Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis

Marc Arbyn, Freija Verdoodt, Peter J F Snijders, Viola M J Verhoef, Eero Suonio, Lena Dillner, Silvia Minozzi, Cristina Bellisario, Rita Banzi, Fang-Hui Zhao, Peter Hillemanns, Ahti Anttila

Lancet Oncol 2014; 15: 172–83

Published Online
January 14, 2014
http://dx.doi.org/10.1016/S1470-2045(13)70570-9

See Comment page 128

Unit of Cancer Epidemiology, Scientific Institute of Public Health, Brussels, Belgium (M Arbyn DrTMH, F Verdoodt PhD); **Department of Pathology, VU University Medical Center, Amsterdam, Netherlands** (Prof P J F Snijders PhD, V M J Verhoef MD); **International Agency for Research on Cancer, Lyon, France** (E Suonio MD); **Karolinska Institute, Stockholm, Sweden** (L Dillner PhD); **Unit of Cancer Epidemiology, Department of Oncology, Piedmont Centre for Cancer Prevention, S Giovanni University Hospital, Turin, Italy** (S Minozzi MD, C Bellisario MSc); **IRCCS-Istituto di Ricerche Farmacologiche Mario Negri, Milan, Italy** (R Banzi PhD); **Department of Cancer Epidemiology, Cancer Institute and Hospital, Chinese Academy of Medical Sciences, Beijing, China** (Prof F-H Zhao PhD); **Department of Gynaecology and Obstetrics, Hannover Medical School, Hannover, Germany** (Prof P Hillemanns PhD); **and Finnish Cancer Registry, Helsinki, Finland** (A Anttila PhD)

Correspondence to:
Dr Marc Arbyn, Unit of Cancer Epidemiology, Scientific Institute of Public Health, J Wytsmanstreet 14, B1050 Brussels, Belgium
marc.arbyn@wiv-isp.be

Summary

Background Screening for human papillomavirus (HPV) infection is more effective in reducing the incidence of cervical cancer than screening using Pap smears. Moreover, HPV testing can be done on a vaginal sample self-taken by a woman, which offers an opportunity to improve screening coverage. However, the clinical accuracy of HPV testing on self-samples is not well-known. We assessed whether HPV testing on self-collected samples is equivalent to HPV testing on samples collected by clinicians.

Methods We identified relevant studies through a search of PubMed, Embase, and CENTRAL. Studies were eligible for inclusion if they fulfilled all of the following selection criteria: a cervical cell sample was self-collected by a woman followed by a sample taken by a clinician; a high-risk HPV test was done on the self-sample (index test) and HPV-testing or cytological interpretation was done on the specimen collected by the clinician (comparator tests); and the presence or absence of cervical intraepithelial neoplasia grade 2 (CIN2) or worse was verified by colposcopy and biopsy in all enrolled women or in women with one or more positive tests. The absolute accuracy for finding CIN2 or worse, or CIN grade 3 (CIN3) or worse of the index and comparator tests as well as the relative accuracy of the index versus the comparator tests were pooled using bivariate normal models and random effect models.

Findings We included data from 36 studies, which altogether enrolled 154 556 women. The absolute accuracy varied by clinical setting. In the context of screening, HPV testing on self-samples detected, on average, 76% (95% CI 69–82) of CIN2 or worse and 84% (72–92) of CIN3 or worse. The pooled absolute specificity to exclude CIN2 or worse was 86% (83–89) and 87% (84–90) to exclude CIN3 or worse. The variation of the relative accuracy of HPV testing on self-samples compared with tests on clinician-taken samples was low across settings, enabling pooling of the relative accuracy over all studies. The pooled sensitivity of HPV testing on self-samples was lower than HPV testing on a clinician-taken sample (ratio 0·88 [95% CI 0·85–0·91] for CIN2 or worse and 0·89 [0·83–0·96] for CIN3 or worse). Also specificity was lower in self-samples versus clinician-taken samples (ratio 0·96 [0·95–0·97] for CIN2 or worse and 0·96 [0·93–0·99] for CIN3 or worse). HPV testing with signal-based assays on self-samples was less sensitive and specific than testing on clinician-based samples. By contrast, some PCR-based HPV tests generally showed similar sensitivity on both self-samples and clinician-based samples.

Interpretation In screening programmes using signal-based assays, sampling by a clinician should be recommended. However, HPV testing on a self-sample can be suggested as an additional strategy to reach women not participating in the regular screening programme. Some PCR-based HPV tests could be considered for routine screening after careful piloting assessing feasibility, logistics, population compliance, and costs.

