HPValidate:
clinical validation of hrHPV test system using self-collected vaginal samples in NHS England commissioned laboratories providing cervical screening services

Core reporting group:

Christopher Mathews, Adam Brentnall, Matejka Rebolj
Wolfson Institute of Population Health, Queen Mary University of London, London

Alexandra Sargent
Cytology Department, Manchester Royal Infirmary, Manchester University NHS Foundation Trust,
Manchester

Kate Cuschieri
Scottish HPV Reference Laboratory, Royal Infirmary of Edinburgh, NHS Lothian, Edinburgh

Karin Denton
Severn Pathology, Southmead Hospital, North Bristol NHS Trust, Bristol

September 2024

This work was commissioned and funded by Public Health England, Department of Health and Social Care (Ref: PHE – Corporate - LOT 1 ITT4424 - Validation of hrHPV Test System Using Self-Collected Vaginal Samples)

Other contributors:

Anne Mackie, Charlotte Mills, Carolina Martinelli, Ann-Marie Wright
Primary Care and Prevention, Department of Health and Social Care, London, UK

Katherine Hunt, Andrew Bird
Severn Pathology, Southmead Hospital, North Bristol NHS Trust, Bristol, UK

Hasit Patel, David Smith
Health Service Laboratories LLP, London, UK

Trudy Johnson, Kay Ellis
NEY Cervical Screening, The NHS Pathology Centre, Queen Elizabeth Hospital, Gateshead Health NHS Foundational Trust, Gateshead, UK

Mark Hunt
Cytopathology Department, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK

Table of Contents

EXECUTIVE SUMMARY **4**

BACKGROUND **6**

RATIONALE FOR THE STUDY **6**

METHODS **6**

DATA REPORT: SUMMARY OF THE FINDINGS **11**

1. Study completion *11*

2. Baseline characteristics of the women whose samples were included in the study *11*
Colposcopy populations (women with CIN2+) 13
Primary care populations (women without CIN2+) 14

3. Primary and exploratory analyses *15*
Relative test sensitivity 16
Relative test specificity 17
Sample invalidity 18
Summary 19

RECOMMENDATIONS FOR THE IMPLEMENTATION IN NHS ENGLAND **20**

EXECUTIVE SUMMARY

NHS England is considering offering human papillomavirus (HPV) testing on self-collected samples for the purpose of primary cervical screening. There are insufficient data to support manufacturer claims for accuracy of self-collection devices with HPV assays currently used in the English programme, and behavioural aspects are poorly studied. To address this, an ethics-approved study (HPValidate) was conducted within NHS England in 2021-2023. Results provide insight into the epidemiological and behavioural aspects of self-collection of cervical screening samples for routine use in the programme.

The main aim of the epidemiological evaluation of the data collected within the study was to assess the performance of various options for self-sampling, comprising the sample collection device and HPV test platform. Pathways assessed by this study may be taken forward for further population evaluation, prior to potential national implementation. This would support laboratories to deliver testing of appropriate quality for service work.

The study evaluated five self-collection workflows. These were combinations of three different self-sampling kits:

- Copan Italia S.p.A's Self Vaginal FLOQswabs,
- Rovers Medical Devices' Evalyn brush, and
- Hologic's Multitest Swab Specimen Collection Kit;

and the two HPV assays used in the programme at present:

- Roche's cobas 6800/8800, and
- Hologic's APTIMA Panther.

Both HPV assays and their instrumentation had been previously validated for use on clinician-collected samples in England.[1]

Recruitment was severely impacted by capacity issues related to the COVID-19 pandemic, but adequate numbers of women were recruited to meet the statistical requirements. No significant practical issues were reported by laboratories. However, the number of specimens processed per week was low, and manual preparation methods which were satisfactory for HPValidate are unlikely to be manageable or practical when scaled up to larger numbers.

The results of the study indicate that four self-collection workflows meet the validation requirements, and may be considered for future evaluation in the context of screening in England:

- Evalyn + cobas,
- FLOQswabs + cobas,
- FLOQswabs + APTIMA, and
- Multitest + APTIMA.

The epidemiological characteristics of women differed between the five workflows. Consequently, the results of HPValidate cannot be used to directly compare the accuracy of the four workflows which met requirements for further evaluation.

The HPValidate study team make the following recommendations to mitigate remaining risks from offering self-sampling in the programme:

[1] NHS England. Guidance. Cervical screening: acceptable HPV tests (Updated 24 November 2023). URL: https://www.gov.uk/government/publications/cervical-screening-acceptable-hpv-tests/cervical-screening-acceptable-hpv-tests. Last accessed: 12 September 2024.

1. Two to four workflows should undergo further evaluation and include both cobas and APTIMA platforms. This will mitigate future risk related to use of a single technology (assay, sampling device) in the programme.
2. HPValidate did not consider test accuracy, or behavioural aspects associated with testing at home. Evaluation of sensitivity is required using at-home samples at screening. This will help quantify the risk of an increased rate of interval cancers in women who choose self-sampling. The validation of sensitivity in HPValidate used a study design where samples from women with the target disease endpoint (CIN2+) were collected up to two years *after* the primary screening test, at the colposcopy appointment with a nurse present. Further, fewer cases with negative cytology were included than routinely present at primary screening. The risk of lower sensitivity in the real world than in HPValidate arises because aspects of the study design could have enriched for the inclusion of samples with higher average viral loads resulting in a more favourable estimate of sensitivity than would be achieved at routine primary screening.
3. Future evaluation should be over a wide geographical area to minimise risk of bias due to the effect of local factors.
4. Criteria for timeliness of sample receipt, processing and appropriate reporting policies for samples which have exceeded these time limits should be included. The study identified a risk to sample adequacy associated with time elapsed between collection and processing in the laboratory.
5. Future evaluation should include aspects of laboratory practice, in particular in relation to aspects of automation.

BACKGROUND

In England, women are invited for cervical screening between the ages of 25 and 64 years, whereby the first invitation is mailed six months before the 25th birthday, at the age of 24.5 years.

