BMJ Oncology

Accuracy of screening tests for cervical precancer in women living with HIV in low-resource settings: a paired prospective study in Lusaka, Zambia

Katayoun Taghavi [ID],[1] Misinzo Moono,[2] Mulindi Mwanahamuntu,[3,4] Marie Roumet,[5] Andreas Limacher,[5] Herbert Kapesa,[2] Thamsanqa Madliwa,[6] Anne Rutjes,[1] Partha Basu,[7] Nicola Low,[1] Albert Manasyan,[2,8] Julia Bohlius[1,9]

ABSTRACT

Objective This study aimed to provide evidence to improve cervical screening for women living with HIV (WLHIV). We assessed the accuracy of screening tests that can be used in low-resource settings and give results at the same visit.

Methods and analysis We conducted a paired, prospective study among consecutive eligible WLHIV, aged 18–65 years, receiving cervical cancer screening at one hospital in Lusaka, Zambia. The histopathological reference standard was multiple biopsies taken at two time points. The target condition was cervical intraepithelial neoplasia grade 2 and above (CIN2+). The index tests were high-risk human papillomavirus detection (hrHPV, Xpert HPV, Cepheid), portable colposcopy (Gynocular, Gynius) and visual inspection with acetic acid (VIA). Accuracy of stand-alone and test combinations were calculated as the point estimate with 95% CIs. A sensitivity analysis considered disease when only visible lesions were biopsied.

Results Women included in the study had well-controlled HIV infection (median CD4 count=542 cells/mm^3) and all except one were on antiretroviral therapy. Among 371 participants with histopathological results, 27% (101/371) women had CIN2+ and 23% (23/101) were not detected by any index test. Sensitivity and specificity for stand-alone tests were: hrHPV, 67.3% (95% CI 57.7% to 75.7%) and 65.3% (95% CI 59.4% to 70.7%); Gynocular 51.5% (95% CI 41.9% to 61.0%) and 80.0% (95% CI 74.8% to 84.3%); and VIA 22.8% (95% CI 15.7% to 31.9%) and 92.6% (95% CI 88.8% to 95.2%), respectively. Combining tests did not improve test accuracy measures. All test accuracies improved in sensitivity analysis.

Conclusion The low accuracy of screening tests assessed might be explained by our reference standard, which reduced verification and misclassification biases. Better screening strategies for WLHIV in low-resource settings are urgently needed.

Trial registration number NCT03931083.

INTRODUCTION

The WHO strategy to eliminate cervical cancer aims to improve prevention and treatment among women living with HIV

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The 2021 WHO guidelines recommend that women living with HIV (WLHIV) receive screening for human papillomavirus (hrHPV) genotypes at 3–5 years intervals, followed by a triage test to determine whether treatment is needed but this is based on low and moderate certainty evidence.

WHAT THIS STUDY ADDS

⇒ This study among WLHIV in Lusaka, Zambia evaluated three screening tests that allow same-day treatment; hrHPV test, portable colposcopy (Gynocular) and visual inspection with acetic acid (VIA), using strict methods to reduce verification and misclassification biases. The test accuracy of the different screening was poor, with sensitivities and specificity for stand-alone tests: hrHPV, 67.3% and 65.3%; Gynocular 51.5% and 80.0%; and VIA 22.8% and 92.6%; respectively.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our findings have implications for research and cervical cancer screening policies among WLHIV if test accuracy in this high-risk population has been overestimated by verification and misclassification biases in a majority of existing studies. Methodologically robust studies are crucial to inform cervical cancer screening practices and policies for the successful implementation of a cervical cancer elimination plan in sub-Saharan Africa, where 85% of women with cervical cancer and HIV live.

infection (WLHIV).[1] A conditional recommendation for WLHIV suggests testing for high-risk human papillomavirus (hrHPV) followed by an additional screening test based on moderate certainty evidence.[1] Cervical cancer remains the leading cause of cancer-related death among women in sub-Saharan African (SSA) countries, where more than half of cervical cancer cases are attributable to HIV.[2] Increased life expectancy on

antiretroviral therapy (ART) increases the number of women with persistent hrHPV infection, which may progress to cervical precancer and cancer.[3–5]

In low-resource settings, tests that give same-day results and lead to decisions about treatment are preferred. An evaluation of alternative same-day screening tests among WLHIV, using methods that minimise verification biases, has not yet been conducted.[1 6 7] Colposcopy is the cornerstone of visual assessment for cervical cancer screening, used in screening pathways of high-resource countries but is rarely accessible in low-resource settings where most WLHIV live. In low-resource settings, visual inspection with acetic acid (VIA) is commonly used,[5 6] but with low accuracy, particularly for WLHIV.[7 8] The WHO strategy recommends molecular tests to detect hrHPV, which were reported to have a sensitivity of 91.6% (95% CI 88%.1 to 94.1%) among WLHIV in a systematic review.[7] Our objectives were to assess the accuracy of molecular and visual screening tests (hrHPV testing, portable colposcopy using the Gynocular and VIA).

METHODS

This study followed a published protocol[9] and is reported according to the Standards for Reporting Diagnostic accuracy studies (STARD) 2015 guideline (online supplemental appendix S1). More details on our methodology have been previously described.[9]

Study design and participants

We conducted a single-site, paired (all women received all tests) prospective test accuracy study among WLHIV in Lusaka, Zambia. Nurses from the cervical cancer screening clinic came to the adjacent HIV clinic, informed them about cervical screening and invited them to volunteer for assessment of eligibility for the study. Consecutive eligible participants received a detailed explanation on the study in a private room where they had the possibility to ask questions. Written information about the study was available in English and local languages. Women who wished to participate, provided written consent and those who declined were encouraged to have standard care. We enrolled women aged 18–65 years with confirmed HIV infection who had ever had sex, gave written consent and agreed to return for a 6-month follow-up visit. We excluded women with a history of cervical cancer or total hysterectomy and those vaccinated against HPV. Women enrolled were a consecutive series who fulfilled eligibility criteria and for whom the research staff could complete all study procedures.