Funding The 7th Framework Programme of the European Commission, the Belgian Foundation against Cancer, the International Agency for Research on Cancer, and the German Guideline Program in Oncology.

Introduction

In view of the strong aetiological link between persistent infection with high-risk types of the human papillomavirus (HPV) and the development of cervical cancer, assays have been developed to detect nucleic acid sequences of the virus.[1,2] Meta-analyses have provided clinical evidence that has led to widely accepted recommendations to use HPV tests to triage women with equivocal cervical cytology and to predict recurrence after treatment of cervical precancer.[3–6] Data from randomised trials have consistently shown that women with a prior negative HPV test have a lower risk of developing grade 3 cervical intraepithelial neoplasia (CIN) and invasive cervical cancer compared with women with a prior normal Pap smear.[7] This new evidence underpins the recommendation to change the policy of secondary prevention of cervical cancer and to use an HPV assay as the primary screening test, used either alone instead of a Pap smear or together with a Pap test.[8,9] Moreover, HPV testing can be done on a vaginal sample taken by the women themselves, which might offer opportunities to reach those who are reluctant to undergo gynaecological examinations.[10,11]

Previous systematic reviews have summarised the performance of HPV testing on self-samples but these reviews were done 6–8 years ago and included mainly small studies, they assessed only virological outcomes, or they did not compare the accuracy for high-grade CIN

with tests on clinician-taken samples.[12–14] A review drew attention to the need for a comprehensive updated meta-analysis in view of the large amount of new data from large studies that have used a wide range of tests and collection devices.[10]

In this meta-analysis, we assessed the clinical accuracy of HPV testing on self-samples to detect underlying high-grade CIN or cancer. Moreover, we compared the accuracy of HPV testing in self-samples with that of HPV testing and cytology processing on samples taken by a clinician. We did not assess cytological processing of self-samples because it was previously shown to be poorly sensitive for picking up high-grade CIN lesions.[10,15,16]

We focused on primary screening for cervical cancer, because self-sampling is most often used in this type of setting and because the absolute accuracy might be different in women who are followed up because of prior cervical abnormalities. However, the relative accuracy of HPV testing in a sample taken by a woman compared with a sample taken by a health professional could be similar across settings. Assessment of the relative accuracy allowed us to include more studies, incorporate randomised trials, and increase the power for explaining heterogeneity about, in particular, the effects of collection devices and HPV assays. We aimed to find out whether an HPV test on a self-sample is as good as a test on a sample taken by a clinician in women attending cervical cancer screening.

Methods

Search strategy and selection criteria

We searched PubMed, Embase, and CENTRAL for eligible studies (see appendix for the clinical questions and search terms used). We searched for papers published between Jan 1, 1990, and June 3, 2013. We also used Scopus to investigate citations of previous systematic reviews on HPV testing of self-samples,[10,12–14] and the reference lists of selected references. Additionally, we searched for unpublished reports in the abstract books of the three most recent international conferences of the Papillomavirus Society (Montreal, July 3–8, 2010; Berlin, Sept 17–22, 2011, and San Juan, Nov 30–Dec 6, 2012). We applied no language restrictions.

Studies were eligible for inclusion if the following criteria were fulfilled: a vaginal sample was self-taken by a woman (self-sample) followed by a sample taken by a clinician (clinician-taken sample) or self-samples were taken in one arm and clinician-taken samples in the other arm of randomised trials; a high-risk HPV DNA or RNA test was done on both samples or the clinician-taken sample was examined microscopically for presence of cytological epithelial lesions; and the presence or absence of CIN grade 2 (CIN2) or worse was verified by colposcopy and biopsy in all enrolled women or in women with at least one positive test. Studies with cytological follow-up for women with negative colposcopy test results at baseline assessment were accepted as fulfilling the third selection criterion but were indexed for sensitivity analyses.

We contacted authors to provide accuracy data separately for the outcome CIN2 or worse when only the outcome CIN grade 3 (CIN3) or worse was reported.