Testing for human papillomavirus (HPV) on clinician-collected liquid-based cytology (LBC) samples is an established method for primary cervical screening. The NHS England cervical screening programme has published a list of HPV tests which have been accepted for use in the programme for primary HPV screening and HPV triage of borderline and low-grade abnormalities, and as a 'test of cure' of treatment according to national protocols.[2] Two HPV tests from this list are currently in use in the English programme, Roche cobas 6800/8800 and Hologic APTIMA Panther.

HPV testing on self-collected samples, as an alternative to LBC clinician-collected samples, has the potential to increase screening uptake and offer women more choice between acceptable screening options. None of the current HPV tests have obtained full regulatory approval for use on a sample taken by a self-sampling device, though a handful of device and test combinations have obtained CE marking. A small number of self-sampling devices are available that have a generic claim/CE mark for use in the home environment for downstream HPV testing.

The UK National Screening Committee has recommended that self-sampling requires further study in well organised pilots and research projects.

RATIONALE FOR THE STUDY

The primary aim of the HPValidate study was to compare the sensitivity and specificity of HPV testing for the detection of high-grade cervical intraepithelial neoplasia (CIN2+) when undertaken on self-collected vs. clinician-collected samples in the English NHS.

This report includes an analysis of the data evaluating the epidemiological aspects of self-collection (LOT 1 ITT4424).[3]

METHODS

The five combinations of self-sampling device and HPV test ("workflows") that were evaluated in HPValidate are presented in Table 1.

[2] NHS England. Guidance. Cervical screening: acceptable HPV tests (Updated 24 November 2023). URL: https://www.gov.uk/government/publications/cervical-screening-acceptable-hpv-tests/cervical-screening-acceptable-hpv-tests. Last accessed: 12 September 2024.

[3] The report evaluating behavioural aspects of self-collection (LOT 2 ITT4424) has been submitted separately by Dr Laura Marlow and Prof Jo Waller.

Table 1. The five self-sampling workflows evaluated in the HPValidate study.

Self-sampling device	HPV test	Comparator HPV test on clinician-collected samples	Laboratory with the respective catchment areas
Evalyn	cobas	cobas	Gateshead NHS Foundation Trust and Norfolk and Norwich University Hospitals NHS Foundation Trust
Evalyn	APTIMA	APTIMA	North Bristol NHS Trust
FLOQSwabs	cobas	cobas	Manchester University NHS Foundation Trust
FLOQSwabs	APTIMA	APTIMA	Health Services Laboratories, London
Multitest	APTIMA	APTIMA	Health Services Laboratories, London and North Bristol NHS Trust

The APTIMA platform reports detection of HPV mRNA for 14 high-risk genotypes in combination, without automatic reporting of individual genotypes. The strength of the signal for HPV detection is available from system logs as the signal-to-cut-off (s/co) ratio. APTIMA also includes a control to adjudicate amplification success. The Roche cobas platform automatically reports detection of HPV DNA on three channels (HPV16, HPV18, and a channel for the remaining 12 non-16/18 high-risk HPV genotypes in combination). The cobas assay also amplifies human beta-globin, which is used as an indication of material sufficiency in the sample (internal control). A readout of individual cycle threshold (CT) values for each channel including the internal control may also be obtained from the testing system logs.

Recruitment into the study

Self-collection took place at the clinic, prior to the clinician appointment. For women recruited at colposcopy clinics, this included a colposcopy to investigate abnormalities detected in primary screening or early recall samples. Here, the referral clinician-collected sample was considered “linked” to the self-collected sample. “Linked” clinician-collected samples were taken up to 56 days prior to self-collection. For women recruited at screening in primary care, the appointment included a LBC sample collected by a clinician from the cervix; these clinician-collected samples were “paired” with the self-collected samples. “Paired” clinician-collected samples were taken on the same day as self-collection. All clinician-led examinations followed the relevant national and/or local guidance and were not affected by the study. The study recommended no additional examinations for women with HPV-negative clinician-collected samples whose self-collected samples returned a positive HPV test result.

Colposcopy clinic inclusion criteria were: women referred to colposcopy, regardless of the degree of cytological abnormality, for further investigation after screening results showing persistent HPV positive/cytology negative, HPV positive/cytology abnormal (≥borderline), or inadequate HPV test/cytology; the referral LBC sample must have been taken at most 8 weeks (≤56 days) before women attended colposcopy. Women were excluded if they were unable to consent themselves into the study; were referred to colposcopy after primary care had not been able to obtain a sample; their reason for referral to colposcopy was a negative HPV test result; or attended for test of cure or continued management.

In primary care, the inclusion criteria were: women who were due for their routine cervical screening test and would have already received their invitation letter, without a limit on how overdue they were for their appointment; any screening ages (24-64 years). Women were excluded if they were unable to consent themselves into the study; presented before their next test due date, including clinical referrals; or were on early recall due to previous HPV positive results.

In addition, women were excluded in the analysis stage if their HPV testing result sets were incomplete, i.e. when either an LBC and/or a self-sample HPV result was missing; and, for reasons explained below, only women with CIN2+ (colposcopy population) and women without CIN2+ (primary screening population) were included in the analyses.

Statistical analysis

All analyses were performed (including subgroups) according to a pre-specified statistical analysis plan.

Baseline information on women included in the primary analyses was tabulated by the type of recruited population and self-sampling workflow. Differences between self-sampling workflows were evaluated by using the Kruskal-Wallis test for continuous variables (age, time to referral) and chi-square test for categorical variables.

Conditional relative sensitivity was defined in this study as the detection, demonstrated by a positive HPV test result, of CIN2+ in the self-collected sample i.e., as (CIN2+ cases with a positive self-collected HPV test)/(Total CIN2+). In this study, all women recruited at colposcopy by definition presented with HPV-positive ("linked") clinician-collected samples. CIN2+ cases among women attending for colposcopy were defined, in line with the clinical practice within NHS England, as routine diagnoses of ungraded CIN, CIN2, CIN3, and cervical cancer. They were retrieved from the initial colposcopy reports. Two-sided 90% CI were calculated using Wilson's method for a binomial parameter, to test non-inferiority at the 1-sided α=0.05 level.