Procedures

Two nurses and one research assistant collected and tested specimens. To ensure independent test results and prevent bias, two different nurses performed procedures in separate rooms and findings were documented on separate case-report forms. Clinical team members did not discuss results, and participants were asked not to communicate findings to staff. After consent, a nurse recorded medical history and sociodemographic information. Blood tests for HIV RNA viral load (Cobas HIV-1/2 Qual; Roche Molecular Systems, New Jersey, USA) and CD4 cell count (Pima CD4 Analyzer, Alere, Waltham, USA) were taken at baseline.

Reference standard

The target condition was the histological presence of cervical intraepithelial neoplasia grade 2 and above (CIN2+) or high-grade squamous intraepithelial lesions (HSIL) at baseline or 6-month follow-up. A study nurse took biopsies during the colposcopy examination. If lesions were seen, she took at least two biopsies from those that looked the most severe. If no lesions were seen, she took four biopsies from clock-face positions 3, 6, 9 and 12 o'clock within the transformation zone. The nurse received training in colposcopy and biopsy taking from a gynaecologist based at the International Agency for Research on Cancer (IARC) and a local senior gynaecologist (online supplemental appendix S2). To further reduce detection bias in the reference standard, we took a second set of biopsies from each woman 6 months later to identify cases of disease missed at baseline. Biopsies were assessed histologically at two independent laboratories in South Africa and Zambia. An expert gynaecological pathologist in each laboratory, blinded to the clinical findings, examined all biopsies and classified them using the Bethesda squamous intraepithelial lesion system.[10] They reviewed all histopathology results via teleconference and reached an agreement on diagnosis. Any sample with CIN2 or ambiguous findings was tested with p16 immunostaining.[10 11] We dichotomised histopathological findings into low-grade and HSIL by the lower anogenital squamous terminology definitions and the WHO Classification of Tumours of Female Reproductive Organs.[10]

Index tests

A trained nurse (online supplemental appendix S2) did a speculum examination and collected specimens, followed by VIA. An endocervical sample was taken using a single-use cytobroom and immediately placed into ThinPrep PreservCyt solution (Hologic, Marlborough, Massachusetts, USA) (online supplemental appendix S2). During screening, an additional swab from the posterior vaginal fornix was also taken and tested for *Trichomonas vaginalis*. The research assistant processed the *T. vaginalis* and hrHPV specimens within 2–4 hours of collection using the GeneXpert platform (Cepheid, Sunnyvale, California, USA) at the study site, as per the manufacturers' instructions. The Xpert HPV test detects 14 hrHPV subtypes, categorised for reporting as HPV16, HPV18/45 (subtypes 18 and/or 45) and HPV other (any of subtypes 31, 33, 35, 39, 51, 52, 56, 58, 59, 66 and 68).

VIA examination followed IARC methodology (online supplemental appendix S2).[12] As per local guidelines, VIA nurses categorised indeterminate findings as abnormal. In a separate room, a different nurse performed a

colposcopic examination using the Gynocular (online supplemental appendix S3) following methods described in the IARC colposcopy manual.[13] We used the Swede score to standardise the documentation of findings on visual inspection with a score from 0 (abnormality not seen) to 2 (most severe)[14] (online supplemental appendix S4).

Treatment

Women who tested positive for VIA, CIN2+ or HSIL during screening were offered treatment as clinically indicated. Women with lesions eligible for cryotherapy or thermoablation could be treated at the time of screening by a trained cervical cancer screening nurse. Larger lesions would be treated with loop electrosurgical excision procedure, which would require a follow-up visit with a gynaecologist. Women with histopathologically confirmed cervical cancer were referred to the University Teaching Hospital in Lusaka for treatment. Women with confirmed *T. vaginalis* on swab were offered treatment with a 7-day course of metronidazole, as per local guidelines.

Interpretations of results

The histological presence of CIN2+ or HSIL at baseline or 6-month follow-up was considered as the disease outcome in the primary analysis. New cases of CIN2+ at follow-up were considered as diagnoses that had been missed at baseline. A positive hrHPV test result was defined as the detection of any of the 14 subtypes detected by the Xpert HPV test. VIA findings were dichotomised as positive (abnormal, suspicious of cancer or indeterminate) and negative (normal). We used receiver operating characteristic curve analysis to calculate the area under the curve (AUC) for each level of the Swede score as assessed by Gynocular colposcopy. We then used the Youden cut-off in the primary analysis, optimising both sensitivity and specificity. In an additional analysis, we used cut-offs maximising either sensitivity (≥90%) or specificity (≥90%).

Sample size and statistical analyses

We required a sample of 350 participants based on estimates of precision for the sensitivity and specificity of Gynocular, hrHPV and VIA as stand-alone tests for detecting CIN2+ lesions with approximately a 10%–15% margin of error (online supplemental appendix S5). We aimed to recruit 450 women to allow for incomplete data.

In our analyses, we used consensus agreement of the reference standard. We assessed agreement between the two pathologists for the reference standard using Cohen's kappa coefficient (κ). The accuracy measures used to assess Gynocular, VIA and hrHPV tests were sensitivity and specificity, positive and negative predictive values, positive and negative likelihood ratios, false positive and false negative rates, and diagnostic ORs (DORs). Screening test accuracy measures were estimated with 95% Wilson CIs. Using the same approach, we evaluated the accuracies of two tests used together. We considered the combination positive if both single tests were positive

and negative otherwise. This mimics the clinical scenario where the second test is used to decide whether treatment is required (triage test).[1]

We also described test accuracy measures in subgroups defined by age (<25, 26–35, 36–45, >46 years and menopausal status), parity, ART status, coinfection with *T. vaginalis*, methods of contraception and CD4 cell count. Sensitivity and specificity were calculated for each subgroup. Estimated sensitivity values were compared with those found in the reference category by calculating the sensitivity ratio. To investigate the occurrence of effect modification by patient characteristics on the association between the diagnostic test and disease status, we used univariable and multivariable logistic regression models and tested for the interaction between the diagnostic test and patient characteristics on disease status. We considered the following patient's characteristics: age, menopause, parity, ART status, *T. vaginalis* at baseline, methods of contraception, HIV RNA, CD4 cell count, history of treatment for precancer and education level. Adjustment was performed considering all before mentioned patient characteristics as predictors and performing a stepwise model selection based on the Akaike information criterion.

