Performance of human papillomavirus testing on self-collected versus clinician-collected samples for the detection of cervical intraepithelial neoplasia of grade 2 or worse: a randomised, paired screen-positive, non-inferiority trial

Nicole J Polman, Renée M F Ebisch, Daniëlle A M Heideman, Willem J G Melchers, Ruud L M Bekkers, Anco C Molijn, Chris J L M Meijer, Wim G V Quint, Peter J F Snijders, Leon F A G Massuger, Folkert J van Kemenade, Johannes Berkhof*

Summary

Background Human papillomavirus (HPV) testing on self-collected samples is a potential alternative to HPV testing on clinician-collected samples, but non-inferiority of its clinical accuracy remains to be assessed in the regular screening population. The IMPROVE study was done to evaluate the clinical accuracy of primary HPV testing on self-collected samples within an organised screening setting.

Methods In this randomised, non-inferiority trial, women aged 29–61 years were invited to participate in the study as part of their regular screening invitation in the Netherlands. Women who provided informed consent were randomly allocated (1:1, with a block size of ten stratified by age) to one of two groups: a self-sampling group, in which women were requested to collect their own cervicovaginal sample using an Evalyn Brush (Rovers Medical Devices BV, Oss, Netherlands); or a clinician-based sampling group, in which samples were collected by a general practitioner with a Cervex-Brush (Rovers Medical Devices BV). All samples were tested for HPV using the clinically validated GP5+/6+ PCR enzyme immunoassay (Labo Biomedical Products BV, Rijswijk, Netherlands). HPV-positive women in both groups were retested with the other collection method and triaged by cytology and repeat cytology in accordance with current Dutch screening guidelines. Primary endpoints were detection of cervical intraepithelial neoplasia (CIN) of grade 2 or worse (CIN2+) and grade 3 or worse (CIN3+). Non-inferiority of HPV testing on self-collected versus clinician-collected samples was evaluated against a margin of 90% for the relative sensitivity and 98% for the relative specificity. This study is registered at the Dutch Trial register (NTR5078) and has been completed.

Findings Of the 187 473 women invited to participate, 8212 were randomly allocated to the self-sampling group and 8198 to the clinician-based sampling group. After exclusion of women who met the exclusion criteria or who did not return their sample, 7643 women were included in the self-sampling group and 6282 in the clinician-based sampling group. 569 (7·4%) self-collected samples and 451 (7·2%) clinician-collected samples tested positive for HPV (relative risk 1·04 [95% CI 0·92–1·17]). Median follow-up duration for HPV-positive women was 20 months (IQR 17–22). The CIN2+ sensitivity and specificity of HPV testing did not differ between self-sampling and clinician-based sampling (relative sensitivity 0·96 [0·90–1·03]; relative specificity 1·00 [0·99–1·01]). For the CIN3+ endpoint, relative sensitivity was 0·99 (0·91–1·08) and relative specificity was 1·00 (0·99–1·01).

Interpretation HPV testing done with a clinically validated PCR-based assay had similar accuracy on self-collected and clinician-collected samples in terms of the detection of CIN2+ or CIN3+ lesions. These findings suggest that HPV self-sampling could be used as a primary screening method in routine screening.

Funding Ministry of Health, Welfare, and Sport (Netherlands), and the European Commission.

Lancet Oncol 2019; 20: 229–38

Published **Online** January 15, 2019 http://dx.doi.org/10.1016/S1470-2045(18)30763-0

See **Comment** page 170

Cancer Centre Amsterdam, Department of Pathology, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, Netherlands (N J Polman MD, D A M Heideman PhD, Prof C J L M Meijer MD, Prof P J F Snijders PhD); **Department of Obstetrics and Gynaecology** (R M F Ebisch MD, R L M Bekkers MD, Prof L F A G Massuger MD) **and Department of Medical Microbiology** (W J G Melchers PhD), **Radboud University Medical Centre, Nijmegen, Netherlands; DDL Diagnostic Laboratory, Rijswijk, Netherlands** (A C Molijn MSc, W G V Quint PhD); **Department of Pathology, Erasmus University Medical Centre, Rotterdam, Netherlands** (Prof F J van Kemenade MD); **and Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, Amsterdam Public Health, Amsterdam, Netherlands** (Prof J Berkhof PhD)

*Prof Snijders died in May, 2018

Correspondence to: Prof J Berkhof, Amsterdam UMC, Vrije Universiteit Amsterdam, Department of Epidemiology and Biostatistics, 1007 MB Amsterdam, Netherlands **h.berkhof@vumc.nl**

Introduction

In several countries, human papillomavirus (HPV) testing is being implemented as a primary method in cervical screening.[1] HPV testing can be done on self-collected cervicovaginal material (HPV self-sampling), allowing the possibility to offer self-sampling to women in cervical screening programmes.[2] Previous studies showed that offering HPV self-sampling to screening non-attendees increased the proportion of participating patients.[3,4] Some countries already offer HPV self-sampling to screening non-attendees,[5–7] and HPV self-sampling is also used in low-resource areas with poor access to screening services.[8]

Implementation of self-sampling as a primary screening option could greatly reduce the workload of physicians doing cervical sampling and could therefore reduce the costs of screening. Most women also find HPV

Research in context

Evidence before this study

The expanding use of human papillomavirus (HPV) testing in cervical screening has led to a growing interest in HPV testing on self-collected samples. Studies have shown that women find HPV self-sampling to be more convenient, less embarrassing, less uncomfortable, and less painful than clinician-based sampling. Offering HPV self-sampling could increase participation, and some countries have implemented HPV self-sampling for screening non-attendees. HPV self-sampling is a potential primary screening method in the general screening population, but its clinical accuracy among screening responders remains to be assessed. We searched PubMed with the terms "self-sampling", "self-collected samples", and "human papillomavirus" for studies published in English between Oct 1, 1999 (initial studies on HPV self-sampling), and June 30, 2014 (drafting of the protocol). Studies reported substantial variation in clinical accuracy of HPV self-sampling, largely attributable to the use of different self-sampling devices and HPV assays. A meta-analysis evaluating the accuracy of HPV testing on self-collected versus clinician-collected samples indicated that HPV self-sampling, when done with an adequate self-sampling device and combined with a PCR-based HPV assay, could achieve accuracy similar to that of HPV testing on clinician-collected samples.

Added value of this study

Published studies comparing the accuracy of HPV testing on self-collected versus clinician-collected samples were small or were done in under-screened women, making generalisation of results to the regular screening population difficult. The IMPROVE study is, to our knowledge, the first large, randomised, non-inferiority trial done in the setting of an organised screening programme. The results show that, in the regular screening population, HPV testing done with a PCR-based HPV assay and an adequate self-sampling device has clinical sensitivity and specificity similar to that of HPV testing on clinician-collected samples.

