, highly sensitive tests are critical to ensure effectiveness of screening programmes by allowing early detection of cervical cancer cases and reducing the need to repeat tests at shorter intervals. In addition, screening tests should be minimally invasive and results should be provided as soon as possible to allow rapid treatment of lesions and minimise the duration and the number of visits. These requirements that might increase adhesion of women to screening are important in sub-Saharan Africa, where strategies based on less frequent screenings seem to be more appropriate. Although not currently recommended by WHO for cervical cancer screening, VILI probably meets the above requirements better than VIA in sub-Saharan Africa.

Interpretation of results

Many reasons could explain the better performance of VILI in sub-Saharan Africa. Compared with VIA, lesions on VILI were judged less equivocal and easily recognised by investigators in Africa.[47] Even experienced African colposcopists have found that colour changes yielded by iodine impregnation of cervical mucosa are more easily detectable than acetowhite switches observed after application of acetic acid.[42] When these tests are performed with the naked eye—by nurses and with lighting probably less adequate than that produced by colposcopes—VILI is understandably preferred to and more accurate than VIA in sub-Saharan Africa. Furthermore, the absence of a time gap between application of iodine on the cervix and appearance of mustard yellow lesion, the greater stability and persistence of lesions, and the reduced need for repeated soaking of the cervix with iodine, partly explain why screeners in Africa are more comfortable with VILI. However, interpretation of VILI is difficult in postmenopausal women, because non-glycogenated atrophied cervical mucosa does not take up iodine and appears yellowish.[13]

The high absolute accuracy estimates of VIA and VILI reported in included studies can be explained by several reasons. In most studies, the reference investigations were applied by newly trained clinicians in techniques such as colposcopy and directed biopsy, and their learning curves possibly evolved as studies progressed. At most study locations, there was either no or limited previous screening and colposcopy experiences, and providers were most often rapidly trained

when the studies were initiated. Similarly, the pathologists involved in the studies probably overcalled CIN 2+ lesions because of their limited experience, recent reorientation, and lack of adequate quality assurance. Moreover, the correlation between visual techniques (VIA, VILI, and colposcopy) might have been another contributory factor. Finally, because most women had never been previously screened, the lesions detected might have been larger in size.

Evaluation of HPV testing's accuracy in this meta-analysis was less precise than that of VIA and VILI, since only three studies in the GSA group fulfilled our inclusion criteria. Furthermore, this review did not include studies of HPV testing on self obtained specimens and could not assess the effect of the sampling device (tampon, brush, lavage, or swab) on test accuracy. The lower interstudy variability for this molecular method could be explained by the fewer reports analysed, but it might also reflect the higher reproducibility of HPV test results, which are not or slightly dependent on the screener. Moreover, only one study using the careHPV assay (which has been designed for use in limited resource settings such as sub-Saharan Africa) was eligible. Although primary results from studies conducted in Asia are promising,[48-51] further trials need to explore the feasibility and efficacy of this new assay, assess the most adequate age range for maximum effectiveness, and determine the optimal threshold for a positive result of careHPV testing in the context of primary cervical cancer screening in Africa.

We could not perform pooled analysis of predictive values of screening tests, because they are closely related to the disease prevalence. However, we used the pooled prevalence of disease obtained among studies to estimate the post-test probability of having CIN2+ or being free of disease by geographical area (fig 4 , web appendix 8). The pooled positive predictive value was lower than 35% for any screening tests wherever the geographical area, suggesting that combining tests might improve the effectiveness of screening programmes. In such situations, the second test can be either an adjunct to the first, or used to select women among those who screened positive to the first test and who are at higher risk for cervical cancer and might need immediate care. Although VILI has been proposed as an adjunctive test to VIA,[13] results from primary studies are controversial. A study in Colombia found that the combination of VIA and VILI performed better than VIA alone,[52] whereas a recent study in Kenya found no difference.[53]