We conducted the following sensitivity analyses to investigate the influence of unverifiable assumptions. First, we explored a possible training effect by assessing the first 10% of participants separately. Second, we explored the impact of the COVID-19 pandemic by conducting primary analyses separately on women who finished the study before 28 March 2020 (study ceased due to the pandemic). Third, we assessed the impact of missing or indeterminate results in the reference standard or screening test results by considering them first as positive cases, then as negative. Finally, acknowledging that biopsy and HPV tests may be performed and interpreted differently, we conducted analyses using a reference standard from a hypothetical scenario in which a biopsy was taken only from visible lesions and using different categories of hrHPV test results.

Role of the funding source

The funders did not contribute to the study design, data collection, analysis, interpretation or writing of the manuscript.

RESULTS

Flow of participants

Between May 2019 and March 2021, we assessed 413 women, enrolled 376 and included 375 in the analysis (1 woman was found to have had a total hysterectomy; figure 1). We had valid reference standard results for 371 women. VIA and hrHPV tests were performed on the 375 enrolled with Gynocular examination conducted on 373. The follow-up period for deriving the reference standard was 6 months. There were no adverse events. Follow-up was completed for 104 women when national COVID-19
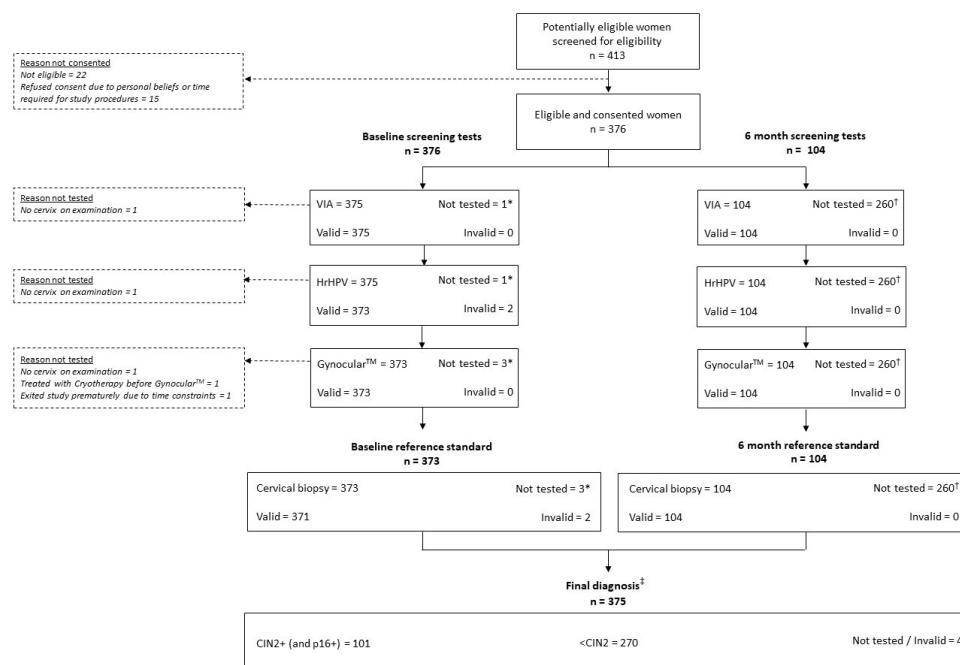
Figure 1 Flow of participants diagram to show the number of women receiving screening tests and reference standard, and analysed in the study. Data are n=number of women. *One women was excluded from the analysis as she did not receive any of the study screening tests - this brings the number of women included in the final diagnosis to 375. †Did not receive further tests as the study stopped following the COVID-19 pandemic. ‡The final diagnosis used in the analyses considers histopathological diagnosis for all women at baseline or 6-month follow-up, disease (CIN2+) was considered as present when biopsies from at least one time point were positive. CIN2+, cervical intraepithelial neoplasia grade 2 and above; <CIN2, cervical intraepithelial neoplasia grade 1 and below; HrHPV, high risk human papillomavirus; p16+, expression of cell cycle regulatory protein 16INK4A; VIA, visual inspection of the cervix with acetic acid.

restrictions on research studies meant that we had to stop the study. From March to December 2021, the official end of the study, we focused our efforts on ensuring that women who needed treatment received it. Study participants were contacted by phone and physically, by a peer educator, and invited to return to the clinic for treatment. However, despite all efforts, many women were unable to return. The study was officially closed in December 2021 as the pandemic persisted.

Patient characteristics

The full baseline characteristics are in table 1. At enrolment, participants had a median age of 37 years (IQR 31–44) and median parity of three (IQR 2–5). Most were not using any contraception (62%, n=231), did not smoke (99%, n=373) or use insunko[15] (a smokeless, carcinogenic tobacco product that can be used vaginally; 95%, n=355), and did not drink alcohol (88%, n=331). Most had never undergone cervical cancer screening (71%, n=267); VIA was the modality among those who had received screening. Seven women (2%) had received previous cryotherapy treatment. Almost all were on ART (99%, n=374) and had well-controlled HIV infection (median CD4 count=542 cells/mm^3). Women with histological CIN2+ were more likely to have a CD4 cell count <200 per mm^3 (7/101, 7%) than women without (5/270, 2%) and viral load ≥50 copies/mL (22/101, 22% and 36/270, 13%, respectively). We report baseline characteristics from routinely available data of all women aged 18–65 years seen at Kanyama HIV clinic in online supplemental appendix S6.

Disease spectrum

A consensus diagnosis of CIN2+ was made in 101 of 371 women with valid histology results (27.2%), of which 44 were CIN2, 56 were CIN3, and 1 was invasive cancer. The pathologists' agreement for determining CIN2+/HSIL was 71% (κ=0.37) at baseline and 82% (κ=0.46) at follow-up. Despite efforts to link all women with CIN2+ to care (online supplemental appendix S7), only 64/101 received treatment. Of these, 50 did not attend follow-up, 4 had positive histology at follow-up and 10 had negative histology results at follow-up. Prevalence of hrHPV was 43.5% (163/371) and *T. vaginalis* 19% (70/371) (online supplemental appendix S8).