Implications of all the available evidence

The non-inferiority of HPV self-sampling versus clinician-based HPV testing shown in this study, together with results of previous studies showing that women have a more positive attitude towards self-sampling than clinician-based sampling, suggest that implementation of HPV self-sampling could be used as a primary screening method in routine screening. Wider use of HPV self-sampling has the potential to greatly reduce the number of screening visits and to lower the barriers to screening in countries with low or moderate HPV screening coverage.

self-sampling to be more convenient, less embarrassing, less uncomfortable, and less painful than sampling done by a clinician, but are concerned about test accuracy.[9–11] Therefore, before self-sampling can be considered for the regular screening population, non-inferiority of the clinical accuracy of HPV testing on self-collected versus clinician-collected samples needs to be established.

Previous studies showed substantial variation in the clinical performance of HPV self-sampling, which was largely attributable to differences in the self-sampling device and the HPV assay.[12] A meta-analysis of clinical self-sampling studies showed that HPV self-sampling, when done with a PCR-based HPV assay, can achieve clinical accuracy similar to that of HPV testing on clinician-collected samples.[12] However, so far, self-sampling studies with suitable self-sampling devices and HPV assays have been small and often conducted in under-screened or never-screened women, making generalisation of the results to the regular screening population difficult.[9,13,14]

To our knowledge, the IMPROVE study is the first large, randomised, non-inferiority trial in the setting of an organised screening programme in which a brush-based self-sampling device is used in combination with a clinically validated, PCR-based HPV assay. We aimed to assess whether HPV testing on self-collected samples is non-inferior to clinician-collected samples in terms of the detection of cervical intraepithelial neoplasia (CIN) of grade 2 or worse (CIN2+) and grade 3 or worse (CIN3+).

Methods

Study design and participants

We did a randomised, non-inferiority trial within the Dutch cervical screening programme. In 2015 and 2016, 187 473 women living in the midwestern, southwestern, and eastern regions of the Netherlands were invited to participate. At that time, cervical screening in the Netherlands was cytology-based, and women aged 30–60 years were invited every 5 years for sample collection for cervical cytology at a general practitioner's practice. The invitation for the IMPROVE study (written, informed consent form and information leaflet) was sent along with the invitation for the regular screening programme, allowing women to choose between regular screening and the IMPROVE study. Women were enrolled in the IMPROVE study if they returned a signed informed consent form between April 15, 2015, and Dec 13, 2016. Exclusion criteria were previous hysterectomy and childbirth less than 6 months ago (exclusion criteria for invitation to routine cervical screening), as well as current pregnancy. During the whole study period, a website, email address, and telephone number were accessible for questions and additional information.

The study was done by the VU University Medical Centre (Amsterdam, Netherlands), the Erasmus University Medical Centre (Rotterdam, Netherlands), and the Radboud University Medical Centre (Nijmegen, Netherlands), in collaboration with screening organisations Midden-West, Zuid-West, and Oost. The IMPROVE study was approved by the Dutch Ministry of

Health, Welfare, and Sport (no 2014/32). The full study protocol is available in the appendix.

Randomisation and masking

Women who submitted a written, signed, informed consent form were registered in the study database. The study database automatically randomly assigned women (1:1) to the intervention group (HPV self-sampling) or the control group (clinician-based sampling). Randomisation was done in blocks of ten and stratified for seven age cohorts (29–33 years, 34–38 years, 39–43 years, 44–48 years, 49–53 years, 54–58 years, and 59–61 years). There was no masking for study participants, physicians, or researchers.

Procedures

Women assigned to the intervention group received a package including a brush-based self-sampling device (Evalyn Brush; Rovers Medical Devices BV, Oss, Netherlands), an explanatory letter about the study, and written and graphical user instructions about the device. Women were requested to self-collect a cervicovaginal sample and return the dry brush to the laboratory in a freepost return envelope. The Evalyn Brush is designed for HPV self-sampling, including wings indicating the depth of insertion and audible clicks for counting the number of rotations. Use of the Evalyn Brush has been described previously.[15]

Women assigned to the control group were invited to their general practitioner's practice to provide a clinician-collected sample. These samples were obtained with the Cervex-Brush (Rovers Medical Devices BV), a brush device used for cervical sampling by a physician during internal examination, and were collected in a vial with 10 mL ThinPrep PreservCyt media (Hologic, Marlborough, MA, USA). The vials were sent to the laboratory by a staff member from the general practioner's practice.

Upon arrival at the laboratory, the dry brushes from the self-sampling devices were suspended in 1·5 mL of ThinPrep PreservCyt media. 200 µL each of the self-sample and the clinician-collected sample were used for DNA isolation.

HPV DNA testing was done with the GP5+/6+ PCR enzyme immunoassay (Labo Biomedical Products BV, Rijswijk, Netherlands), which detects 14 high-risk HPV types (HPV16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68), according to manufacturer's instructions.[16]

Samples were visually inspected to assess whether sufficient material had been obtained. If no cell material was visible, a β-globin PCR to assess the presence of human DNA was done. In case of a negative β-globin PCR and a negative HPV assay, the sample was considered invalid, and the woman was requested to provide a new sample.

Participating women received a letter with the HPV test result and follow-up advice. Women whose results were negative for HPV were directed to routine screening.

Women in the intervention group with a positive HPV test were referred to their general practitioner to give a liquid-based cytology sample for cytological assessment. In the control group, reflex cytology was done for women with a positive HPV test result based on the available clinician-collected sample.

See Online for appendix

In accordance with the screening guidelines of the Dutch primary HPV screening programme, women with normal cytology at baseline were advised to undergo repeat cytological testing after 6 months. Women with abnormal cytology (borderline or mild dyskaryosis or worse) at baseline or at repeat testing were referred for colposcopy.

Slides were prepared from liquid-based cytology samples with a ThinPrep 5000 processor (Hologic). Cytology slides were Pap-stained and scored according to the CISOE-A classification,[17] which is easily translatable into the Bethesda 2001 classification. Cytotechnicians were informed about the HPV test result.

At the colposcopy visit, biopsies were taken from suspected areas according to standard procedures in the Netherlands.[18,19] If no abnormalities were seen, the gynaecologist was requested to take two random biopsies. Histological specimens were examined in local pathology laboratories in the Netherlands according to standard guidelines.[19] Histology samples were classified as no CIN, CIN1, CIN2, CIN3, or invasive cancer.[20] Carcinoma in situ and adenocarcinoma in situ were classified under CIN3.