Limitations of study

VILI has never been implemented as a single test for primary screening, and has always been evaluated following VIA. Despite all precautions to ensure independence between the two tests in selected studies (such as different providers, and blinding of providers towards each test), some level of contamination cannot be eliminated. In view of the uncertainties for VILI as a standalone test, the value and limitations of using both visual methods in a sequential manner may be
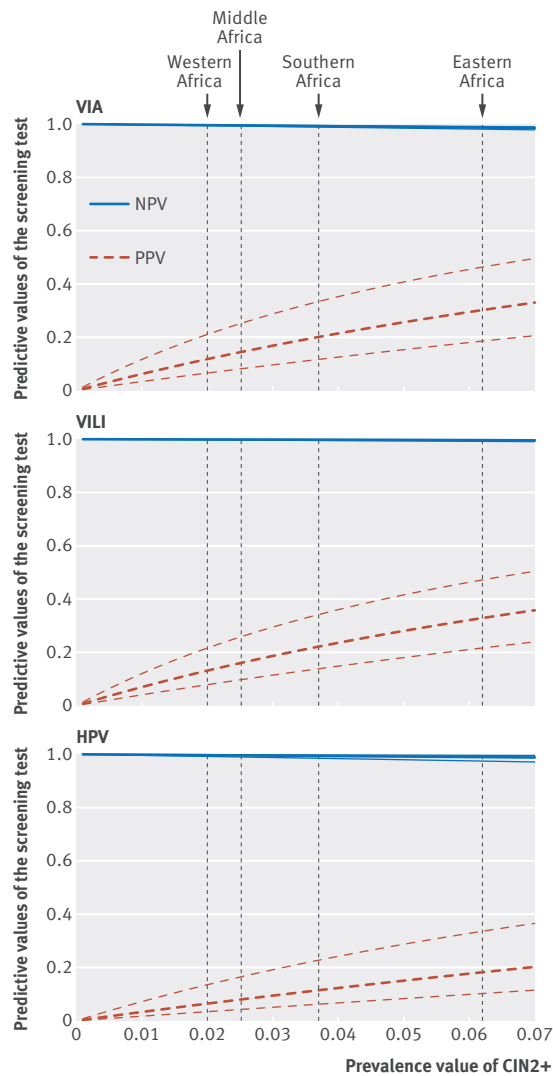
Fig 4 | Positive and negative predictive values of VIA, VILI, and HPV testing in detecting CIN2+ in sub-Saharan Africa. Curves=variation of positive and negative predictive values as a function of the prevalence of disease (CIN2+) using the Bayes'theorem, by screening test and by geographical region (western Africa, middle Africa, southern Africa, and eastern Africa). We considered the absolute sensitivity and specificity of each test to be constant and equal to the pooled values calculated with a random effects model. Prevalence of CIN2+ is shown on the x axis, and the resulting positive or negative predictive values are noted on the y axis. Heavy dotted curves=fitted predictive values; light dotted curves=95% confidence bands; NPV=negative predictive value; PPV=positive predictive value

considered. On the other hand, studies did not provide enough data to compare screening methods' performance by HIV status, and experience and background of screeners.

HIV infection is still very common in Africa, and the prevalence of cervical dysplasia in HIV positive women reaches 50-76% in some settings.[54][55] It is important to identify which screening methods are more appropriate for this high risk population. No study used histology of random cervical biopsy alone as a reference test. Although the use of a colposcopy based gold standard

is likely to overestimate the accuracy of visual methods (VIA and VILI) and underestimate HPV testing's accuracy,[56][57] the effect of QUADAS-2 items regarding the reference test (imperfect gold standard bias, incorporation bias, and masking towards index test) on screening tests' accuracy could not be assessed.

Conclusions and policy implications

Our findings suggest that tests qualities for VIA, VILI, and HPV testing are accurate enough for cervical cancer screening in sub-Saharan Africa, provided that adequate training and quality assurance are achieved. However, high absolute accuracy measures obtained for these tests are probably misleading, given the average quality of studies and the limitations of the gold standard. Among visual methods, VILI seems to be the most sensitive test to use in the African continent. It is deemed easy to perform by Africans clinicians, and is an acceptable alternative to VIA for cervical cancer screening in this region. Finally, further studies are needed to assess the effect of screening strategies using these tests on incidence and mortality from cervical cancer in the region.

1   Castellsague X, de Sanjose S, Aguado K, et al. HPV and cervical cancer in the 2007 report. Vaccine 2007;25 suppl 3:C1-230.
2   Arbyn M, Castellsague X, de Sanjose S, et al. Worldwide burden of cervical cancer in 2008. Ann Oncol 2011;22:2675-86.
3   Blumenthal P, Gaffikin L, Deganus S, Lewis R, Emerson M, Adadevoh S. Cervical cancer prevention:safety, acceptability, and feasibility of a single-visit approach in Accra, Ghana. Am J Obstet Gynecol 2007;196:407 e1-8; discussion 407 e8-9.
4   Katz I, Wright A. Preventing cervical cancer in the developing world. N Engl J Med 2006;354:1110.