Conditional relative specificity was defined among women without CIN2+ as the number of screened women with an HPV-negative self-sampling test result compared with (i.e., divided by) the number of HPV-negative women on the paired clinician-collected LBC sample. CIN2+ cases, when recorded after direct colposcopy referral (i.e., after an HPV-positive primary screening test showing at least borderline abnormalities on cytology triage), were excluded from the analysis of the relative specificity. The study did not attempt full ascertainment of CIN2+ from this population because to do so would require at least two more years of follow-up to ascertain cases diagnosed after the two early recalls. Two-sided 90% CI was based on methods given by Hayen et al.,[4] to test non-inferiority at the 1-sided α=0.05 level.

Conditional relative sensitivity and specificity were evaluated based on pre-defined thresholds for non-inferiority, being 90% ("the higher threshold") and 75% ("the lower threshold") for test sensitivity, and 95% for test specificity. The two sensitivity thresholds were selected based on anticipation of a scenario in which one or more of the self-sampling

[4] Hayen A, Macaskill P, Irwig L, Bossuyt P. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. J Clin Epidemiol 2010;63:883-891. Available on: https://pubmed.ncbi.nlm.nih.gov/20079607/.

workflows might not meet the non-inferiority threshold for relative sensitivity of 0.90. In that case, the protocol recommended to establish whether the sensitivity of a self-sampling workflow was non-inferior to cytology. In the English HPV screening pilot, HPV testing using clinician-collected samples was 60% more sensitive for the detection of CIN2+ compared with cytology.[3] This means that the sensitivity of cytology relative to HPV testing was approximately 0.63 (=1/1.6). Allowing for uncertainty in the relative sensitivity of cytology, a secondary non-inferiority threshold was set at 0.75.

The latest raw datasets were shared with the data analysis team on the 19th of December 2023. The latest data errors were corrected on the 25th of January 2024.

Sample size considerations

For the colposcopy component of the study, the target study size was ≥60 women with CIN2+ per self-sampling workflow, diagnosed at direct colposcopy referral of women with HPV-positive/cytology-abnormal screening samples, or after the 12- or 24-month early recall of women with HPV-positive/cytology-negative screening samples (in these women, primary screening took place up to two years before recruitment into the study). Assuming independence between self- and clinician-sampling (i.e., kappa is zero), if both tests have a true absolute sensitivity of 97.5% for CIN2+ detection, then n=60 cases would provide approximately 80% power to reject inferiority of self-sampling using a 90% threshold at the one-sided 5% level. If the absolute sensitivity is 95% for both, then power is approximately 99% when testing non-inferiority at the 75% cut point. If the absolute sensitivity of clinician-sample HPV testing is 95%, and it is 90% for self-sampling (true relative sensitivity 90/95=94.7%) then n=60 would have approximately 90% power to reject inferiority of self-sampling using a threshold of 75%. Therefore, n=60 cases per self-sampling workflow was felt to provide sufficient information to decide whether to further investigate different self-sampling tests in subsequent studies, as was the main objective of HPValidate.

Per self-sampling workflow, in order to achieve at least n=60 CIN2+ cases, the study sought to recruit 350 consecutive consented women in colposcopy clinics based on the findings from the English HPV screening pilot.[5] In the pilot, 4.0% of unvaccinated women screened with HPV testing on clinician samples for the first time were referred to colposcopy directly (with a positive predictive value, PPV, for CIN2+ of 43%), 1.0% at the 12-month early recall (PPV: 37%), and 1.4% at the 24-month early recall (PPV: 21%). On average, the PPV of a colposcopy was 37%. For HPValidate, it was conservatively assumed that the PPV could be halved compared to the HPV screening pilot. This is because HPValidate was going to be undertaken while the English programme was offering screening to women vaccinated against HPV16/18[6] and because some of the women recruited into the study may have undergone previous HPV tests in areas that introduced HPV-based screening prior to the national roll-out. Hence, the study sought to recruit 350 consecutive consented women (including ≥60 CIN2+ cases, resulting in a PPV of ≥17%) attending for colposcopy indicated by routine cervical screening, per self-sampling workflow.

[5] Rebolj M, Mathews CS, Pesola F, Cuschieri K, Denton K, Kitchener H. Age-specific outcomes from the first round of HPV screening in unvaccinated women: Observational study from the English cervical screening pilot. BJOG 2022;129:1278-1288. Available on: https://pubmed.ncbi.nlm.nih.gov/34913243/.
[6] Rebolj M, Pesola F, Mathews C, Mesher D, Soldan K, Kitchener H. The impact of catch-up bivalent human papillomavirus vaccination on cervical screening outcomes: an observational study from the English HPV primary screening pilot. Br J Cancer 2022;127:278-287. Available on: https://pubmed.ncbi.nlm.nih.gov/35347326/.

For the primary care component of the study, recruitment of 1000 consecutive consented women per self-sampling workflow was considered achievable. In a well-screened population in Scotland in which the relative specificity of self-sampling (cobas 4800 testing of samples collected with PCR female swab sample packets) vs. clinician sampling, estimated as the ratio of test-negatives in women without CIN2+, was 0.98 (86% vs. 88%), relative test positivity was 1.17 (14% vs. 12%), and kappa was 0.73.[7] A meta-analysis of primary screening studies combining previously well-screened and unscreened populations estimated the test positivity ratio as 1.01 and kappa as 0.65.[8] Varying test positivity for self-sampling between 8% (assuming a lower positivity in a partially vaccinated HPValidate population) and 14% and that of clinician sampling between 8% and 12%, and kappa values between ~0.60 and ~0.75, a study size with 1000 recruited women has >90% power to reject inferiority of the relative specificity of self-sampling at the threshold level of 0.95 at the one-sided 5% significance level. The power remains at ≥80% when 800 women are recruited.