According to the randomised, paired screen-positive study design,[21] HPV-positive women in the intervention self-sampling group were tested for HPV in the clinician-collected sample, and HPV-positive women in the clinician-based sampling group were requested to obtain a self-sample for HPV testing. Cross-testing was done before colposcopy and HPV cross-testing results were not disclosed to study participants and were not used for screening management. Adverse events during follow-up were not recorded as per protocol because the study was embedded within the screening programme. Adverse events related to the HPV testing procedure were not expected, but participants were able to report pain or difficulty with the procedure.

Follow-up cytology, colposcopy, and histology in HPV-positive women were collected directly from pathology laboratories and gynaecologists. The nationwide network and registry of histopathology and cytopathology in the Netherlands (PALGA) was consulted to complete cytology and histology when missing.[22] Cytological and histological findings were included in our analysis when recorded before March 1, 2018.

Outcomes

Primary endpoints were detection of CIN2+ and CIN3+. Detection of invasive cancer after 1–2 years was also a main endpoint. The number of invasive cancer cases was reported, but non-inferiority analyses for this endpoint were not done because of low numbers of cases. Cost-effectiveness of HPV self-sampling was also specified as

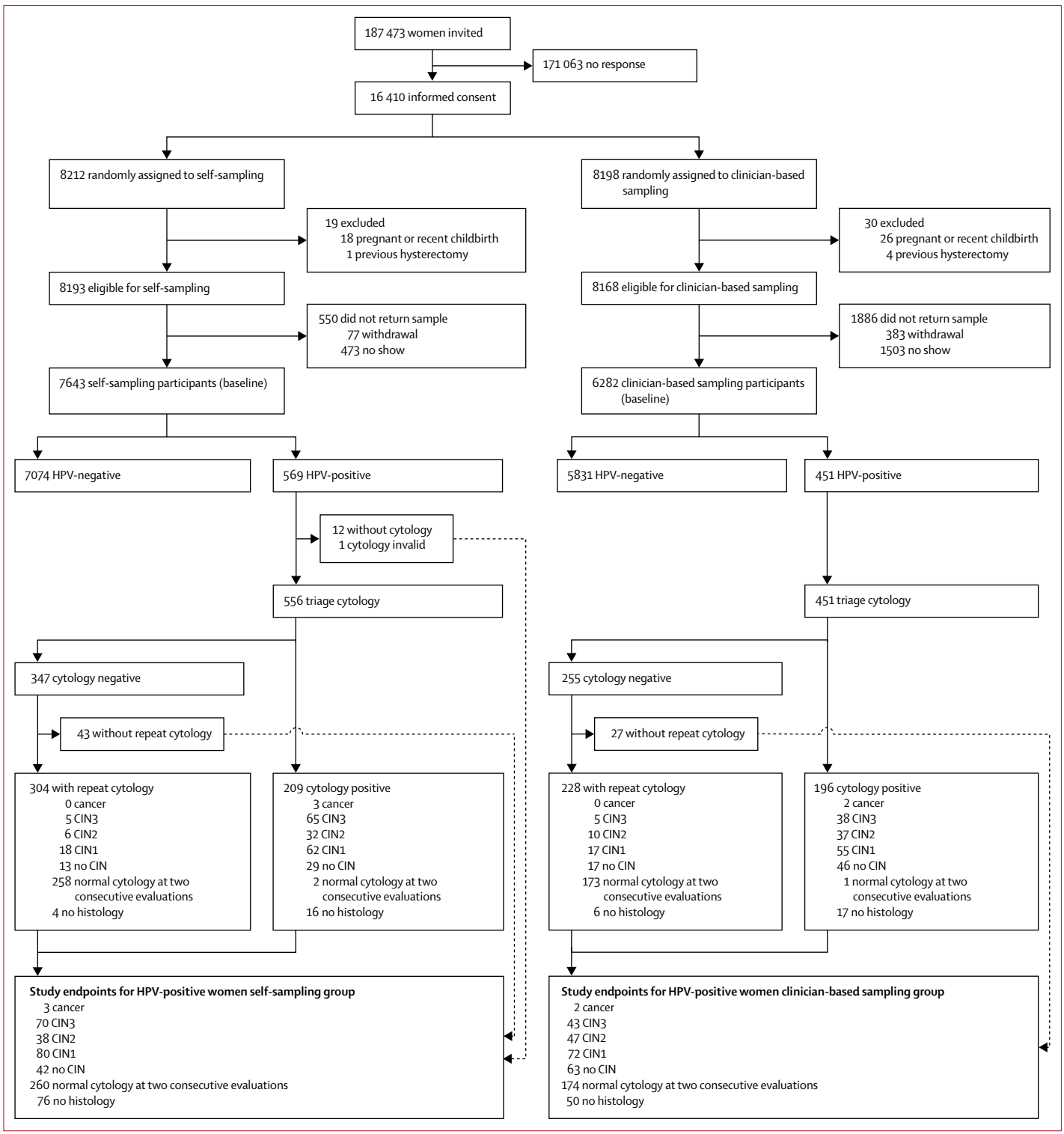

Figure: **Study flowchart**
HPV=human papillomavirus. CIN=cervical intraepithelial neoplasia.

a secondary outcome in the protocol, but will be studied and published separately. Detection of CIN2+ and CIN3+ after 5 years will be also compared when the next screening round has taken place.

Statistical analysis

We hypothesised that HPV testing on self-collected samples is non-inferior to HPV testing on clinician-collected samples for the detection of CIN2+ and CIN3+. To show with at least 80% power that the sensitivity of HPV testing on self-collected samples is non-inferior to the sensitivity of HPV testing on clinician-collected samples for detection of CIN2+ and CIN3+, 14 000 women needed to be tested.[21] The non-inferiority margin of the clinical sensitivity was set at 90%, as recommended in consensus guidelines for new HPV DNA tests.[23] The results of the two HPV testing methods were assumed to be independent in women with CIN2+, the sensitivities of both HPV tests were assumed to be 0·95, and the proportion of women with CIN2+ was set at 0·012 and CIN3+ at 0·008, based on a large population-based HPV screening trial.[24] Around 30 000 women were invited between April 15 and Dec 31, 2015; however, because of an unexpectedly low participation rate of 8·8%, the total number of invitations was increased to 185 000 in 2016, inclusive of the 30 000 women already invited. A licence was provided by the Dutch Ministry of Health, Welfare, and Sport, for the enrolment of women from Jan 1, 2015, until Dec 31, 2016. Because of logistics, the number of invitations could not be exactly determined upfront and, eventually, 187 473 invitations were sent out. The final date of enrolment was Dec 13, 2016.