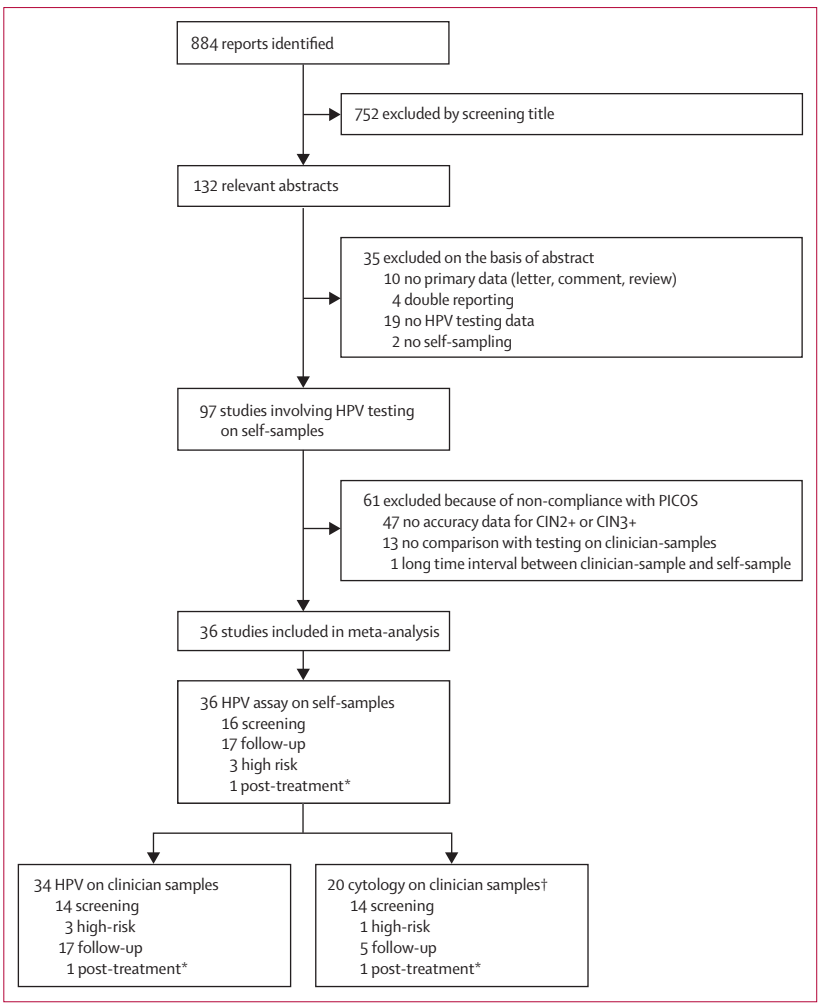

Figure 1: **Study flow chart**
CIN2 or worse=cervical intraepithelial neoplasia of grade 2 or worse. CIN3 or worse=cervical intraepithelial neoplasia of grade 3 or worse. HPV=human papillomavirus. PICOS=Population/Index/Comparator/Outcome/Study. *One study[48] included women in a follow-up setting and a post-treatment surveillance setting. †18 of 20 studies included both cytology and HPV testing on the clinician-taken samples, whereas two trials[47,50] included only cytology on the clinician-taken samples.

See **Online** for appendix

Clinical questions and data extraction

We aimed to answer two questions. The first was to establish the absolute accuracy of HPV testing on a self-sample (index test) and the accuracy of cytological processing or HPV testing on a cervical cell specimen taken by a clinician (comparator tests). The second was to establish the relative accuracy of HPV testing on a self-sample compared with the comparator tests on a clinician-taken sample. Accuracy was determined for the