Stand-alone screening test accuracy

Of 101 women with CIN2+, 23 (22.8%) had a negative result on all three screening tests (table 2). The stand-alone test with the highest point estimate for sensitivity was hrHPV testing (67.3%, 95% CI 57.7% to 75.7%) (figure 2, table of results in online supplemental appendix S9). Specificity was 65.3% (95% CI 59.4% to 70.7%). Women with CIN2+ were almost four times more likely to test positive for hrHPV than those without (DOR hrHPV 3.9, 95% CI 2.4 to 6.3). Using the Swede score, the AUC

Table 1   Baseline characteristics

Patient characteristics	Overall, n=375	CIN2+, n=101	<CIN2, n=270	Not tested/invalid, n=4
Age				
Median (IQR)	37 (31–44)	36 (28–42)	37 (31–44)	46 (42–54)
Missing	7	2	5	0
Menopause				
No	321 (86%)	90 (89%)	229 (85%)	2 (50%)
Yes	54 (14%)	11 (11%)	41 (15%)	2 (50%)
Marital status				
In a relationship/married	244 (65%)	70 (69%)	173 (64%)	1 (25%)
Separated/divorced	69 (18%)	12 (12%)	55 (20%)	2 (50%)
Single	12 (3%)	4 (4%)	8 (3%)	0 (0%)
Widowed	50 (13%)	15 (15%)	34 (13%)	1 (25%)
Employment				
Working	157 (42%)	39 (39%)	116 (43%)	2 (50%)
Not working	218 (58%)	62 (61%)	154 (57%)	2 (50%)
Education				
Did not finish secondary	308 (82%)	86 (85%)	219 (81%)	3 (75%)
Finished secondary	54 (14%)	12 (12%)	41 (15%)	1 (25%)
More than secondary	13 (4%)	3 (3%)	10 (4%)	0 (0%)
Income* (Kwacha/month)				
None	216 (58%)	61 (60%)	153 (57%)	2 (50%)
1–500	50 (13%)	16 (16%)	34 (13%)	0 (0%)
501–1000	74 (20%)	15 (15%)	57 (21%)	2 (50%)
1001–2500	32 (9%)	9 (9%)	23 (9%)	0 (0%)
2501–5000	3 (1%)	0 (0%)	3 (1.1%)	0 (0%)
Smoking				
No	373 (99%)	101 (100%)	268 (99%)	4 (100%)
Yes	2 (1%)	0 (0%)	2 (1%)	0 (0%)
Alcohol				
No	331 (88%)	92 (91%)	235 (87%)	4 (100%)
Yes	44 (12%)	9 (8.9%)	35 (13%)	0 (0%)
Insunko				
No	355 (95%)	95 (94%)	256 (95%)	4 (100%)
Yes	20 (5%)	6 (6%)	14 (5%)	0 (0%)
Age at sexual debut				
Median (IQR)	17 (16–19)	17 (15–19)	17 (16–19)	17 (16–18)
Missing	3	0	3	0
Contraception use				
No	231 (62%)	66 (65%)	161 (60%)	4 (100%)
Yes	144 (38%)	35 (35%)	109 (40%)	0 (0%)
Gravidity				
Median (IQR)	4 (2–5)	3 (2–5)	4 (2–5)	3 (2–6)
Parity				
Median (IQR)	3 (2–5)	3 (2–4)	4 (2–5)	3 (2–6)
On ART				
No	1 (0.3%)	0 (0%)	1 (0.4%)	0 (0%)

Table 1 Continued

Patient characteristics	Overall, n=375	CIN2+, n=101	<CIN2, n=270	Not tested/invalid, n=4
Yes	374 (100%)	101 (100%)	269 (100%)	4 (100%)
CD4 cell count (cells/mm3)				
Median (IQR)	542 (418–759)	532 (406–742)	543 (421–763)	1111 (928–1123)
Missing	2	0	1	1
HIV RNA load (copies/mL)				
<50	316 (84%)	79 (78%)	234 (87%)	3 (100%)
50–1000	29 (8%)	10 (10%)	19 (7%)	0 (0%)
1001–10 000	11 (3%)	6 (6%)	5 (2%)	0 (0%)
10 001 and more	18 (5%)	6 (6%)	12 (4%)	0 (0%)
Missing	1	0	0	1
History of previous cervical screening				
No	267 (71%)	71 (70%)	194 (72%)	2 (50%)
Yes	108 (29%)	30 (30%)	76 (28%)	2 (50%)
History of treatment for precancer				
No	368 (98%)	99 (98%)	265 (98%)	4 (100%)
Yes	7 (2%)	2 (2%)	5 (2%)	0 (0%)

Data are n=number of participants (%).
*Income in 3 months before enrolment.
ART, antiretroviral therapy; CD4, cluster of differentiation 4; cells/mm3, cell per cubic millimetre; CIN2+, cervical intraepithelial neoplasia grade 2 and above; <CIN2, cervical intraepithelial neoplasia grade 1 and below; copies/mL, copies per millilitre; hrHPV, high-risk human papillomavirus; VIA, visual inspection of the uterine cervix after application of 3%–5% acetic.

for Gynocular was 0.69 (95% CI 0.63 to 0.75) (online supplemental appendix S10). When dichotomised using the Youden index (Swede score 3), the test had a sensitivity of 51.5% (95% CI 41.9% to 61.0%), a specificity of 80.0% (95% CI 74.8% to 84.3%) and DOR of 4.25 (95% CI 2.6 to 6.9, figure 2, online supplemental appendix S9). When using the Swede score 1 (threshold yielding sensitivity≥90%), we reached a sensitivity of 97.0% (95% CI 92.0% to 99.0%) with a specificity of 3.3% (95% CI 1.8% to 6.2%). When using the Swede score 6 (threshold

yielding specificity≥90%), specificity reached 94.1% (95% CI 90.6% to 96.3%) with a sensitivity of 29.7% (95% CI 21.7% to 39.2%). VIA had the lowest sensitivity (22.8%, 95% CI 15.7% to 31.9%) and highest specificity (92.6%, 95% CI 88.8% to 95.2%), with a DOR of 3.7 (95% CI 1.9 to 7.1).