Among women with a valid HPV test result, risk ratios (RRs) for HPV positivity and the detection of CIN2+ and CIN3+ between the two study groups were calculated together with 95% CIs calculated by the Wald method. A difference between study groups was considered significant if the 95% CI of the RR was completely below or above 1. As a post-hoc analysis, RRs for the detection of CIN2+ and CIN3+ were calculated among all women randomly allocated to the study groups.

Sensitivity and specificity of both sampling methods were calculated with the crude data and an adjusted dataset. Crude sensitivity was estimated by the number of positive HPV cross-test results among women with detected disease. Crude specificity was estimated by the number of negative HPV test results among women without detected disease. Adjusted data were obtained by imputing the expected number of CIN2+ and CIN3+ in HPV-positive women without histology or two-times normal cytology, based on their cytology and colposcopy results. Relative sensitivity and specificity are presented with Wald 95% CIs. A difference in sensitivity or specificity between study groups was considered significant if the 95% CI of the relative sensitivity or specificity was completely below or above 1. Non-inferiority of the HPV test on self-collected samples versus clinician-collected samples was evaluated by non-inferiority testing against a margin of 90% for the relative sensitivity and a margin of 98% for the relative specificity. These non-inferiority criteria are used for evaluation and validation of new HPV assays, according to international consensus criteria for validation of HPV assays.[23] Non-inferiority was assessed by one-sided Wald tests. In additional post-hoc analyses, differences in relative sensitivity and specificity for CIN2+ between age cohorts were assessed with the Mantel-Haenszel test of homogeneity, and non-inferiority of HPV testing on self-collected versus clinician-collected samples was evaluated after exclusion of under-screened women. A woman was classified as under-screened if she was 34 years or older (and would have had a previous screening invitation) and the time since last cervical cytology was at least 7 years. The screening interval is 5 years in the Netherlands, but the threshold for being under-screened was set at 7 years because of variability in the timing of the screening invitation and in the time between the invitation and the test evaluation.

Statistical analyses were done with IBM SPSS (version 22.0) and STATA (version 14.1). p values less than 0·05 were considered to indicate statistical significance. This study is registered at the Dutch Trial Register, number NTR5078.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to submit the paper for publication. NJP and JB had access to the raw data. The corresponding author had full access to all the data and had the final responsibility to submit for publication.

	Patients eligible for self-sampling (n=8193)	Patients eligible for clinician-based sampling (n=8168)
Already participated or preferred to participate in regular screening programme	49 (0·6%)	128 (1·6%)
Only agreed to participate if allocated to other sampling method	2 (<0·1%)	131 (1·6%)
Logistical reasons at the general practitioner's practice*	1 (<0·1%)	21 (0·3%)
Did not do self-sampling or have clinician-based sampling done (eg, because of difficulty or pain)	7 (0·1%)	19 (0·2%)
Had a cervical sample taken because of gynaecological complaints within the past 12 months	2 (<0·1%)	14 (0·2%)
Died	0 (<0·1%)	1 (<0·1%)
Other reasons	16 (0·2%)	69 (0·8%)
Subtotal		
Reason known	77 (0·9%)	383 (4·7%)
Reason unknown	473 (5·8%)	1503 (18·4%)
Total withdrawal	550 (6·7%)	1886 (23·1%)

Data are n (%). *For example, sample sent to routine screening laboratory instead of study laboratory.

Table 1: **Reasons for study withdrawal by randomisation group**

	Self-sampling group (n=7643)	Clinician-based sampling group (n=6282)
Age, years		
Mean (SD)	45·5 (9·4)	45·7 (9·3)
Median (IQR)	45 (39–54)	45 (39–54)
29–33	745 (9·7%)	600 (9·6%)
34–38	888 (11·6%)	712 (11·3%)
39–43	1055 (13·8%)	839 (13·4%)
44–48	1394 (18·2%)	1149 (18·3%)
49–53	1154 (15·1%)	957 (15·2%)
54–58	1333 (17·4%)	1143 (18·2%)
59–61	1074 (14·1%)	882 (14·0%)
Region		
Midwestern	2668 (34·9%)	2219 (35·3%)
Southwestern	2377 (31·1%)	1891 (30·1%)
Eastern	2598 (34·0%)	2172 (34·6%)

Data are n (%) unless otherwise specified.

Table 2: **Baseline characteristics**

	Total	HPV-positive	CIN2+	CIN3+
Self-sampling group	7643	569 (7·4%)	111 (1·5%)	73 (1·0%)
29–33 years	745	129 (17·3%)	34 (4·6%)	22 (3·0%)
34–38 years	888	96 (10·8%)	26 (2·9%)	19 (2·1%)
39–43 years	1055	68 (6·4%)	17 (1·6%)	13 (1·2%)
44–48 years	1394	82 (5·9%)	14 (1·0%)	7 (0·5%)
49–53 years	1154	82 (7·1%)	11 (1·0%)	9 (0·8%)
54–58 years	1333	69 (5·2%)	6 (0·5%)	2 (0·2%)
59–61 years	1074	43 (4·0%)	3 (0·3%)	1 (0·1%)
Clinician-based sampling group	6282	451 (7·2%)	92 (1·5%)	45 (0·7%)
29–33 years	600	98 (16·3%)	33 (5·5%)	19 (3·2%)
34–38 years	712	75 (10·5%)	12 (1·7%)	3 (0·4%)
39–43 years	839	58 (6·9%)	14 (1·7%)	7 (0·8%)
44–48 years	1149	74 (6·4%)	9 (0·8%)	5 (0·4%)
49–53 years	957	65 (6·8%)	17 (1·8%)	6 (0·6%)
54–58 years	1143	56 (4·9%)	5 (0·4%)	4 (0·3%)
59–61 years	882	25 (2·8%)	2 (0·2%)	1 (0·1%)

Data are n (%). HPV=human papillomavirus. CIN2+=cervical intraepithelial neoplasia grade 2 or worse. CIN3+=cervical intraepithelial neoplasia grade 3 or worse.

Table 3: **HPV prevalence and cumulative CIN2+ and CIN3+ detection among participants by study group and age group**

	Total	Self-sampling group	Clinician-based sampling group
CIN2 or worse	184/194 (95%)	106/110 (96%)	78/84 (93%)
CIN3 or worse	108/113 (96%)	69/72 (96%)	39/41 (95%)
CIN1 or less*	459/671 (68%)	256/381 (67%)	203/290 (70%)
No histology	70/99 (71%)	34/57 (60%)	36/42 (86%)

Data are n/N (%). Results are only shown for HPV-positive women who returned a HPV cross-test (21 HPV-positive women in the self-sampling group and 35 HPV-positive women in the clinician-based sampling group did not return a HPV cross-test). *Women with histologically confirmed CIN1 or no CIN, and women with two consecutive normal cytology results. CIN=cervical intraepithelial neoplasia.