	Patient selection		Index and comparator tests		Reference test			Flow and timing						Concerns of applicability: risk of bias		
	P1	P2	T1	T2	R1	R2	R3	F1	F2	F3	F4	F5	F6	Patient selection	Index and comparator test	Reference test
Morrison et al, 1992[30]	U	U	U	U	Y	U	Y	Y	Y	Y	Y	Y	Y	Moderate	Moderate	Low
Hillemanns et al, 1999[31]	U	U	U	U	Y	U	Y	U	Y	Y	U	N	N	Moderate	Moderate	Low
Sellors et al, 2000[32]	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Low	Low	Low
Wright et al, 2000[33]	U	U	Y	Y	Y	Y	Y	N	N	Y	N	N	N	Moderate	Low	Low
Belinson et al, 2001[34]	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Low	Low	Low
Lorenzato et al, 2002[35]	Y	Y	Y	U	Y	U	Y	Y	Y	Y	Y	Y	N	Low	Low	Low
Nobbenhuis et al, 2002[16]	Y	U	U	U	Y	U	Y	Y	Y	Y	Y	Y	Y	Low	Moderate	Low
Garcia et al, 2003[15]	Y	U	Y	U	Y	Y	Y	Y	Y	Y	N	Y	N	Low	Low	Low
Salmeron et al, 2003[36]	U	U	Y	U	Y	N	Y	U	N	Y	Y	Y	Y	Moderate	Low	Moderate
Brink, 2006[37]	U	U	Y	Y	Y	Y	Y	Y	N	Y	N	Y	N	Moderate	Low	Low
Daponte et al, 2006[38]	Y	U	U	U	Y	U	Y	U	Y	Y	U	N	N	Low	Moderate	Low
Girianelli et al, 2006[39]	U	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Moderate	Low	Low
Holanda et al, 2006[40]	U	U	Y	U	Y	Y	Y	Y	Y	Y	Y	N	N	Moderate	Low	Low
Seo et al, 2006[41]	U	U	Y	U	Y	U	Y	Y	Y	Y	N	N	N	Moderate	Low	Low
Szarewski et al, 2007[42]	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Low	Low	Low
Qiao et al, 2008[43]	Y*	U	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Low	Low	Low
Bhatla et al, 2009[44]	Y	U	Y	U	Y	U	Y	Y	Y	Y	Y	N	Y	Low	Low	Low
Balasubra et al, 2010[45]	Y	U	Y	U	Y	U	Y	Y	Y	Y	Y	Y	Y	Low	Low	Low
Gustavsson et al, 2011[46]	U	U	U	U	Y	Y	Y	Y	Y	Y	Y	N	N	Moderate	Moderate	Low
Lazcano-P, 2011[47]	Y	U	Y	U	Y	U	Y	U	N	Y	Y	N	N	Low	Low	Low
Taylor et al, 2011[48]	Y	U	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Low	Low	Low
Twu et al, 2011[49]	U	U	Y	U	Y	U	Y	U	N	Y	N	Y	Y	Moderate	Low	Low
Wikstrom et al, 2011[50]	Y	U	Y	U	Y	U	Y	N	N	Y	Y	N	N	Low	Low	Low
Belinson et al, 2012[51]	U	U	Y	U	Y	U	Y	Y	N	Y	Y	N	Y	Moderate	Low	Low
Dijkstra et al, 2012[52]	U	U	U	Y	Y	U	Y	Y	Y	Y	N	N	N	Moderate	Low	Low
Longatto-F et al, 2012[53]	Y	U	Y	Y	Y	N	Y	U	Y	Y	Y	U	U	Low	Low	Moderate
van Baars et al, 2012[54]	U	U	U	U	N†	U	U	Y	N	U	Y	N	N	Moderate	Moderate	High
Zhao et al, 2012‡[55]	Y	Y	Y	U	Y	U	Y	Y	Y	Y	Y	N	N	Low	Low	Low
Darlin et al, 2013[56]	U	U	U	U	Y	U	Y	Y	Y	Y	Y	Y	N	Moderate	Moderate	Low
Geraets et al, 2013[57]	Y	U	U	U	Y	U	Y	Y	Y	Y	N	N	N	Low	Moderate	Low
Guan et al, 2013[58]	N	U	U	U	Y	U	Y	Y	Y	Y	N	N	N	High	Moderate	Low
Jentschke et al, 2013a[60]	U	U	U	U	Y	U	Y	U	Y	Y	Y	N	N	Moderate	Moderate	Low
Jentschke et al, 2013b[59]	U	U	Y	U	Y	U	Y	U	Y	Y	Y	N	N	Moderate	Low	Low
Nieves et al, 2013[61]	Y	Y	U	Y	Y	N	Y	U	Y	Y	Y	Y	Y	Low	Low	Moderate

QUADAS items:[19] P1=acceptable enrolment method, P2=inappropriate exclusions avoided, T1=prespecified test cut-off, T2=results of the index test and the comparator are masked towards each other and both the index and comparator tests are masked towards the reference test, R1=acceptable reference test, R2=results of the reference test are masked towards the index and comparator tests, R3=incorporation bias avoided, F1=acceptable delay between triage tests and reference test, F2=partial verification avoided, F3=differential verification avoided, F4=withdrawals explained, F5=uninterpretable results reported for tests, F6=uninterpretable results reported for reference test. Each quality item is judged with the following: Y=fulfilled, U=unclear, and N=not fulfilled. *Participants recruited from randomly selected communes. †Only 44/134 women were verified histologically, the rest of the participants were verified cytologically. ‡Only data for SPOCCS III-1, III-2, and III-3 are included in this meta-analysis; data also reported jointly by Belinson et al.[64]