Sensitivity analyses
We did not detect a strong training effect (online supplemental appendix S11a). Test accuracy measures were

Table 2 Tests results and CIN status

Screening test Gynocular	VIA	hrHPV	No Neoplasia n=128	CIN1 n=141	CIN2 n=44	CIN3 n=56	Cancer n=1
neg	neg	neg	82 (64%)	62 (44%)	18 (41%)	5 (8.9%)	0 (0%)
pos	neg	neg	9 (7%)	14 (10%)	5 (11%)	4 (7%)	0 (0%)
neg	pos	neg	3 (2%)	2 (1%)	1 (2%)	0 (0%)	0 (0%)
neg	neg	pos	28 (22%)	35 (25%)	9 (20%)	16 (29%)	0 (0%)
pos	pos	neg	2 (2%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)
neg	pos	pos	0 (0%)	2 (1%)	0 (0%)	0 (0%)	0 (0%)
pos	neg	pos	2 (2%)	16 (11%)	5 (11%)	16 (29%)	0 (0%)
pos	pos	pos	2 (2%)	8 (6%)	6 (14%)	15 (27%)	1 (100%)

Women with three valid tests are included in this table, total=370. Data are n=number of participants; (%).
CIN1, cervical intraepithelial neoplasia grade 1; CIN2, cervical intraepithelial neoplasia grade 2; CIN3, cervical intraepithelial neoplasia grade 3; hrHPV, high-risk human papillomavirus; neg, test negative; pos, test positive; VIA, visual inspection of the uterine cervix after application of 3%–5% acetic.
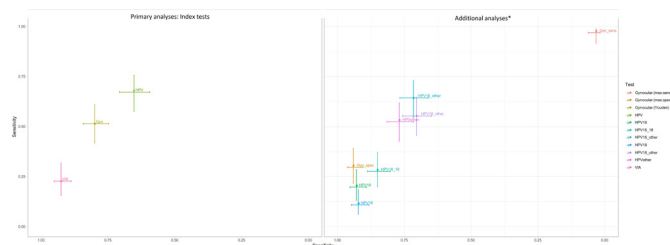
Figure 2 Sensitivity and specificity of single test screening strategies for prevalent CIN2+. *Secondary analysis (Gynocular) sensitivity analysis (HPV subtypes). Gyn, Gynocular; Max.spec, using a threshold that maximises specificity; Max.sens, using a threshold that maximises sensitivity; HPV16, human papillomavirus subtype 16; HPV18, human papillomavirus subtypes 18 and 45; HPVother, human papillomavirus other high-risk subtypes pooled –31, 33, 35, 39, 51, 52, 56, 58, 59, 66 and 68; hrHPV, high risk human papillomavirus; VIA, visual inspection of the uterine cervix after application of 3%–5% acetic.

similar whether or not participants stopped the study because of the COVID-19 pandemic (online supplemental appendix S11b), and results replacing missing and indeterminate test results and reference standards did not substantially affect estimates of accuracy (online supplemental appendix S11c). Using different categories of HPV subtypes showed similar results, with the best combination being HPV16 with 'other' (sensitivity 64.4%, 95% CI 54.6% to 73.0%, specificity 71.6%, 95% CI 66.0% to 76.7% and DOR 4.6, 95% CI 2.8% to 7.4%, table 2). In the hypothetical scenario where biopsies were taken only from visible lesions (n=106), sensitivity increased for all tests and specificity remained at similar levels (online supplemental appendix S11d). For hrHPV, sensitivity was 85.7% (95% CI 73.3% to 92.9%), specificity was 62.7% (95% CI 57.3% to 67.8%), and DOR was 10.1 (95% CI 4.4 to 23.2). Sensitivity of Gynocular increased to 93.9% (95% CI 83.5% to 97.9%), specificity to 81.5% (95% CI 76.9% to 85.3%) and DOR to 67.5 (95% CI 20.3 to 22.4). The sensitivity for VIA increased to 44.9% (95% CI 31.9% to 58.7%), specificity to 93.5% (95% CI 90.3% to 95.7%) and DOR to 11.8 (95% CI 5.75 to 24.1).

Two tests in combination

When we examined combinations of two tests with positive results, we found the specificities improved to above 90% for all test combinations but found a higher proportion of false negatives than when using single screening tests. Among combinations of tests, hrHPV followed by Gynocular yielded the most favourable balance of sensitivity 42.6% (95% CI 33.4% to 52.3%) and specificity 90.0% (95% CI 85.3% to 92.7%) (table 3, online supplemental appendix S12). Other analyses of test combinations are reported in online supplemental appendix S13.

Subgroup analyses

In a subgroup analysis, we found no clear differences in sensitivity and specificity according to age, menopause, education, contraception, parity, *T. vaginalis* result, ART

status, HIV RNA viral load, CD4 cell count and previous treatment for precancerous disease (online supplemental appendix S14), and we did not detect effect modification by patient characteristics on the association between diagnostic test and disease status (online supplemental appendix S15).

DISCUSSION

We found a high prevalence of CIN2+precancerous lesions and hrHPV in WLHIV, almost all of whom were on ART. Stand-alone hrHPV, Gynocular and VIA testing missed almost a quarter of precancerous disease. Among visual screening tests, the Gynocular performed better than VIA. Combining tests did not improve test accuracy measures. In a sensitivity analysis in which only CIN2+ detected from visible lesions was used as the reference standard, all accuracy measures improved.

The study has several strengths. First, we tested a novel magnification device (Gynocular) among WLHIV with limited access to conventional colposcopy. Second, the index tests and reference standards were relevant to the context and performed by local experts. Third, we optimised the study methods with several strategies. Local and international experts contributed to protocol development and training staff. A data safety and monitoring board provided oversight.[9] All women received the reference standard, preventing partial verification biases. We reduced detection bias by obtaining 2–4 biopsies from each woman and considering the presence of disease at two time points 6 months apart. We used objective measures of HIV severity and concurrent *T. vaginalis* to examine associations between coexisting conditions and test performance.[15 16] We safeguarded blinding of screening tests and the reference standard. Furthermore, p16 immunostaining was used to determine HSIL objectively.[11 17] Because screening results often include indeterminate and missing results, we included a sensitivity analysis to understand the impact of these on test accuracy.