Ethics

The study was sponsored by Department of Health's Office for Health Improvement and Disparities (OHID). It was included in the NIHR register (CPMS ID: 47399). IRAS record number: 286052. Ethics approval was granted by Stanmore NHS Health Research Authority on 27 October 2020 (reference: 20/LO/1009). All women provided written informed consent.

Hologic Gen-Probe Inc, Copan Italia S.p.A, and Rovers Medical Devices B.V., donated self-sampling devices to the study. The study utilised HPV testing platforms already in use and under contract by the English HPV laboratories (Roche Holding AG and Hologic Gen-Probe Inc). The companies had no role in the design, conduct, or analysis of the study, and findings remain independent.

Look ahead

This report will be followed by one or more manuscripts, in which the highlights from the study will be reported for the wider national and international audiences. The manuscripts will be submitted for peer review and publication to scientific journals.

[7] Stanczuk GA, Currie H, Forson W, Baxter G, Lawrence J, Wilson A, Palmer T, Arbyn M, Cuschieri K. Self-sampling as the principal modality for population based cervical screening: Five-year follow-up of the PaVDaG study. Int J Cancer 2022:150;1350-1356. Available on: https://pubmed.ncbi.nlm.nih.gov/34850395/.
[8] Arbyn M, Castle PE, Schiffman M, Wentzensen N, Heckman-Stoddard B, Sahasrabuddhe VV. Meta-analysis of agreement/concordance statistics in studies comparing self- vs clinician-collected samples for HPV testing in cervical cancer screening. Int J Cancer 2022:151;308-312. Available on: https://pubmed.ncbi.nlm.nih.gov/35179777/.

DATA REPORT: SUMMARY OF THE FINDINGS

1. Study completion

HPValidate was undertaken during a challenging period for the English NHS. Although these challenges affected the study and some of the study sites could not meet the target recruitment numbers, ultimately the collected information was deemed sufficient to address the pre-specified primary study questions. The numbers of women recruited in the study whose data were suitable for analysis are reported in Table 2.

Table 2. The number of women included in the analysis of the data collected in the HPValidate study.

Self-sampling workflow	Colposcopy (with CIN2+)[a]	Primary care (without CIN2+)[b]
Evalyn + cobas	102	940
FLOQswabs + cobas	120	994
Evalyn + APTIMA	86	909
FLOQswabs + APTIMA	66	878
Multitest + APTIMA	78	1006

Abbreviations. CIN: cervical intraepithelial neoplasia.

[a] Reasons for exclusion of the recruited women from the analysis (for all five workflows combined): 2 (0.1%) with incomplete or withdrawn consent; 19 (1.2%) because of incomplete HPV test results on the self-collected samples; 2 (0.1%) because of incomplete clinician-collected HPV test results; 24 (1.5%) with the referral LBC sample taken >56 days prior or at an unknown date; 19 (1.2%) presenting for an unrelated reason; 4 (0.3%) because of other unresolved errors; and 1035 (66.5%) because their diagnosis was <CIN2.

[b] Reasons for exclusion of the recruited women from the analysis (for all five workflows combined): 10 (0.2%) with incomplete or withdrawn consent; 6 (0.1%) because they were outside of the target age range for screening; 66 (1.4%) because of incomplete HPV test results on the self-collected samples; 4 (0.1%) because of incomplete clinician-collected HPV test results; and 42 (0.9%) because they had a CIN2+ diagnosis after direct referral.

2. Baseline characteristics of the women whose samples were included in the study

For both the populations recruited in colposcopy (Table 3) or in primary care clinics (Table 4), the baseline epidemiological characteristics differed between the five workflows. This heterogeneity makes it difficult to directly compare the outcomes between the five self-sampling workflows.

Table 3. Baseline information on women with CIN2+ recruited at colposcopy clinics.

	Self-sampling workflow					
	Evalyn + cobas	**FLOQswabs + cobas**	**Evalyn + APTIMA**	**FLOQswabs+ APTIMA**	**Multitest + APTIMA**	**P**
Total	102 (100%)	120 (100%)	86 (100%)	66 (100%)	78 (100%)	
Age group (years)						
24-29	26 (25.5%)	29 (24.2%)	28 (32.6%)	24 (36.4%)	29 (37.2%)	
≥30[a]	76 (74.5%)	91 (75.8%)	58 (67.4%)	42 (63.6%)	49 (62.8%)	
Range, Md (IQR)	24-59, 34.5 (29.3-40.0)	24-67, 35.0 (30.0-40.0)	24-58, 33.5 (28.0-40.5)	24-60, 32.0 (28.0-36.0)	24-63, 31.5 (27.0-39.8)	0.12
Reason for colposcopy referral						<0.001
HPV+/cyt+ at primary screening (direct referral)	89 (87.3%)	70 (58.3%)	74 (86.0%)	66 (100%)	74 (94.9%)	
HPV+/cyt+ at 12-month early recall	5 (4.9%)	36 (30.0%)	5 (5.8%)	0 (0%)	0 (0%)	
HPV+ at 24-month early recall	8 (7.8%)	14 (11.7%)	7 (8.1%)	0 (0%)	4 (5.1%)	
Referral cytology						
Negative	7 (6.9%)	8 (6.7%)	7 (8.1%)	0 (0%)	4 (5.1%)	
Borderline or low-grade abnormal	18 (17.6%)	9 (7.5%)	40 (46.5%)	0 (0%)	19 (24.4%)	
High-grade abnormal	77 (75.5%)	102 (85.0%)	39 (45.3%)	66 (100%)	55 (70.5%)	
(Any abnormal)[b]	95 (93.1%)	111 (92.5%)	79 (91.9%)	66 (100%)	74 (94.9%)	<0.001
Inadequate	0 (0%)	1 (0.8%)	0 (0%)	0 (0%)	0 (0%)	
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Time from referral LBC sample to colposcopy appointment (days)						
Range, Md (IQR)	11-56, 44.0 (37.0-49.0)	35-56, 50.5 (48.0-54.0)	14-56, 30.0 (23.0-46.5)	21-55, 32.0 (29.0-38.0)	13-55, 36.0 (29.3-43.5)	<0.001
HPV test result on the linked clinician sample						
Positive	102 (100%)	120 (100%)	86 (100%)	66 (100%)	78 (100%)	
HPV test result on the self-collected sample						
Negative	10 (9.8%)	3 (2.5%)	20 (23.3%)	5 (7.6%)	5 (6.4%)	
Positive	92 (90.2%)	113 (94.2%)	66 (76.7%)	61 (92.4%)	72 (92.3%)	<0.001
Invalid	0 (0%)	4 (3.3%)	0 (0%)	0 (0%)	1 (1.3%)	

Abbreviations. cyt+: abnormal cytology (≥borderline). HPV+: a positive (high-risk) human papillomavirus test. IQR: interquartile range. Md: median.