Table 4: **HPV-positive cross-test results by study group and outcome**

Results

Between April 15, 2015, and Dec 13, 2016, 16 410 (8·8%) of 187 473 women returned a signed informed consent form and were randomly assigned to either self-sampling (8212 [50·0%]) or clinician-based sampling (8198 [50·0%]). 49 (0·3%) of 16 410 women were excluded (figure). Samples were received for 7643 (93·3%) of 8193 women eligible for self-sampling and 6282 (76·9%) of 8168 eligible for clinician-based sampling. Reasons for not providing a sample are shown in table 1. The mean age of women not providing a sample was 44·2 years (median 45, IQR 39–54) in the self-sampling group and 44·7 years (45, 39–54) in the clinician-based sampling group. The proportion of women who did not provide a sample was lower in the self-sampling group (550 [6·7%] of 8193) than in the clinician-sampling group (1886 [23·1%] of 8168).

Baseline characteristics are shown in table 2. Time between randomisation and receipt of the sample at the laboratory was shorter in the self-sampling group (median 27·0 days, IQR 19·0–44·0) than in the clinician-based sampling group (44·0 days, 30·0–65·0).

569 (7·4%) of 7643 self-collected and 451 (7·2%) of 6282 clinician-collected samples tested positive for HPV (RR 1·04 [95% CI 0·92–1·17]; table 3). HPV prevalence in the self-sampling group and clinician-based sampling group were similar within every 5-year age cohort (table 3). Among HPV-positive women aged 34 years and older, the proportion of under-screened women was slightly higher in the self-sampling group (48 [13%] of 382) than in the clinician-based sampling group (21 [7%] of 314). Median follow-up duration for HPV-positive women was 20 months (IQR 17–22).

Of the 569 HPV-positive women in the self-sampling group, 556 (98%) had a valid cytology triage result. In the clinician-based sampling group, cytology results were available for all 451 HPV-positive women. Of those with valid cytology results, 209 (38%) women in the self-sampling group and 196 (43%) women in the clinician-based sampling group had abnormal cytology at baseline and were referred for colposcopy. In the self-sampling group, two women (<1%) were diagnosed with squamous-cell carcinoma, one (<1%) with adenocarcinoma, 65 (11%) with CIN3, and 32 (6%) with CIN2. In the clinician-based sampling group, one (<1%) woman was diagnosed with squamous-cell carcinoma, one (<1%) with adenocarcinoma, 38 (8%) with CIN3, and 37 (8%) with CIN2 (figure).

304 (88%) of 347 women in the self-sampling group with normal triage cytology at baseline had repeat cytology. Of those, five were diagnosed with CIN3 and six with CIN2. In the clinician-based sampling group, 228 (89%) of 255 women with normal cytology at baseline had repeat cytology, of whom five were diagnosed with

CIN3 and ten with CIN2. Among women with a valid HPV test result, CIN2+ detection was similar between the self-sampling group (111 [1·5%] of 7643) and clinician-based sampling group (92 [1·5%] of 6282; RR 0·99 [95% CI 0·75–1·31]; table 3). CIN3+ detection was also similar between the self-sampling group (73 [1·0%] of 7643) and the clinician-based sampling group (45 [0·7%] of 6282; RR 1·33 [0·92–1·93]). Post-hoc analyses showed that among all randomised women, detection of CIN2+ was similar in the self-sampling group (111 [1·4%] of 8193) and in the clinician-based sampling group (92 [1·1%] of 8168; RR 1·20 [0·91–1·58]), but detection of CIN3+ was higher in the self-sampling group (73 [0·9%] of 8193) than in the clinical-based sampling group (45 [0·6%] of 8168; RR 1·62 [1·12–2·34]).

In the self-sampling group, HPV cross-tests (on clinician-collected samples) were available for 548 (96·3%) of 569 HPV-positive women. In the clinician-based sampling group, HPV cross-tests (on self-collected samples) were available for 416 (92·2%) of 451 HPV-positive women. The mean time between baseline test and HPV cross-test was 38·8 days (SD 23·7; median 33·0 days, IQR 27·0–42·0) in the self-sampling group and 38·1 days (25·6; median 32·0 days, 26·0–40·75) in the clinician-based sampling group, and the difference in means, measured with an independent samples *t* test, was not significant (difference 0·7 days, 95% CI –2·4 to 3·8). Cross-test results are shown in table 4.

The crude sensitivity of HPV testing on self-collected samples for the detection of CIN2+ was similar to that of clinician-collected HPV testing (relative sensitivity 0·96 [95% CI 0·90–1·03]; table 5). The crude specificity of self-sampling was also similar to that of clinician-based sampling (relative specificity 1·00 [0·99–1·01]). For endpoint CIN3+, the sensitivity and specificity of HPV testing on self-collected samples were similar to those of clinician-collected HPV testing (relative sensitivity 0·99 [0·91–1·08]; relative specificity 1·00 [0·99–1·01]). Results were similar for the adjusted data (table 5).

Non-inferiority of the sensitivity of the HPV test on self-collected versus clinician-collected samples at a relative sensitivity margin of 90% was observed for the CIN2+ (crude data p=0·0285; adjusted p=0·0162) and CIN3+ (p=0·0111; adjusted p=0·0064) endpoints. Furthermore, non-inferiority of the specificity of the HPV test on self-collected versus clinician-collected samples at a relative specificity margin of 98% was observed for an endpoint less than CIN2 (p<0·0001; adjusted p<0·0001) and less than CIN3 (p<0·0001; adjusted p<0·0001; table 5).