5   Goldie S, Kuhn L, Denny L, Pollack A, Wright T. Policy analysis of cervical cancer screening strategies in low-resource settings: clinical benefits and cost-effectiveness. *JAMA* 2001;285:3107-15.

6   Mandelblatt J, Lawrence W, Gaffikin L, et al. Costs and benefits of different strategies to screen for cervical cancer in less-developed countries. *J Natl Cancer Inst* 2002;94:1469-83.

7   Goldie S, Gaffikin L, Goldhaber-Fiebert J, et al. Cost-effectiveness of cervical-cancer screening in five developing countries. *N Engl J Med* 2005;353:2158-68.

8   Fokom-Domgue J, Vassilakos P, Petignat P. Is screen-and-treat approach suited for screening and management of precancerous cervical lesions in Sub-Saharan Africa? *Prev Med* 2014;65:138-40.

9   Adefuye P, Broutet N, de Sanjose S, Denny L. Trials and projects on cervical cancer and human papillomavirus prevention in sub-Saharan Africa. *Vaccine* 2013;31 suppl 5:F53-9.

10 Cronje H, Parham G, Cooreman B, de Beer A, Divall P, Bam R. A comparison of four screening methods for cervical neoplasia in a developing country. *Am J Obstet Gynecol* 2003;188:395-400.

11 Ajenifuja K, Gage J, Adepiti A, et al. A population-based study of visual inspection with acetic acid (VIA) for cervical screening in rural Nigeria. *Int J Gynecol Cancer* 2013;23:507-12.

12 Sauvaget C, Fayette J, Muwonge R, Wesley R, Sankaranarayanan R. Accuracy of visual inspection with acetic acid for cervical cancer screening. *Int J Gynaecol Obstet* 2011;113:14-24.

13 Sankaranarayanan R, Nessa A, Esmy P, Dangou J. Visual inspection methods for cervical cancer prevention. *Best Pract Res Clin Obstet Gynecol* 2012;26:221-32.

14 Cuzick J, Arbyn M, Sankaranarayanan R, et al. Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries. *Vaccine* 2008;26 suppl 10:K29-41.

15 Bradford L, Goodman A. Cervical cancer screening and prevention in low-resource settings. *Clin Obstet Gynecol* 2013;56:76-87.

16 Sankaranarayanan R, Wesley R. A practical manual on visual screening for cervical neoplasia. IARC Press and WHO Distribution and Sales, 2003:1-49.

17 Hadzic B, Hadzic M, Curcin N. Histologic classification and terminology of precancerous lesions of the cervix. *Med Pregl* 1999;52:151-5.

18 Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006-12.

19 Jeronimo J, Bansil P, Lim J, et al. A multicountry evaluation of careHPV testing, visual inspection with acetic acid, and papanicolaou testing for the detection of cervical cancer. *Int J Gynecol Cancer* 2014;24:576-85.

20 Whiting P, Rutjes A, Westwood M, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529-36.

21 Rucker G, Schwarzer G, Carpenter J, Olkin I. Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells. *Stat Med* 2009;28:721-38.

22 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.

23 Reitsma J, Glas A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982-90.

24 Rutter C, Gatsonis C. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med* 2001;20:2865-84.

25 Harbord R, Deeks J, Egger M, Whiting P, Sterne J. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239-51.

26 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.

27 Jackson D, Riley R, White IR. Multivariate meta-analysis: potential and promise. *Stat Med* 2011;30:2481-98.

28 Barendregt J, Doi S, Lee Y, Norman R, Vos T. Meta-analysis of prevalence. *J Epidemiol Community Health* 2013;67:974-8.

29 Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000;56:455-63.

30 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Software* 2010;36:1-48.

31 Doebler P, Holling H, Bohning D. A mixed model approach to meta-analysis of diagnostic studies with binary test outcome. *Psychol Methods* 2012;17:418-36.

32 Gasparrini A, Armstrong B, Kenward M. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med* 2012;31:3821-39.

33 Megevand E, Van Wyk W, Knight B, Bloch B. Can cervical cancer be prevented by a see, screen, and treat program? A pilot study. *Am J Obstet Gynecol* 1996;174:923-8.

34 University of Zimbabwe. Visual inspection with acetic acid for cervical-cancer screening: test qualities in a primary-care setting. University of Zimbabwe/JHPIEGO Cervical Cancer Project. *Lancet* 1999;353:869-73.

35 Denny L, Kuhn L, Pollack A, Wainwright H, Wright T. Evaluation of alternative methods of cervical cancer screening for resource-poor settings. *Cancer* 2000;89:826-33.