[a] May include a small number of women older than 64 years, particularly if they were referred to colposcopy after an early recall.

[b] Sum of borderline, low-grade, and high-grade abnormal cases.

Table 4. Baseline information on women without CIN2+ recruited at primary care.

	Self-sampling workflow					
	Evalyn + cobas	**FLOQswabs + cobas**	**Evalyn + APTIMA**	**FLOQswabs + APTIMA**	**Multitest + APTIMA**	**P**
Total	940 (100%)	994 (100%)	909 (100%)	878 (100%)	1006 (100%)	
Age group (years)						
24-29	142 (15.1%)	179 (18.0%)	117 (12.9%)	240 (27.3%)	222 (22.1%)	
30-49	539 (57.3%)	605 (60.9%)	577 (63.9%)	508 (57.9%)	627 (62.3%)	
50-64	259 (27.6%)	210 (21.1%)	215 (23.7%)	130 (14.8%)	157 (15.6%)	
Range, Md (IQR)	24-64, 41.0 (33.0-50.0)	24-64, 39.0 (32.0-48.0)	24-64, 41.0 (33.0-49.0)	24-64, 35.0 (29.0-45.0)	24-64, 37.0 (30.0-45.0)	<0.001
Screening history in last 3/5 years						<0.001
No abnormality	929 (98.8%)	974 (98.0%)	880 (96.8%)	795 (90.5%)	905 (90.0%)	
Retest after an inadequate	1 (0.1%)	10 (1.0%)	5 (0.6%)	2 (0.2%)	9 (0.9%)	
Cytological or histological abnormalities	10 (1.1%)	10 (1.0%)	18 (2.0%)	81 (9.2%)	90 (8.9%)	
Unknown	0 (0%)	0 (0%)	6 (0.7%)	0 (0%)	2 (0.2%)	
HPV test result on the paired clinician sample						
Negative	828 (88.1%)	912 (91.8%)	811 (89.2%)	770 (87.7%)	904 (89.9%)	
Positive	112 (11.9%)	82 (8.2%)	93 (10.2%)	108 (12.3%)	101 (10.0%)	0.030
Invalid	0 (0%)	0 (0%)	5 (0.6%)	0 (0%)	1 (0.1%)	
HPV test result on the self-collected sample						
Negative	796 (84.7%)	833 (83.8%)	801 (88.1%)	687 (78.2%)	784 (77.9%)	
Positive	135 (14.4%)	105 (10.6%)	107 (11.8%)	191 (21.8%)	221 (22.0%)	<0.001
Invalid	9 (1.0%)	56 (5.6%)	1 (0.1%)	0 (0%)	1 (0.1%)	
Cytology outcome on the paired HPV-positive clinician sample						
HPV-positive	112 (100%)	82 (100%)	93 (100%)	108 (100%)	101 (100%)	
Negative	89 (79.5%)	58 (70.7%)	64 (68.8%)	56 (51.9%)	53 (52.5%)	
Abnormal (≥borderline)	21 (18.8%)	19 (23.2%)	29 (31.2%)	45 (41.7%)	42 (41.6%)	<0.001
Inadequate	2 (1.8%)	5 (6.1%)	0 (0%)	7 (6.5%)	5 (5.0%)	
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.0%)	

Abbreviations. HPV: human papillomavirus. IQR: interquartile range. Md: median.

Colposcopy populations (women with CIN2+)

Age distributions were similar across the five self-sampling workflows (Table 3), but the following differences were observed:

- The median time between the referral sample and the colposcopy appointment (during which the self-collected sample was provided) varied between 30 and 51 days. While study recruitment was ongoing, some of the laboratories and colposcopy clinics were experiencing significant backlogs in the delivery of their services. In some cases, this made women attending their routine colposcopy appointments ineligible for the study if the appointment was booked for more than 8 weeks (56 days) after the referral sample. Because the national guidance requires that women with high-grade abnormal cytology are seen in colposcopy within two weeks and those with

less severely abnormal cytology within six weeks,[9] some of the workflows showed a higher proportion of recruited women with high grade abnormalities.
- The proportions of women with any grade of abnormal cytology on the linked referral sample varied between 92% and 100%. For reasons explained above, however, there were larger differences between the workflows at the level of cytological grade, so that the proportions with high-grade cytological abnormalities varied between 45% and 100%, while the proportions with borderline or low-grade abnormalities varied between 0% and 47%.
- Related to the above, most women were recruited after an HPV-positive/cytology-abnormal primary screening sample (direct referral), but this varied between 58% and 100% of cases by workflow.