Table 1: **Quality assessment of all included studies**

disease outcomes CIN2 or worse or CIN3 or worse. We analysed the following groups: those attending routine cervical cancer screening, high-risk women, and those referred to colposcopy because of previous positive screening results. We designed a protocol in accordance with PRISMA guidelines for reporting of meta-analyses.[17,18]

Eligibility of studies for inclusion was checked independently by at least two investigators and, subsequently, the numbers of the true and false positives and negatives for each combination of sampling, test, and disease outcome were extracted for all studies which had complete verification of test results with a reference standard. For randomised trials comparing self-samples with clinician-taken samples in which verification was restricted to patients being positive for one test, we extracted only the number of true-positives in those screened. Data from these trials was only used in the pooling of the relative sensitivity. Information about study participants, setting, tests, sampling

	Test cutoff	Number of studies (number of test-device combinations)		Sensitivity (%; 95% CI)		Specificity (%; 95% CI)	
		CIN2 or worse	CIN3 or worse	CIN2 or worse	CIN3 or worse	CIN2 or worse	CIN3 or worse
Primary screening							
Self-sample							
HPV	As defined by manufacturer	14 (16)	6 (8)	76% (69–82)	84% (72–92)	86% (83–89)	87% (84–90)
Clinician-taken sample							
HPV	As defined by manufacturer	14 (16)	6 (8)	91% (87–94)	95% (91–97)	88% (85–91)	89% (87–92)
Cytology	ASC-US or worse	12	6	83% (75–89)	91% (85–95)	91% (87–94)	89% (86–91)
	LSIL or worse	8	5	71% (66–76)	78% (72–85)*	97% (97–98)	97% (96–97)*
Screening of high-risk group							
Self-sample							
HPV	As defined by manufacturer	3 (4)	1	75% (58–87)	42% (27–57)†	86% (77–92)	81% (76–87)†
Clinician-taken sample							
HPV	As defined by manufacturer	3 (4)	1	88% (78–93)	80% (67–93)†	88% (81–93)	82% (77–88)†
Cytology	ASC-US or worse	1	0	77% (64–91)†	..	87% (84–90)†	..
	LSIL or worse	1	0	70% (55–85)†	..	95% (93–97)†	..
Follow-up							
Self-sample							
HPV	As defined by manufacturer	17 (19)	5 (7)	84% (78–89)	85% (76–91)	56% (49–63)	45% (36–54)
Clinician-taken sample							
HPV	As defined by manufacturer	17 (19)	5 (7)	91% (86–94)	96% (92–1 00)*	58% (48–67)	46% (35–57)*
Cytology	ASC-US or worse	5	0	85% (77–91)	..	69% (57–80)	..

ASC-US=atypical squamous cells of undetermined significance. CIN2=cervical intraepithelial neoplasia of grade 2. CIN3=cervical intraepithelial neoplasia of grade 3. HPV=assay that picks up high-risk types of the human papillomavirus. LSIL=low-grade squamous intraepithelial lesion. *Sensitivity and specificity pooled separately because lack of fit of the bivariate normal model. †No meta-analytical pooling because only one study.

Table 2: **Absolute sensitivity and specificity of human papillomavirus testing on self-samples, and human papillomavirus testing and cytology on clinician-taken samples, by clinical setting and grade of cervical intraepithelial neoplasia**

devices, and verification by the reference standard was coded and compiled in a comprehensive table (appendix). We assessed the quality of the selected studies using the QUADAS-2 checklist.[19] To define test positivity of the HPV test, we accepted the cutoff proposed by the manufacturer, whereas for cytological tests we considered two cutoffs: atypical squamous cells of undetermined significance (ASC-US) or worse and low-grade squamous intraepithelial lesions (LSIL) or worse.[20]