We acknowledge the limitations of our study methods. First, we used an index test (Gynocular) to guide biopsy samples for the reference standard. However, partial verification bias was avoided because all women received multiple biopsies irrespective of whether a lesion was seen.[18 19] Second, the COVID-19 pandemic interrupted follow-up, and only 104 (28%) women had a second reference test by the time the study had to close. We found five additional cases of CIN2+ among these, presumably missed at baseline. Were we able to complete follow-up on all women, disease prevalence may have been higher, affecting the predictive values of the tests performed at baseline.[20] Third, while we considered 6 months a short enough interval for the second reference standard test to detect missed disease, a 12-week time frame has also been used in previous studies.[7] Fourth, we used GeneXpert as the hrHPV testing platform, but an additional laboratory-based method would have enhanced quality control.

Table 3 Diagnostic accuracy of tests in combination and their precision

Test 1	Test 2	N	tp	fp	tn	fn	Sensitivity 95% CI	Specificity 95% CI	PPV 95% CI	NPV 95% CI	PLR 95% CI	NLR 95% CI
Gynocular and VIA												
Gynocular Youden	VIA	371	22	13	257	79	0.22 0.15 to 0.31	0.95 0.92 to 0.97	0.63 0.46 to 0.77	0.77 0.72 to 0.81	0.77 0.72 to 0.81	4.52 2.37 to 8.64
Gynocular max. sens	VIA	371	23	20	250	78	0.23 0.16 to 0.32	0.93 0.89 to 0.95	0.54 0.39 to 0.68	0.76 0.71 to 0.81	0.76 0.71 to 0.81	3.07 1.77 to 5.35
Gynocular max. spec	VIA	371	19	6	264	82	0.19 0.12 to 0.28	0.98 0.95 to 0.99	0.76 0.57 to 0.89	0.76 0.72 to 0.81	8.47 3.48 to 20.59	0.83 0.76 to 0.91
HPV and Gynocular												
HPV	Gynocular Youden	369	43	28	240	58	0.43 0.33 to 0.52	0.90 0.85 to 0.93	0.61 0.49 to 0.71	0.81 0.76 to 0.85	4.08 2.69 to 6.19	0.64 0.54 to 0.76
HPV16*	Gynocular Youden	369	17	5	263	84	0.17 0.11 to 0.25	0.98 0.96 to 0.99	0.77 0.57 to 0.90	0.76 0.71 to 0.8	9.02 3.41 to 23.81	0.85 0.78 to 0.93
HPV18/45*	Gynocular Youden	369	8	8	260	93	0.08 0.04 to 0.15	0.97 0.94 to 0.99	0.50 0.28 to 0.72	0.74 0.69 to 0.78	2.65 1.02 to 6.88	0.95 0.89 to 1.01
HPVother*	Gynocular Youden	369	31	18	250	70	0.30 0.23 to 0.40	0.93 0.90 to 0.96	0.63 0.49 to 0.75	0.78 0.73 to 0.82	4.57 2.68 to 7.79	0.74 0.65 to 0.85
HPV16 HPV18/45*	Gynocular Youden	369	22	13	255	79	0.22 0.15 to 0.31	0.95 0.92 to 0.97	0.63 0.46 to 0.77	0.76 0.72 to 0.81	4.49 2.35 to 8.57	0.82 0.74 to 0.91
HPV16 HPVother*	Gynocular Youden	369	41	22	246	60	0.41 0.32 to 0.50	0.92 0.88 to 0.95	0.65 0.53 to 0.76	0.80 0.76 to 0.85	4.95 3.11 to 7.87	0.65 0.55 to 0.76
HPV18/45† HPVother*	Gynocular Youden	369	33	24	244	68	0.33 0.24 to 0.42	0.91 0.87 to 0.94	0.58 0.45 to 0.70	0.78 0.73 to 0.82	3.65 2.27 to 5.86	0.734 0.64 to 0.85
HPV and VIA												
HPV	VIA	369	22	12	256	79	0.22 0.15 to 0.31	0.96 0.92 to 0.97	0.65 0.48 to 0.79	0.76 0.72 to 0.81	4.87 2.50 to 9.46	0.82 0.74 to 0.91
HPV16*	VIA	369	10	2	266	91	0.10 0.06 to 0.17	0.99 0.97 to 0.10	0.83 0.55 to 0.95	0.75 0.70 to 0.79	13.27 2.96 to 59.51	0.92 0.85 to 0.97
HPV18/45*	VIA	369	5	1	267	96	0.05 0.02 to 0.11	0.10 0.98 to 0.10	0.83 0.44 to 0.97	0.74 0.69 to 0.78	13.27 1.57 to 112.18	0.95 0.91 to 0.10
HPVother*	VIA	369	13	10	258	88	0.13 0.08 to 0.21	0.96 0.93 to 0.98	0.57 0.37 to 0.74	0.75 0.70 to 0.79	3.45 1.56 to 7.62	0.91 0.84 to 0.98
HPV16 HPV18/45*	VIA	369	13	3	265	88	0.13 0.08 to 0.21	0.99 0.97 to 0.10	0.81 0.57 to 0.93	0.75 0.70 to 0.79	11.50 3.35 to 39.51	0.88 0.82 to 0.95
HPV16 HPVother*	VIA	369	20	11	257	81	0.20 0.13 to 0.29	0.96 0.93 to 0.98	0.65 0.47 to 0.79	0.76 0.71 to 0.80	4.82 2.40 to 9.71	0.84 0.76 to 0.92

Continued

Table 3 Continued

Test 1	Test 2	N	tp	fp	tn	fn	Sensitivity 95% CI	Specificity 95% CI	PPV 95% CI	NPV 95% CI	PLR 95% CI	NLR 95% CI
HPV18/45 HPVother*	VIA	369	15	11	257	86	0.15 0.09 to 0.23	0.96 0.93 to 0.98	0.58 0.39 to 0.75	0.75 0.71 to 0.79	3.62 1.72 to 7.61	0.89 0.82 to 0.97

Data are n=number of participants.