Primary care populations (women without CIN2+)

The following differences between the primary care populations recruited into the study were observed between the five workflows (Table 4):
- The median ages varied between 35 and 41 years.
- Almost all, >97%, of the recruited women in three workflows (Evalyn + cobas, FLOQswabs + cobas, and Evalyn + APTIMA) had no recent abnormalities. In the remaining two workflows (FLOQswabs + APTIMA and Multitest + APTIMA, both recruiting in the London catchment areas), laboratory records revealed that around 9% of the recruited women had cytological or histological abnormalities in the last 3 or 5 years. The reasons for this difference between the London and non-London recruitment sites are not known but may relate to non-standard setting of the recall date on discharge from colposcopy, so that women remained on early recall outside of the guidance that is in place for the routine screening services.
- The proportion of women with a positive HPV test on a paired clinician-collected sample ranged between 8.2% and 12.3%. These differences were not examined further but may be related to the women's age distributions, the HPV vaccination coverage among those attending screening, previous HPV-based screening in the area (e.g., in the English HPV pilot, or during early regional roll-out in 2018-2019 to mitigate cytology capacity), and other factors. Similar variations are seen in routine programme data collected from laboratories.
- In the Evalyn + cobas, FLOQswabs + cobas, and Evalyn + APTIMA workflows, self-collected samples showed a ~20% higher HPV positivity than their paired clinician-collected samples, with self-sampling positivity ranging between 10.5% and 14.3%. In the FLOQswabs + APTIMA and Multitest + APTIMA workflows, 22% of the self-collected samples returned a positive HPV test result, which was about doubled when compared with their paired clinician-collected samples. The reasons for the observation are as yet unclear. The study team contacted the London laboratory for additional information, but there appears to have been no obvious technical issue that could drive the observed discrepancy between the results of the clinician- and self-collected samples. Potential contributing factors related to (unknown) population characteristics are discussed below.
- Between 19% and 42% of women with positive HPV tests on the paired clinician-collected samples had abnormal cytology. This variation is well recognised in the routine programme. It is likely that much of the difference relates to what proportion of the population have been subject to previous rounds of HPV-based screening. This

[9] Gov.uk. Guidance. Cervical screening: programme and colposcopy management. URL: https://www.gov.uk/government/publications/cervical-screening-programme-and-colposcopy-management. Last accessed: 12 September 2024.

varies dramatically between the HPValidate sites, due to variations in the implementation date for HPV primary screening ranging between 2013 and 2019. Other factors including age distribution may also have an impact. It is considered highly unlikely, however, that differences in the performance of cytology influence the differences between the estimates of test specificity between the five self-sampling workflows.

3. Primary and exploratory analyses

The relative sensitivity and specificity estimates for the detection of CIN2+ of the five studied workflows (as compared to clinician collection) are reported in Table 5. Table 6 provides estimates from additional exploratory analyses.

Table 5. Estimates of conditional relative sensitivity and specificity, by self-sampling workflow, as estimated in the primary analysis.

Self-sampling workflow	N_{ss}/N_{LBC}, relative sensitivity (90% CI)	N_{ss}/N_{LBC}, relative specificity (90% CI)
Evalyn + cobas	92/102, 90.2% (84.3-94.0)	796/828, 96.1% (94.7-97.6)
FLOQswabs + cobas	113/120, 94.2% (89.6-96.8)	833/912, 91.3% (89.6-93.1)
Evalyn + APTIMA	66/86, 76.7% (68.5-83.4)	801/811, 98.8% (96.8-100.7)
FLOQswabs + APTIMA	61/66, 92.4% (85.2-96.3)	687/770, 89.2% (87.0-91.5)
Multitest + APTIMA	72/78, 92.3% (85.8-96.0)	784/904, 86.7% (84.7-88.8)

Abbreviations. CI: confidence interval. LBC: liquid-based cytology samples, collected by a clinician. SS: self-collected samples.
Legend. Dark grey: likely non-inferior at the threshold of 90% (relative sensitivity) or 95% (relative specificity); light grey: likely non-inferior at the threshold of 75% (relative sensitivity).
Note 1. Two-sided 90% confidence intervals were calculated, to test non-inferiority at the 1-sided α=0.05 level.
Note 2. For N_{ss}/N_{LBC} (number with the relevant HPV test result on the self-collected test / number with the relevant HPV test result on the clinician-collected LBC test), compare with Tables 3 (relative sensitivity) and 4 (relative specificity). A “relevant” test result is an HPV-positive test for the analysis of the relative sensitivity and an HPV-negative test for the analysis of the relative specificity.

Table 5 shows that none of the five self-sampling workflows could be validated as non-inferior on both the relative sensitivity (at the “higher” 90% threshold) and the relative specificity (at the 95% threshold) simultaneously.

Table 6. Summary of the estimates for the conditional relative sensitivity and specificity for self-sampling vs. clinician sampling from exploratory analyses, by self-sampling workflow.

	Self-sampling workflow				
	Evalyn + cobas	**FLOQswabs + cobas**	**Evalyn + APTIMA**	**FLOQswabs + APTIMA**	**Multitest + APTIMA**
Relative sensitivity	**N_{SS}/N_{LBC}, relative sensitivity (90% CI)**	**N_{SS}/N_{LBC}, relative sensitivity (90% CI)**	**N_{SS}/N_{LBC}, relative sensitivity (90% CI)**	**N_{SS}/N_{LBC}, relative sensitivity (90% CI)**	**N_{SS}/N_{LBC}, relative sensitivity (90% CI)**
Add CIN2+ from primary care[a]	95/106, 0.896 (0.837-0.935)	121/128, 0.945 (0.902-0.970)	74/96, 0.771 (0.693-0.833)	68/73, 0.932 (0.866-0.966)	85/91, 0.934 (0.878-0.966)
CIN3+ cases	41/45, 0.911 (0.816-0.959)	75/77, 0.974 (0.925-0.991)	36/47, 0.766 (0.652-0.851)	17/19, 0.895 (0.727-0.965)	31/31, 1.00 (0.920-1.00)
Only CIN2+ with high-grade abnormal cytology	71/77, 0.922 (0.856-0.959)	96/102, 0.941 (0.890-0.969)	27/39, 0.692 (0.562-0.798)	61/66, 0.924 (0.852-0.963)	50/55, 0.909 (0.825-0.955)
Relative specificity	**N_{SS}/N_{LBC}, relative specificity (90% CI)**	**N_{SS}/N_{LBC}, relative specificity (90% CI)**	**N_{SS}/N_{LBC}, relative specificity (90% CI)**	**N_{SS}/N_{LBC}, relative specificity (90% CI)**	**N_{SS}/N_{LBC}, relative specificity (90% CI)**
Exclude invalid self-collected samples	796/820, 0.971 (0.957-0.984)	833/857, 0.972 (0.959-0.985)	801/810, 0.989 (0.970-1.008)	687/770, 0.892 (0.870-0.915)	784/903, 0.868 (0.848-0.889)
Exclude samples from women with recent abnormalities	790/823, 0.960 (0.946-0.974)	827/904, 0.915 (0.898-0.932)	788/797, 0.989 (0.969-1.008)	640/710, 0.901 (0.879-0.925)	737/842, 0.875 (0.855-0.896)