Post-hoc analyses showed that the relative sensitivity and specificity for detection of CIN2+ did not differ between age cohorts (Mantel-Haenszel test of homogeneity, p=0·22 and p=0·54, respectively). Nine women with CIN2+ were found to be under-screened: one with CIN2, six with CIN3, and one with cancer in the self-sampling group, and one with cancer in the clinician-based sampling group. After exclusion of under-screened

	Unadjusted data		Adjusted data*	
	n/N (% [95% CI])	Relative accuracy (95% CI)	% (95% CI)	Relative accuracy (95% CI)
CIN2 or worse				
Sensitivity				
Self-sampling	78/84 (92·9% [87·3–98·4])	0·96 (0·90–1·03)	93·1% (88·1–98·0)	0·97 (0·91–1·03)
Clinician-based sampling	106/110 (96·4% [92·9–99·9])		96·3% (93·0–99·7)	
Specificity				
Self-sampling	7074/7532 (93·9% [93·4–94·5])	1·00 (0·99–1·01)	94·0% (93·5–94·6)	1·00 (0·99–1·01)
Clinician-based sampling	5831/6190 (94·2% [93·6–94·8])		94·3% (93·7–94·9)	
Sensitivity (no under-screened)				
Self-sampling	72/78 (92·3% [86·4–98·2])	0·97 (0·89–1·04)	92·7% (87·4–98·1)	0·97 (0·90–1·04)
Clinician-based sampling	87/91 (95·6% [91·4–99·8])		95·4% (91·3–99·5)	
CIN3 or worse				
Sensitivity				
Self-sampling	39/41 (95·1% [88·5–100])	0·99 (0·91–1·08)	95·2% (89·1–100)	0·99 (0·92–1·07)
Clinician-based sampling	69/72 (95·8% [91·2–100])		95·8% (91·3–100)	
Specificity				
Self-sampling	7074/7570 (93·4% [92·9–94·0])	1·00 (0·99–1·01)	93·5% (93·0–94·1)	1·00 (0·99–1·01)
Clinician-based sampling	5831/6237 (93·5% [92·9–94·1])		93·5% (93·0–94·2)	
Sensitivity (no under-screened)				
Self-sampling	36/38 (94·7% [87·6–100])	1·00 (0·91–1·10)	95·0% (88·4–100)	1·00 (0·92–1·10)
Clinician-based sampling	54/57 (94·7% [88·9–100])		94·5% (88·8–100)	

CIN=cervical intraepithelial neoplasia. *Adjusted for HPV-positive women without histology or two times normal cytology.

Table 5: **Clinical performance of self-sampling compared with clinician-based sampling**

women, HPV testing on self-collected samples was still non-inferior to HPV testing on clinician-collected samples with regard to sensitivity for detection of CIN2+ (crude data p=0·0385; adjusted data p=0·0182) and CIN3+ (p=0·0164; adjusted p=0·0084; table 5).

Discussion

The results of this randomised, non-inferiority trial showed that HPV testing on self-collected samples and clinician-collected samples has similar sensitivity and similar specificity for detection of CIN2+ and CIN3+. A non-inferiority assessment, in which international consensus guidelines for clinical validation of new HPV DNA assays were followed,[23] showed that use of the combination of a self-sampling device (Evalyn brush) and HPV assay (GP5+/6+ PCR enzyme immunoassay) is non-inferior to clinician-collected HPV testing with respect to clinical sensitivity and specificity. Invasive cancer is not included in consensus guidelines for validation of HPV assays, but was a main endpoint in our study. The number of cancers was small, but similar in the intervention and control group.

HPV self-sampling has potential for use as a primary screening test, but its clinical accuracy needs to be established with a high level of confidence.

We chose a randomised, paired, screen-positive design in which screen-positive women were requested to provide a second sample.[21] This design is preferable to comparing self-sampling with clinician-based sampling in an unpaired manner. Although a fully paired study design is recommended in diagnostic studies,[25] we did not choose this design because compliance with the second sampling method is expected to be poor among women who test negative on the first sample. An alternative is to collect two samples on the same day,[13,26] but a negative bias in the accuracy of the second sample cannot be ruled out because cervical sampling causes minor damage to the cervical tissue. Furthermore, a paired screen-positive design is expected to yield higher power for non-inferiority testing than that of the fully paired design when the study size is constrained by the number of HPV tests.[21,27] The paired screen-positive design provides a valid estimation procedure for relative sensitivity. However, the relative specificity might be biased because CIN2+ cases in HPV-negative women were not observed. This bias is expected to be minimal because only 1·5% of women were found to be CIN2+ in both study groups, and few CIN2+ cases had discordant HPV test results.

The analyses in this study are valid under the assumption that the quality of the cross-test sample and the cross-test HPV result are not influenced by baseline sampling several weeks earlier. If cervical sampling leads to tissue damage, the damage from clinician-based sampling will probably be more severe than that from self-sampling, in which case a bias in the estimation of the relative sensitivity will support clinician-based sampling rather than self-sampling.

A strength of our study, which was intended to reflect a potential routine screening setting with HPV self-sampling as the primary screening method, was that self-sampling was done at home without the help of a clinician.[5] Furthermore, women were invited to participate in the IMPROVE study as part of their regular invitation in the setting of primary HPV screening.

The proportion of women providing informed consent in our opt-in study was 8·8%, which raises concerns about the generalisability of the results beyond the study population. The risk of an underlying high-grade lesion is associated with both age and screening history, and these factors deserve careful scrutiny. The mean age in the study was about 45 years in all women providing informed consent and in women with a valid HPV result. This age is similar to the mean age of screening-eligible women in the Netherlands. We did not observe any associations between age and the relative sensitivity and relative specificity in our study, supporting the clinical accuracy of HPV self-sampling in age groups with different disease risks. The comparability of our study population and the general screening population is further supported by the CIN2+ prevalence of 1·5% in the study population, which was also predicted for the new Dutch HPV-based screening programme.[28] Among women who were previously invited for screening, attendance at the previous screen was slightly lower in the self-sampling group than in the clinician-based sampling group. When we repeated the analyses after exclusion of under-screened women, the sensitivity and specificity of HPV testing on self-collected and clinician-collected samples remained similar. This observation, together with the absence of an association between age and relative sensitivity and specificity, provides support for the accuracy of HPV self-sampling among regularly screened women. We also found that the proportion of women who withdrew from the study was higher in the clinician-based sampling group than in the self-sampling group, which could be explained by women in the clinician-based sampling group choosing to withdraw when they were not randomised to the preferred sampling method. Indeed, among the subset of women who gave a reason for their withdrawal, allocation to the non-preferred sampling group was the most frequently mentioned reason. These women might represent a subgroup of women with irregular attendance at screening, but the aim of our study was to show that HPV self-sampling can also be applied to regular screening attendees.

In addition to test accuracy, the overall effectiveness of HPV self-sampling within a routine screening setting is influenced by the acceptability of self-sampling in this setting and the compliance with follow-up among HPV-positive women. Previous studies have evaluated these factors, although mostly among screening non-attendees. Multiple studies reported high acceptability of

self-sampling when self-sampling devices were sent directly to women instead of using an opt-in strategy as in our study.[4,29] Moreover, questionnaire studies showed that women find self-sampling less uncomfortable and more convenient than clinician-based sampling.[9,10,15] Regarding the compliance to follow-up recommendations, varying estimates have been reported. Studies done in France, Sweden, and the UK showed low compliance with follow-up recommendations (<60%),[30–32] whereas studies in the Netherlands and Denmark showed high compliance (about 90%) with follow-up recommendations among self-sampling, HPV-positive women.[3,15] This suboptimal compliance in some studies indicates that HPV self-sampling cannot be adopted without a clear implementation and communication strategy. We expect that an active role of a health-care professional, with regard to promoting the use of the self-sampling device, would lead to increased compliance.