Test accuracies are reported with the point estimate and the 95% CIs in italics.

*Sensitivity analysis.

†n missing HPV tests and missing histopathology.

AUC, area under the curve; DOR, diagnostic OR; fn, false negative; fp, false positive; HPV16, human papillomavirus subtype 16; HPV18/45, human papillomavirus subtypes 18 and 45; HPVother, human papillomavirus other high-risk subtypes pooled −31, 33, 35, 39, 51, 52, 56, 58, 59, 66 and 68; hrHPV, high-risk human papillomavirus; Max sens, using threshold that maximises sensitivity; Max.spec, using threshold that maximises specificity; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; Tn, true negative; Tp, true positive; VIA, visual inspection of the uterine cervix after application of 3%–5% acetic.

Fifth, the study assessed VIA, but many sites in SSA use an amended method, including cervicography.[21 22] The results of this study are, therefore, not applicable to the Cervical Cancer Prevention Programme in Zambia.

The sensitivity of testing for hrHPV was lower in our study than in many others.[6 7] In contrast to many previous studies, we took four biopsies from women with no visible lesions and repeated testing 6 months later to avoid partial verification bias when only acetowhite lesions are sampled. We found the sensitivity of hrHPV was 65.3% (95% CI 59.4% to 70.7%) when biopsies were obtained from all women and 85.7% (95% CI 73.3% to 92.9%) if only biopsies from visible lesions were considered. Kelly *et al*'s systematic review of cervical cancer screening strategies among WLHIV in studies published up to July 2022 found that the sensitivity of VIA was overestimated in studies with a risk of partial verification bias.[7] They did not, however, do a subgroup analysis stratified by the risk of verification bias for hrHPV testing. Studies in which the reference standard is obtained only from visible lesions during colposcopy[22 23] have higher estimates of sensitivity and specificity than when all women have biopsies.[6 24 25] We also found a prevalence of precancer among WLHIV that was higher than in another Zambian study, in which CIN2+ prevalence was 16% among 200 women screened at the University Teaching Hospital in 2016.[8] A systematic review evaluating diagnostic accuracy of cervical cancer screening strategies among WLHIV found a pooled prevalence of 12% (range 2%–26%),[9] with higher prevalence in tertiary settings where referral for abnormal cervical smear or positive HPV test suggested a high risk for CIN2+. Our reference standard methods, taking 2–4 biopsies at two time points, might have detected more CIN2+ cases than in studies taking one biopsy from the most severe cervical lesion[26 27] or a maximum of two biopsies.[6 7] Wentzensen *et al*[20] found that sensitivities for detecting CIN2+ increased from 61% (95% CI 55% to 67%) in a single biopsy to 86% (95% CI 80% to 90%) with two biopsies to 96% (95% CI 91% to 99%) with three biopsies.[27] In contrast to previous studies that calculated combined test accuracy using the denominator of women testing positive from the first test, we considered all women in our denominator so as not to miss any disease in the target population. This better emulates a real-life situation highlighting that combining tests does not improve accuracy when the sensitivity of the primary screening test is low. In different contexts, the choice of screening tests that prioritise sensitivity or specificity may vary depending on the resources and infrastructure available.[28 29] For example, if there is already a system in place to ensure women receive timely follow-up and treatment, providers can prioritise a test with lower sensitivity and higher specificity, to avoid unnecessary treatments. However, if this infrastructure is not available, a test that prioritises high sensitivity and enables a point-of-care strategy to link screening and treatment, may be preferred to ensure that fewer women with the potential to develop cervical cancer are missed. Ideally, a screening sequence should aim for a sensitivity

of 90%–95% and specificity of 85% to detect CIN3+ during one screening interval.[30] Although p16-positivity indicates a higher cancer potential than CIN2+alone, our results cannot be directly compared with this target. Larger test accuracy studies among WLHIV which minimise bias, would strengthen estimates of accuracy and enable improve the healthcare for women.

In our study, hrHPV testing, Gynocular colposcopy and VIA performed poorly as stand-alone screening tests among WLHIV, and 22.9% of cases were not detected by any test. Combining two tests did improve specificity but not overall accuracy when all women (and all disease) were considered in the denominator. Our findings have implications for research and cervical cancer screening policies among WLHIV if test accuracy in this high-risk population has been overestimated. According to our sensitivity analysis, the assumption that taking biopsies from visible lesions on colposcopy is an acceptable reference standard might need reassessment. WHO recommends 3–5 years screening intervals for WLHIV, based on the assumption that suboptimal screening tests at sufficiently frequent intervals will still prevent cancer because of the long precancerous phase. However, if accuracy measures informing modelling studies are overestimated, these screening intervals might be too long. Larger studies, among WLHIV, in countries with the highest disease burden and using methods that reduce verification bias are urgently required. Our robust descriptive study results can be used in future modelling studies and randomised controlled trials of screening effectiveness, both of which are needed to determine improved strategies for cervical cancer screening among WLHIV.

Author affiliations
[1]Institute of Social and Preventative Medicine, University of Bern, Bern, Switzerland
[2]Center for Infectious Disease Research in Zambia, Lusaka, Zambia
[3]University of Zambia University Teaching Hospital, Lusaka, Zambia
[4]Levy Mwanawasa Medical University Hospital, Lusaka, Zambia
[5]Clinical Trials Unit, University of Bern, Bern, Switzerland
[6]Department of histopathology, Lancet Laboratories, Richmond, South Africa
[7]International Agency for Research on Cancer, Lyon, France
[8]University of Alabama at Birmingham (UAB), Birmingham, Alabama, USA
[9]University of Basel, Basel, Switzerland

Twitter Katayoun Taghavi @KatiTaghavi

ORCID iD
Katayoun Taghavi http://orcid.org/0000-0003-0812-0069

REFERENCES

1 World Health Organization. *WHO guideline for screening and treatment of cervical pre-cancer lesions for cervical cancer prevention., ISBN: 9789240030824*second edition. 2021. Available: https://www.who.int/initiatives/cervical-cancer-elimination-initiative
2 Stelzle D, Tanaka LF, Lee KK, *et al*. Estimates of the global burden of cervical cancer associated with HIV. *Lancet Glob Health* 2021;9:e161–9.
3 Castellsagué X, Díaz M, de Sanjosé S, *et al*. Worldwide human papillomavirus etiology of cervical adenocarcinoma and its cofactors:

implications for screening and prevention. *J Natl Cancer Inst* 2006;98:303–15.