Abbreviations. CI: confidence interval. CIN: cervical intraepithelial neoplasia. LBC: liquid-based cytology samples, collected by a clinician. SS: self-collected samples.
Note. For N_{SS}/N_{LBC}, see Table 5.
[a] In total across all five workflows, 42 cases of CIN2+ were excluded from the primary analyses of relative specificity in the primary screening populations. Those 42 cases were added to this exploratory analysis, which was pre-specified in the statistical analysis plan.

Relative test sensitivity

Four workflows:
- Evalyn + cobas,
- FLOQswabs + cobas,
- FLOQswabs + APTIMA, and
- Multitest + APTIMA

had point estimates of relative sensitivity higher than 90%. One of these workflows, FLOQswabs + cobas, could be considered non-inferior to clinician collection using the 90% non-inferiority threshold. The other three showed non-inferiority at the lower (75%) non-inferiority threshold. The study also observed a high relative sensitivity for the detection of CIN3+ with these four workflows (Table 6, exploratory analysis).

The Evalyn + APTIMA workflow did not meet either non-inferiority threshold for CIN2+ (Table 5, point estimate of 77% with a lower bound of the confidence interval <75%). It was also relatively low for CIN3+ (Table 6, exploratory analysis, point estimate: 77%). This was also the case when only women with high-grade referral cytology were included in the analysis (point estimate: 69%).

In terms of the test sensitivity, therefore, four workflows (Evalyn + cobas, FLOQswabs + cobas, FLOQswabs + APTIMA, and Multitest + APTIMA) can be considered potentially useful for screening within the English NHS.

Relative test specificity

Two workflows:

- Evalyn + cobas, and
- Evalyn + APTIMA

met the pre-specified non-inferiority threshold for relative specificity. One of these (Evalyn + cobas) also met the lower non-inferiority threshold for relative sensitivity.

A third workflow:

- FLOQswabs + cobas

appeared to have been uniquely negatively affected by a large number of invalid tests, which may have been a consequence of delayed testing of a proportion of self-collected samples, see below. The estimate of the relative specificity might have been higher if stricter conditions were imposed on the maximum acceptable sample turn-around times (Table 6).

The final two workflows:

- FLOQswabs + APTIMA and
- Multitest + APTIMA

showed a lower relative specificity than clinician collection. Although issues related to assay technology or sample processing procedures cannot be excluded, this may be related to population characteristics which were not measured in the study. This conclusion was based on the following information:

- Both workflows were studied in the catchment areas belonging to the same laboratory (London), and they both showed a doubling in HPV positivity between the paired clinician-collected and self-collected samples obtained on the same day.
- This appears to be consistent with the findings from the YouScreen study, which included catchment areas belonging to the same laboratory and used the FLOQswabs + cobas self-sampling workflow: in YouScreen, only half of the women with positive HPV self-sampling tests (cobas) also had positive HPV tests (APTIMA) once they presented for clinician collection for the purpose of cytological triage testing.[10]
- In HPValidate, the same FLOQswabs + cobas workflow was studied in the catchment areas of the Manchester laboratory, but the difference in HPV positivity between the paired self-collected and clinician-collected samples was much smaller, around 28% (Table 4), and similar to what was seen for the remaining two workflows studied in the catchment areas of the Gateshead and Bristol laboratories.

In terms of the test specificity, therefore, one workflow with non-inferior sensitivity at the lower 75% threshold (Evalyn + cobas) met the pre-specified specificity non-inferiority threshold. For the remaining three workflows with non-inferior sensitivity (FLOQSwabs + cobas, FLOQswabs + APTIMA, and Multitest + APTIMA), the possibility that lower estimates of the relative specificity may have been an artefact of the study and/or population characteristics cannot be excluded and require further validation.

[10] Lim AWW, Deats K, Gambell J, Lawrence A, Lei J, Lyons M, North B, Parmar D, Patel H, Waller J, Warwick J, Sasieni PD. Opportunistic Offering of Self-Sampling to Non-Attenders within the English Cervical Screening Programme: A Pragmatic, Multicentre, Implementation Feasibility Trial with Randomly Allocated Cluster Intervention Start Dates (Youscreen). eClinicalMedicine 2024;73:102672. Available on: https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(24)00251-7/fulltext.

Sample invalidity

Sample invalidity rates[11] observed in the study (Table 7) cannot be directly compared between the workflows. This is because sample invalidity is reported using different mechanisms for the cobas (where an internal control requires amplification of a human housekeeping gene) vs. the APTIMA (where a control checks for amplification/inhibition only) assays. On top of that, workload backlogs and batching of self-collected samples in preparation for HPV testing led to delays in resuspension and testing in some but not all laboratories.

Where data were available, self-collected samples were on average resuspended within around a week of collection; note that the workflow using the wet Multitest sampling device required resuspension at the point of collection. In a proportion of cases, however, resuspension took place within up to a month, particularly in the case of the FLOQswabs + cobas workflow. In this workflow, self-collected samples were tested on the day they were resuspended. Invalidity increased from 3.2% in the first week (N=439 tested), to 6.2% in the second week (N=452), 10.9% in the third week (N=92), and 36.4% in the fourth week (N=11).

[11] See also Cervical screening: implementation guide for primary HPV screening - GOV.UK (www.gov.uk). Last accessed: 12 September 2024.

Table 7. Proportions of women recruited in primary care with invalid HPV test results by time to resuspend a dry swab from sample taken date.