Statistical analysis

The pooled absolute sensitivity and specificity of the tests were estimated jointly with metandi, a procedure in STATA, based on the bivariate normal model for the logit transforms of sensitivity and specificity, taking the intrinsic correlation between true-positive and false-positive rates and the variability between studies into account.[21,22] We computed the relative sensitivity and specificity of HPV testing on self-samples compared with cytology or HPV testing on clinician-taken samples using metadas, a SAS macro for meta-analysis of diagnostic accuracy studies that allows the inclusion of type of test as a covariate, making comparison of tests possible.[18,23] We also applied this model to assess the effect of study characteristics on the absolute test accuracy. We assessed heterogeneity in the relative sensitivity and specificity (testing on self-sample *vs* clinician-taken samples), in particular the effect of the HPV assays and sampling devices, separately with subgroup meta-analyses and sensitivity analyses.[24,25] We identified influential reports by repeating the meta-analysis and consecutively omitting each individual study.[26] We applied the effective sample-size funnel plot and associated regression test to assess publication bias or small study effects in the meta-analyses of the absolute accuracy.[27] To verify small-sample effects in the relative sensitivity and specificity of HPV testing in self-samples versus clinician-taken samples, we checked asymmetry in funnel plots visually and subsequently tested them statistically by a linear regression based on the efficient score and its variance.[28,29]

Statistical tests were two-sided and statistical significance was defined as p values of less than 0·05. We used STATA (version 10.1) and SAS (version 9.3) for statistical analyses.

Role of the funding source

The sponsor of the study had no role in study design, data collection, data analysis, data interpretation, or

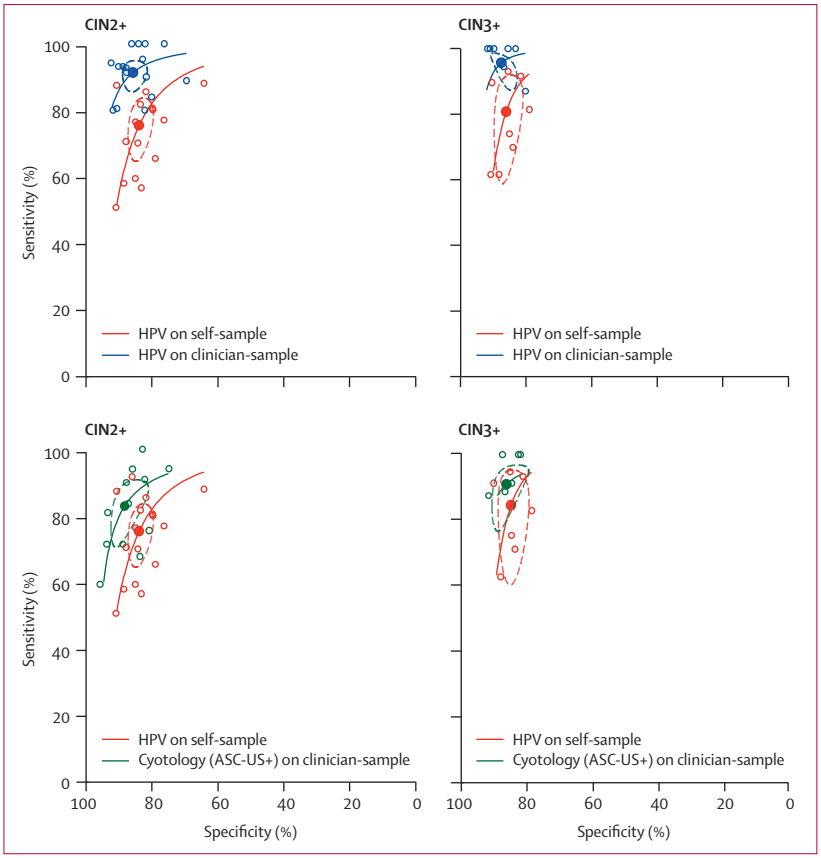

Figure 2: **The accuracy in primary cervical cancer screening, by collection method and grade of cervical intraepithelial neoplasia**

Hollow circles are individual studies, the full curved line is the summary ROC curve, solid circles are the pooled accuracy measures, surrounded by the 95% confidence ellipse (dashed line). CIN2+=cervical intraepithelial neoplasia of grade 2 or worse. CIN3+=cervical intraepithelial neoplasia of grade 3 or worse. HPV=human papillomavirus. ASC-US+=atypical squamous cells of undetermined significance or worse.