Another important aspect of the adoption of HPV self-sampling is the cost of screening. HPV self-sampling is expected to reduce the workload of clinicians doing cervical sampling. In the Netherlands, about 500 000 cervical samples are taken annually. We expect a reduction of at least 90% by HPV self-sampling because only HPV-positive women require cervical sampling by a clinician. However, the costs also depend on the price of the HPV self-sampling device, which can be higher than the price of a brush used at the clinician's office. Furthermore, the adoption of HPV self-sampling could affect the costs of transporting cervical samples to the screening laboratories. Cost-effectiveness analyses prespecifed in our protocol will be reported elsewhere.

There were several other limitations to our study. First, the study leaflet was sent together with the regular screening invitation, and, although it was clearly explained that a choice had to be made between participation in the IMPROVE study or routine cytological screening, some women withdrew. Second, some general practitioners accidentally sent the cervical sample to a routine screening laboratory instead of a study laboratory. These cases were also reported in table 1. Third, some women referred for colposcopy did not have a histological endpoint; however, we do not expect that this factor biased our results because women were managed in the same way in both study groups. Furthermore, more than 90% of women referred for colposcopy had a histological result, and the adjusted data in which we corrected for incomplete follow-up showed results similar to the crude data. Fourth, original diagnoses by local pathologists were used and might have been subject to misclassification. However, high consistency between original CIN3+ diagnosis and independent review diagnoses by other pathologists has previously been observed in a large HPV screening study in the Netherlands.[24] CIN2 is a more ambiguous diagnosis than CIN3,[33] and that the IMPROVE study showed non-inferiority of self-sampling for both CIN2+ and CIN3+ is reassuring.

To the best of our knowledge, the IMPROVE study is the first randomised trial to show non-inferiority of clinical performance of HPV self-sampling versus clinician-collected HPV testing for the detection of CIN2+ and CIN3+. These findings suggest that HPV self-sampling—with a suitable combination of self-sampling device and PCR-based HPV assay—could be used as a primary screening method in nationwide screening programmes. Wider use of HPV self-sampling could greatly reduce the workload of clinicians doing cervical sampling, reducing the costs of screening, and lowering barriers to screening in countries with low or moderate screening coverage.

Contributors

PJFS, CJLMM, LFAGM, FJvK, and JB contributed to study design. Data collection was done by NJP, RMFE, DAMH, WJGM, RLMB, ACM, WGVQ, and PJFS. NJP and JB were responsible for data analysis and prepared the first draft of the manuscript. All authors reviewed the manuscript and approved the final version.

Declaration of interests

DAMH is minority stock holder of Self-Screen BV, a spin-off company of VU University Medical Centre. Self-screen BV holds patents related to the work; has been on the speaker's bureau of Qiagen; and serves occasionally on the scientific advisory boards of Pfizer and Bristol-Meyer Squibb. RLMB reports grants from Roche diagnostics. ACM reports grants from DDL Diagnostic Laboratory. CJLMM has received speakers' fees from Sanofi Pasteur MSD/Merck; has served occasionally on the scientific advisory boards (expert meetings) of Qiagen, GlaxoSmithKline, and Sanofi Pasteur MSD/Merck; has been co-investigator on a Sanofi Pasteur MSD-sponsored HPV vaccination trial, for which his institute received research funding; is part-time director of and minority stock holder in Self-Screen BV, a spin-off company of VU University Medical Centre, which owns patents on methylation markers and HPV detection; and holds some shares in Qiagen, and until April, 2016, had minority stock of Diassay BV. WGVQ is shareholder of LBP. PJFS was on the speakers' bureaus of Roche diagnostics, Gen-Probe, Abbott, Qiagen, and Seegene; was a consultant for Crucell BV; and was a minority shareholder in Self-Screen BV. JB reports personal fees from GlaxoSmithKline and Merck; and non-financial support from DDL; the fees from GlaxoSmithKline and Merck were collected by his employer. NJP, RMFE, WJGM, LFAGM, and FJvK declare no competing interests.

Data sharing

Individual participant data will be made available to investigators who provide a methodologically sound proposal for meta-analyses. Proposals should be directed to h.berkhof@vumc.nl. The approval of the Dutch study protocol is available at the website of The Dutch Health Council (https://www.gezondheidsraad.nl). The English translation of the study protocol is available as an online appendix to this Article.

Acknowledgments

We thank all women who participated in the IMPROVE study; Johan Bulten for his contribution to the study design and implementation; the screening organisations Midden-West, Zuid-West, and Oost; the National Institute for Public Health and the Environment; the physicians, gynaecologists, and pathologists in the study regions; and the PALGA Foundation for their help with the PALGA data collection. We also thank the technicians and research staff at the Unit of Molecular Pathology, Department of Pathology, VU University Medical Centre (particularly Dénira Agard); at the Department of Medical Microbiology and Department of Pathology, Radboud University Medical Centre; and at the DDL Diagnostic Laboratory. This study was funded by the Ministry of Health, Welfare, and Sport (number 325973) and the European Commission 7th Framework Programme for Research and Technological Development (CoheaHr, number 603019).

References

1 Wentzensen N, Arbyn M, Berkhof J, et al. Eurogin 2016 Roadmap: how HPV knowledge is changing screening practice. *Int J Cancer* 2017; **140:** 2192–200.

2 Brink AA, Meijer CJ, Wiegerinck MA, et al. High concordance of results of testing for human papillomavirus in cervicovaginal samples collected by two methods, with comparison of a novel self-sampling device to a conventional endocervical brush. *J Clin Microbiol* 2006; **44:** 2518–23.

3 Gök M, Heideman DA, van Kemenade FJ, et al. HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. *BMJ* 2010; **340:** c1040.

4 Verdoodt F, Jentschke M, Hillemanns P, Racey CS, Snijders PJ, Arbyn M. Reaching women who do not participate in the regular cervical cancer screening programme by offering self-sampling kits: a systematic review and meta-analysis of randomised trials. *Eur J Cancer* 2015; **51:** 2375–85.

5 Dutch Health Council. Law on population-based screening: renewal of national population-based screening programme for cervical cancer. 2016 (in Dutch). https://www.gezondheidsraad.nl/documenten/adviezen/2016/06/20/wet-op-het-bevolkingsonderzoek-vernieuwing-landelijk-bevolkingsonderzoek-baarmoederhalskanker (accessed Sept 23, 2018).