	Number of days between sample taken and resuspended/ tested							Unknown date of resuspension
	1-7	**8-14**	**15-21**	**22-30**	**>30**	**Total**	**P**	
Evalyn + cobas								
Total tested	911 (100%)	29 (100%)	0	0	0	940 (100%)		0
Valid result	903 (99.1%)	28 (96.6%)	0	0	0	931 (99.0%)		0
Invalid result	8 (0.9%)	1 (3.4%)	0	0	0	9 (1.0%)	0.247	0
FLOQswabs + cobas								
Total tested	439 (100%)	452 (100%)	92 (100%)	11 (100%)	0	994 (100%)		0
Valid result	425 (96.8%)	424 (93.8%)	82 (89.1%)	7 (63.6%)	0	938 (94.4%)		0
Invalid result	14 (3.2%)	28 (6.2%)	10 (10.9%)	4 (36.4%)	0	56 (5.6%)	<0.001	0
Evalyn + APTIMA								
Total tested	404 (100%)	21 (100%)	0	0	1 (100%)	426 (100%)		483 (100%)
Valid result	404 (100%)	2 (100%)	0	0	1 (100%)	426 (100%)		482 (99.8%)
Invalid result	0 (0%)	0 (0%)	0	0	0 (0%)	0 (0%)	x	1 (0.2%)
FLOQswabs + APTIMA								
Total tested	807 (100%)	69 (100%)	1 (100%)	1 (100%)	0	878 (100%)		0
Valid result	807 (100%)	69 (100%)	1 (100%)	1 (100%)	0	878 (100%)		0
Invalid result	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0	0 (0%)	x	0

Note. The Multitest+APTIMA workflow is not included; this is a wet device and requires no resuspension. P for differences in proportions of invalid results between the paired self- and clinician-collected samples within each workflow were calculated using the McNemar's test.

Summary

- Four workflows had point estimates of relative sensitivity of >90%, of which one met the higher (90%) and three met the lower (75%) pre-specified non-inferiority threshold using the confidence intervals. One of these four workflows, Evalyn + cobas, also met the pre-specified non-inferiority threshold for test specificity. Further validation is required to establish the relative specificity of the remaining three workflows.

- Only one of the five included self-sampling workflows (Evalyn + APTIMA) failed to meet both pre-specified non-inferiority thresholds for test sensitivity.

- Self-sample invalidity appeared to increase with time between the sample was taken until the sample was resuspended and tested.

RECOMMENDATIONS FOR THE IMPLEMENTATION IN NHS ENGLAND

Four workflows:

- Evalyn + cobas,
- FLOQswabs + cobas,
- FLOQswabs + APTIMA, and
- Multitest + APTIMA

are suitable candidates for further validation.

The following conditions should be considered for further validation:

1. In HPValidate, self-collection took place at primary care and colposcopy clinics. If NHS England plans to offer at-home self-collection, then these workflows should be evaluated when self-collection takes place **at home** and the samples are **posted by mail**. At-home self-collection introduces several additional variables which need to be understood prior to a national roll-out, such as the adequacy of the instructions for use,[12] reliability and timeliness of the transport to the laboratory, requirements for maximum turn-around times to avoid sample invalidity.

2. In HPValidate, the evaluation of the sensitivity for the detection of CIN2+ was undertaken on self-collected samples obtained when women attended their colposcopy appointments. In some cases, and in accordance with the national screening and clinical management pathways, this was a year or two after primary screening. Due to operational challenges which only eased towards the end of the recruitment period, the study samples were also dominated by a very high proportion of women with high-grade cytology. Hence, the data can be interpreted as suggesting that the sensitivity of the four self-sampling workflows may be at least as good as cytology but could not provide a definitive answer whether the achieved sensitivity is also as high as that of HPV testing on clinician samples collected at primary screening (=the intended population). It is therefore recommended that **further validation studies** are **undertaken in primary care** and adopt one of the designs that were discussed in a recent publication.[13] This would enable a more representative distribution of viral loads in samples collected for primary screening.

3. Although HPValidate evaluated the relative specificity in the intended population (women attending for their clinician-collected primary screening samples), the study could not provide a robust conclusion on this aspect of test accuracy. Future validation studies should be **extended across multiple catchment areas** to better understand whether the apparently poor specificity of certain testing technologies is real or an artefact of population characteristics in specific geographies.

[12] Note, for example, various observations reported from the behavioural part of the HPValidate study: here, some women felt reassured by the nurse (“which made all the difference”), and nurses helped some of the women use the device correctly (“Nurse explained fully the process whilst showing me the instructions - if she hadn't done this I may have struggled.”). See: Marlow L, Waller J: Acceptability of self-collecting vaginal samples in HPValidate and attitudes to self-sampling as a choice in future cervical screening. Unpublished report, December 2023, pages 26 and 27.

[13] Brentnall AR, Cuschieri K, Sargent A, Berkhof J, Rebolj M. Staged design recommendations for validating relative sensitivity of self-sample human papillomavirus tests for cervical screening. J Clin Epidemiol 2024;166:111227. Available on: https://pubmed.ncbi.nlm.nih.gov/38065518/.

4. **Standard operating procedures and protocols** that relate to the pre-analytical aspects of self-sampling should reflect these (and future, related) data.[14] **Explicit laboratory acceptance criteria** including the maximum time a sample can remain unprocessed/tested after collection should be created to minimise invalidity rate and thus the requirement for repeat samples. Additionally, in the absence of bespoke automation for the addition/processing of samples collected using dry devices to the platform at scale by the assay manufacturer, efforts should be made to find alternative compatible **automation**. Manual processing of individual samples would not be feasible for any larger implementation.

[14] See, for example: Connor L, Elasifer H, Sargent A, Bhatia R, Graham C, Cuschieri K. Influence of resuspension volume on dry sampling devices taken for human papillomavirus testing: implications for self-sampling. Biotechniques 2023;74(2):77-84. Available on: https://pubmed.ncbi.nlm.nih.gov/36655599/.