	Number of studies	Relative sensitivity (95% CI)	Relative specificity (95% CI)
HPV on self-samples *vs* cytology (ASC-US or worse) on clinician-taken samples			
Outcome of CIN2 or worse	19	0·95 (0·91–0·99)*	0·92 (0·90–0·94)*
Outcome of CIN3 or worse	6	0·99 (0·94–1·06)	0·98 (0·97–0·99)*
HPV on self-samples *vs* cytology (LSIL or worse) on clinician-taken samples			
Outcome of CIN2 or worse	11	1·14 (1·07–1·21)*	0·88 (0·86–0·90)*
Outcome of CIN3 or worse	6	1·19 (1·09–1·29)*	0·90 (0·87–0·94)*
HPV on self-samples *vs* HPV on clinician-taken samples			
Outcome of CIN2 or worse	34	0·88 (0·85–0·91)*	0·96 (0·95–0·97)*
Outcome of CIN3 or worse	12	0·89 (0·83–0·96)*	0·96 (0·93–0·99)*

ASC-US=atypical squamous cells of undetermined significance. CIN2=cervical intraepithelial neoplasia of grade 2. CIN3=cervical intraepithelial neoplasia of grade 3. HPV=assay that picks up high-risk types of the human papillomavirus. *Statistically significant (p<0·05).

Table 3: **Pooled relative sensitivity and specificity of human papillomavirus testing on self-samples versus cytology or human papillomavirus testing on clinician-taken samples, by grade of cervical intraepithelial neoplasia**

writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

We identified 884 articles, of which 34 were papers assessing the clinical accuracy of HPV DNA or RNA testing in self-samples (figure 1, appendix).[15,16,30–61] Because one paper contained three separate studies,[55] data from 36 studies were used in the meta-analysis, altogether including 154 556 women. In all but two selected studies,[47,50] the comparator test was HPV testing on a clinician-taken sample, whereas in 20 reports, the clinician-taken samples were examined cytologically.[15,16,32–34,36,37,39,42–44,47,48,50,51,53,55,61] We requested and obtained unreported accuracy data for the disease threshold CIN2 or worse from three studies.[51,55,61] 16 studies (from 14 papers) were in primary screening of generally healthy women,[33,34,36,39,40,42,43,47,50,51,53,55,58,61] whereas three studies screened high-risk populations.[35,44,45] In 17 reports, women referred to colposcopy were enrolled.[15,16,30–32,37,38,41,46,48,49,52,54,56,57,59,60] One study recruited women under follow-up and women who were treated for cervical precancer.[48] An overview of design, population, and test characteristics and an overview of HPV assays and sampling devices used, which are grouped into five categories (brush, lavage, spatula, swab, and tampon) are available in the appendix.

The methodological quality of the 36 included studies was, overall, moderate to good with average negative scores for the 13 QUADAS items varying between 0% and 31% (4/13), equivocal scores varying between 0% and 58% (7/13), and positive scores varying between 17% (2/13) and 100% (table 1). Risk of bias regarding enrolment of patients was low in 19 (53%), moderate in 16 (44%), and high in one (3%) of 36 studies. Reporting and execution of tests (description of cut-off or blinding of the index test towards comparator and reference test) was adequate in 26 (72%) studies and unclear in ten (28%) studies, but the risk of bias was never assessed as high. The quality of the verification with a reference standard (acceptable validity, blinding towards tests, avoidance of incorporating test results in final conclusion of disease outcome) was good in 32 (89%), moderate in three (8%), and possibly problematic in one (3%) of the 36 studies. The delay between self-sampling, clinician-sampling, and verification with the reference standard was short (6 months or less) in 25 (69%), unreported in nine (25%), and long in two (6%) studies. Partial verification was avoided in 28 (78%) studies but clearly present in eight (22%), whereas differential verification was absent in all but one study.[54] Withdrawal of patients was explained appropriately in 25 (69%) but not in nine (25%) studies. In most studies, uninterpretable results were poorly reported (20 [56%] studies for the assessed tests and 22 [61%] studies for the reference standard).

Table 2 shows the pooled absolute sensitivity and specificity of HPV DNA testing on self-samples and clinician-taken samples and of cytology at cutoff ASC-US or worse and LSIL or worse on clinician-taken samples for detecting underlying CIN2 or worse and CIN3 or worse for each setting separately. Because of the substantial and