36 Wright T, Denny L, Kuhn L, Pollack A, Lorincz A. HPV DNA testing of self-collected vaginal samples compared with cytologic screening to detect cervical cancer. *JAMA* 2000;283:81-6.

37 Womack S, Chirenje Z, Blumenthal P, et al. Evaluation of a human papillomavirus assay in cervical screening in Zimbabwe. *BJOG* 2000;107:33-8.

38 Kuhn L, Denny L, Pollack A, Lorincz A, Richart R, Wright T. Human papillomavirus DNA testing for cervical cancer screening in low-resource settings. *J Natl Cancer Inst* 2000;92:818-25.

39 Denny L, Kuhn L, Pollack A, Wright T. Direct visual inspection for cervical cancer screening: an analysis of factors influencing test performance. *Cancer* 2002;94:1699-707.

40 Sankaranarayanan R, Basu P, Wesley RS, et al. Accuracy of visual screening for cervical neoplasia: Results from an IARC multicentre study in India and Africa. *Int J Cancer* 2004;110:907-13.

41 De Vuyst H, Claeys P, Njiru S, et al. Comparison of pap smear, visual inspection with acetic acid, human papillomavirus DNA-PCR testing and cervicography. *Int J Gynaecol Obstet* 2005;89:120-6.

42 Sangwa-Lugoma G, Mahmud S, Nasr S, et al. Visual inspection as a cervical cancer screening method in a primary health care setting in Africa. *Int J Cancer* 2006;119:1389-95.

43 Muwonge R, Manuel Mda G, Filipe A, Dumas J, Frank M, Sankaranarayanan R. Visual screening for early detection of cervical neoplasia in Angola. *Int J Gynaecol Obstet* 2010;111:68-72.

44 Ngoma T, Muwonge R, Mwaiselage J, Kawegere J, Bukori P, Sankaranarayanan R. Evaluation of cervical visual inspection screening in Dar es Salaam, Tanzania. *Int J Gynaecol Obstet* 2010;109:100-4.

45 Ibrahim A, Aro A, Rasch V, Pukkala E. Cervical cancer screening in primary health care setting in Sudan: a comparative study of visual inspection with acetic acid and Pap smear. *Int J Womens Health* 2012;4:67-73.

46 Mahmud S, Sangwa-Lugoma G, Nasr S, 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* 2012;124:286-91.

47 Sankaranarayanan R, Chatterji R, Shastri S,et al. Accuracy of human papillomavirus testing in primary screening of cervical neoplasia: results from a multicenter study in India. *Int J Cancer* 2004;112:341-7.

48 Ying H, Jing F, Fanghui Z, Youlin Q, Yali H. High-risk HPV nucleic acid detection kit—the careHPV test—a new detection method for screening. *Sci Rep* 2014;4:4704.

49 Labani S, Asthana S, Sodhani P, et al. CareHPV cervical cancer screening demonstration in a rural population of north India. *Eur J Obstet Gynecol Reprod Biol* 2014;176:75-9.

50 Kang L, Jeronimo J, Qiao Y, et al. Optimal positive cutoff points for careHPV testing of clinician- and self-collected specimens in primary cervical cancer screening: an analysis from rural China. *J Clin Microbiol* 2014;52:1954-61.

51 Qiao Y, Sellors J, Eder P, 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* 2008;9:929-36.

52 Murillo R, Luna J, Gamboa O, Osorio E, Bonilla J, Cendales R. Cervical cancer screening with naked-eye visual inspection in Colombia. *Int J Gynaecol Obstet* 2010;109:230-4.

53 Huchko M, Sneden J, Leslie H, et al. A comparison of two visual inspection methods for cervical cancer screening among HIV-infected women in Kenya. *Bull World Health Organ* 2014;92:195-203.

54 Parham G, Sahasrabuddhe V, Mwanahamuntu M, et al. Prevalence and predictors of squamous intraepithelial lesions of the cervix in HIV-infected women in Lusaka, Zambia. *Gynecol Oncol* 2006;103:1017-22.

55 Moodley J, Hoffman M, Carrara H, et al. HIV and pre-neoplastic and neoplastic lesions of the cervix in South Africa: a case-control study. *BMC Cancer* 2006;6:135.

56 Kohn M, Carpenter C, Newman T. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad Emerg Med* 2013;20:1194-206.

57 Pretorius R, Bao Y, Belinson J, Burchette R, Smith J, Qiao Y. Inappropriate gold standard bias in cervical cancer screening studies. *Int J Cancer* 2007;121:2218-24.

Web appendix: Supplementary material