6 Medical Services Advisory Committee Australia. Effectiveness modelling and economics evaluation in the Australian setting. In: National cervical screening program renewal: executive summary. November, 2013. http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/E6A211A6FFC29E2CCA257CED007FB678/$File/Executive%20Summary%20notated%2013.6.14.pdf (accessed Sept 23, 2018).

7 Danish Health Council. Screening for cervical cancer—recommendations. 2017 (in Danish). https://prodstoragehoeringspo.blob.core.windows.net/fdaf8cf3-e83f-4c62-87ea-611ade1dd743/SST%20Anbefalinger%20livmoderhalskr%C3%A6ftscreening%20-%20H%C3%B8ring%20nov%202017.pdf (accessed Sept 23, 2018).

8 Zhao FH, Lewkowitz AK, Chen F, et al. Pooled analysis of a self-sampling HPV DNA test as a cervical cancer primary screening method. *J Natl Cancer Inst* 2012; **104:** 178–88.

9 Ketelaars PJW, Bosgraaf RP, Siebers AG, et al. High-risk human papillomavirus detection in self-sampling compared to physician-taken smear in a responder population of the Dutch cervical screening: results of the VERA study. *Prev Med* 2017; **101:** 96–101.

10 Sultana F, Mullins R, English DR, et al. Women's experience with home-based self-sampling for human papillomavirus testing. *BMC Cancer* 2015; **15:** 849.

11 Bosgraaf RP, Ketelaars PJ, Verhoef VM, et al. Reasons for non-attendance to cervical screening and preferences for HPV self-sampling in Dutch women. *Prev Med* 2014; **64:** 108–13.

12 Arbyn M, Verdoodt F, Snijders PJ, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. *Lancet Oncol* 2014; **15:** 172–83.

13 Belinson JL, Du H, Yang B, et al. Improved sensitivity of vaginal self-collection and high-risk human papillomavirus testing. *Int J Cancer* 2012; **130:** 1855–60.

14 Lam JUH, Rebolj M, Ejegod DM, et al. Prevalence of human papillomavirus in self-taken samples from screening nonattenders. *J Clin Microbiol* 2017; **55:** 2913–23.

15 Bosgraaf RP, Verhoef VM, Massuger LF, et al. Comparative performance of novel self-sampling methods in detecting high-risk human papillomavirus in 30,130 women not attending cervical screening. *Int J Cancer* 2015; **136:** 646–55.

16 Jacobs MV, Snijders PJ, van den Brule AJ, Helmerhorst TJ, Meijer CJ, Walboomers JM. A general primer GP5+/GP6(+)-mediated PCR-enzyme immunoassay method for rapid detection of 14 high-risk and 6 low-risk human papillomavirus genotypes in cervical scrapings. *J Clin Microbiol* 1997; **35:** 791–95.

17 Bulk S, Van Kemenade FJ, Rozendaal L, Meijer CJ. The Dutch CISOE-A framework for cytology reporting increases efficacy of screening upon standardisation since 1996. *J Clin Pathol* 2004; **57:** 388–93.

18 Hopman EH, Rozendaal L, Voorhorst FJ, Walboomers JM, Kenemans P, Helmerhorst TJ. High risk human papillomavirus in women with normal cervical cytology prior to the development of abnormal cytology and colposcopy. *BJOG* 2000; **107:** 600–04.

19 Oncoline. Guideline CIN, AIS and VAIN. 2015 (in Dutch). https://www.oncoline.nl/cin-ais-en-vain (accessed Sept 23, 2018).

20 Anderson B. Premalignant and malignant squamous lesions of the cervix. In: Fox H, Wells M, eds. Haines and Taylor obstetrical and gynaecological pathology, 4th edn. New York: Churchill Livingstone, 1995: 273–322.

21 Alonzo TA, Kittelson JM. A novel design for estimating relative accuracy of screening tests when complete disease verification is not feasible. *Biometrics* 2006; **62:** 605–12.

22 Casparie M, Tiebosch AT, Burger G, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. *Cell Oncol* 2007; **29:** 19–24.

23 Meijer CJ, Berkhof J, Castle PE, et al. Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. *Int J Cancer* 2009; **124:** 516–20.

24 Rijkaart DC, Berkhof J, Rozendaal L, et al. Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. *Lancet Oncol* 2012; **13:** 78–88.

25 Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* 2016; **6:** e012799.

26 Salmerón J, Lazcano-Ponce E, Lorincz A, et al. Comparison of HPV-based assays with Papanicolaou smears for cervical cancer screening in Morelos State, Mexico. *Cancer Causes Control* 2003; **14:** 505–12.

27 van de Ven PM, Berkhof J. Non-inferiority and superiority tests for partially matched data when a constraint is placed on the number of new tests. *Stat Med* 2013; **32:** 2515–28.

28 Naber SK, Matthijsse SM, Jansen EEL, de Kok IMCM, de Koning HJ, van Ballegooijen M. Effects of the new cervical screening programme in the Netherlands following recent developments. September, 2016 (in Dutch). https://www.rijksoverheid.nl/documenten/rapporten/2016/09/01/effecten-en-kosten-van-het-vernieuwde-bevolkingsonderzoek-baarmoederhalskanker (accessed Sept 23, 2018).

29 Giorgi Rossi P, Marsili LM, Camilloni L, et al. The effect of self-sampled HPV testing on participation to cervical cancer screening in Italy: a randomised controlled trial (ISRCTN96071600). *Br J Cancer* 2011; **104:** 248–54.

30 Sancho-Garnier H, Tamalet C, Halfon P, et al. HPV self-sampling or the Pap-smear: a randomized study among cervical screening nonattenders from lower socioeconomic groups in France. *Int J Cancer* 2013; **133:** 2681–87.

31 Broberg G, Gyrd-Hansen D, Miao Jonasson J, et al. Increasing participation in cervical cancer screening: offering a HPV self-test to long-term non-attendees as part of RACOMIP, a Swedish randomized controlled trial. *Int J Cancer* 2014; **134:** 2223–30.

32 Cadman L, Wilkes S, Mansour D, et al. A randomized controlled trial in non-responders from Newcastle upon Tyne invited to return a self-sample for Human Papillomavirus testing versus repeat invitation for cervical screening. *J Med Screen* 2015; **22:** 28–37.

33 van Zummeren M, Leeman A, Kremer WW, et al. Three-tiered score for Ki-67 and $p16^{ink4a}$ improves accuracy and reproducibility of grading CIN lesions. *J Clin Pathol* 2018; **71:** 981–88.