4. Moscicki A-B, Schiffman M, Burchell A, *et al*. Updating the natural history of human papillomavirus and anogenital cancers. *Vaccine* 2012;30 Suppl 5:F24–33.

5. Morhason-Bello IO, Odedina F, Rebbeck TR, *et al*. Challenges and opportunities in cancer control in Africa: a perspective from the African Organisation for research and training in cancer. *Lancet Oncol* 2013;14:e142–51.

6. Smith SK, Nwosu O, Edwards A, *et al*. Performance of screening tools for cervical neoplasia among women in low- and middle-income countries: a systematic review and meta-analysis. *PLOS Glob Public Health* 2023;3:e0001598.

7. Kelly H, Jaafar I, Chung M, *et al*. Diagnostic accuracy of cervical cancer screening strategies for high-grade cervical intraepithelial Neoplasia (Cin2+/Cin3+) among women living with HIV: a systematic review and meta-analysis. *EClinicalMedicine* 2022;53:101645.

8. Chibwesha CJ, Frett B, Katundu K, *et al*. Clinical performance validation of 4 point-of-care cervical cancer screening tests in HIV-infected women in Zambia. *J Low Genit Tract Dis* 2016;20:218–23.

9. Taghavi K, Moono M, Mwanahamuntu M, *et al*. Screening test-accuracy to improve detection of precancerous lesions of the cervix in women living with HIV: a study protocol. *BMJ Open* 2020;10:e037955.

10. Darragh TM, Colgan TJ, Cox JT, *et al*. The lower anogenital squamous terminology standardization project for HPV-associated lesions: background and consensus recommendations from the college of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *J Low Genit Tract Dis* 2012;16:205–42.

11. Ozaki S, Zen Y, Inoue M. Biomarker expression in cervical intraepithelial neoplasia: potential progression predictive factors for low-grade lesions. *Hum Pathol* 2011;42:1007–12.

12. Colposcopy and treatment of Cervical intraepithelial Neoplasia: a beginners' manual. ISBN-13 (database) 978-92-832-0412-1. 2004. Available: http://screening.iarc.fr/colpochap.php?lang=1&chap=4

13. Kalubula M, Shen H, Khanam T. Assessment of carcinogenic and toxic substances in "Insunko" herb. *Toxicology Reports* 2020;7:468–74.

14. Bowring J, Strander B, Young M, *et al*. The Swede score. *J Low Genit Tract Dis* 2010;14:301–5.

15. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926–30.

16. Whiting PF, Rutjes AWS, Westwood ME, *et al*. A systematic review classifies sources of bias and variation in diagnostic test-accuracy studies. *J Clin Epidemiol* 2013;66:1093–104.

17. McGrath CJ, Garcia R, Trinh TT, *et al*. Role of P16 testing in cervical cancer screening among HIV-infected women. *PLoS One* 2017;12:e0185597.

18. Leeflang MMG, Rutjes AWS, Reitsma JB, *et al*. Variation of a test's sensitivity and specificity with disease prevalence. *CMAJ* 2013;185:E537–44.

19. Huh WK, Sideri M, Stoler M, *et al*. Relevance of random biopsy at the transformation zone when colposcopy is negative. *Obstet Gynecol* 2014;124:670–8.

20. Wentzensen N, Walker JL, Gold MA, *et al*. Multiple biopsies and detection of Cervical cancer precursors at Colposcopy. *J Clin Oncol* 2015;33:83–9.

21. World Health Organisation. WHO Prequalification of in vitro diagnostics: public report. Pqdx 0268-070-00 WHO PQ public report. December 2017/version 3.0. 2017. Available: https://www.who.int/diagnostics_laboratory/evaluations/pq-list/hiv-vrl/171221_final_pq_report_pqdx_0268_070_00.pdf?ua=1

22. Kuhn L, Saidu R, Boa R, *et al*. Clinical evaluation of modifications to a human papillomavirus assay to optimise its utility for cervical cancer screening in low-resource settings: a diagnostic accuracy study. *Lancet Glob Health* 2020;8:e296–304.

23. Firnhaber C, Mayisela N, Mao L, *et al*. Validation of cervical cancer screening methods in HIV positive women from Johannesburg South Africa. *PLoS One* 2013;8:e53494.

24. Kremer WW, van Zummeren M, Breytenbach E, *et al*. The use of molecular markers for cervical screening of women living with HIV in South Africa. *AIDS* 2019;33:2035–42.

25. Chung MH, McKenzie KP, De Vuyst H, *et al*. Comparing papanicolau smear, visual inspection with acetic acid and human papillomavirus cervical cancer screening methods among HIV-positive women by immune status and antiretroviral therapy. *AIDS* 2013;27:2909–19.

26. Wentzensen N, Walker J, Smith K, *et al*. A prospective study of risk-based colposcopy demonstrates improved detection of cervical precancers. *Am J Obstet Gynecol* 2018;218:604.

27. Baasland I, Hagen B, Vogt C, *et al*. Colposcopy and additive diagnostic value of biopsies from colposcopy-negative areas to detect cervical dysplasia. *Acta Obstet Gynecol Scand* 2016;95:1258–63.

28. Parham GP, Mwanahamuntu MH, Kapambwe S, *et al*. Population-level scale-up of cervical cancer prevention services in a low-resource setting: development, implementation, and evaluation of the cervical cancer prevention program in Zambia. *PLoS One* 2015;10:e0122169.

29. Mwanahamuntu MH, Sahasrabuddhe VV, Blevins M, *et al*. Monitoring the performance of "screen-and-treat" cervical cancer prevention programs. *Int J Gynaecol Obstet* 2014;126:88–9.

30. Castle PE. The potential utility of HPV genotyping in screening and clinical management. *J Natl Compr Canc Netw* 2008;6:83–95.