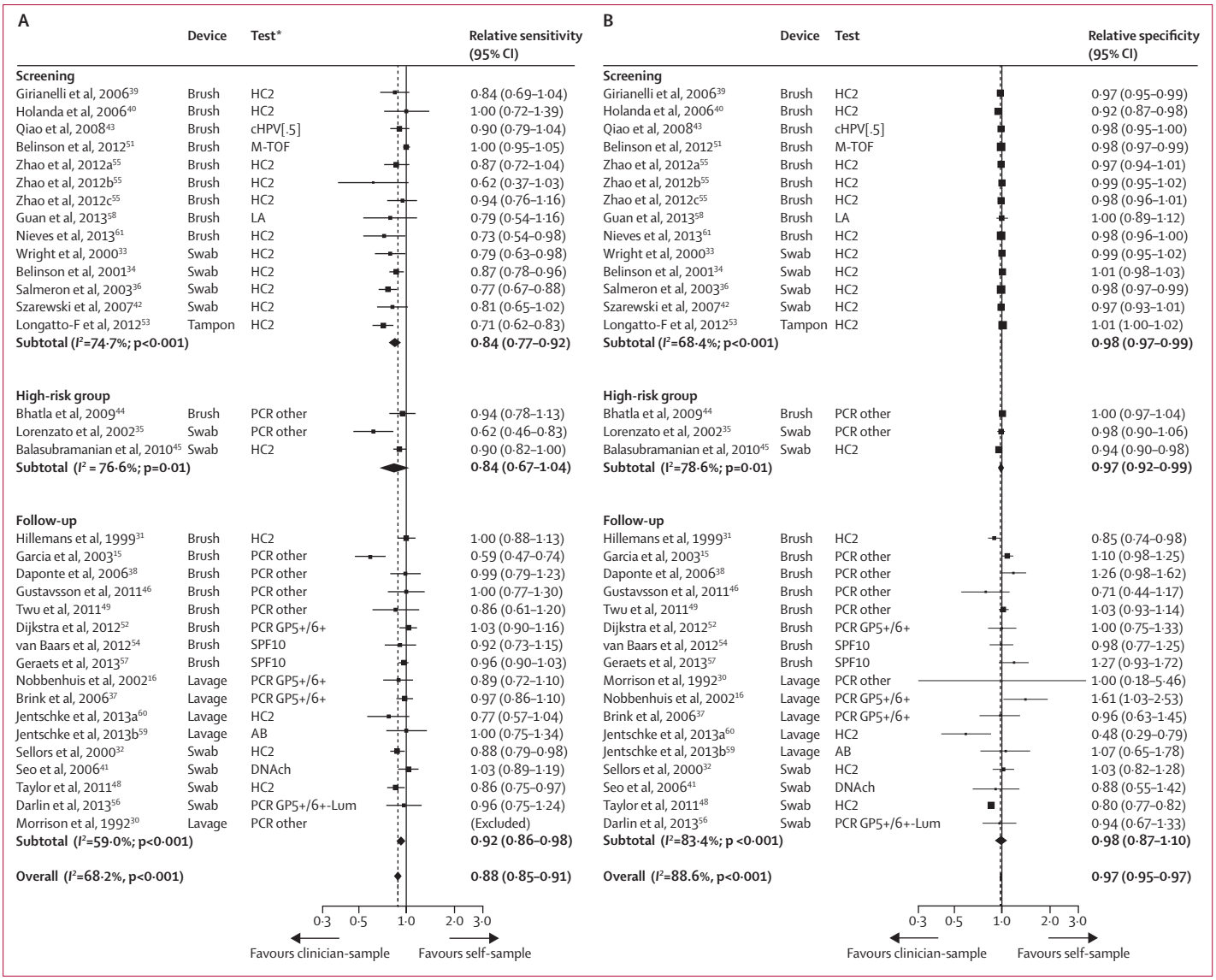

Figure 3: **Relative sensitivity (A) and specificity (B) of human papillomavirus on self-samples versus clinician-taken samples, by clinical setting, for outcome CIN2 or worse**
*See appendix for full test names.

significant variation by setting (p<0·0001; data not shown), we computed no overall pooled accuracy values. In primary screening, the sensitivity of HPV testing on self-samples ranged between 51%[61] and 93%[51] for CIN2 or worse and between 63%[61] and 94%[51] for CIN3 or worse. The pooled sensitivity estimates were 76% (95% CI 69–82) for CIN2 or worse and 84% (72–92) for CIN3 or worse (figure 2, table 2). The specificity for excluding CIN2 or worse ranged between 67%[40] and 93%.[55] HPV testing on clinician-taken samples showed high pooled sensitivity estimates for both CIN2 or worse and CIN3 or worse, and the pooled specificity for excluding CIN2 or worse was 88% (85–91). The sensitivity of cytological testing at cut-off ASC-US or worse on clinician-taken samples was between that of HPV testing in self-samples and clinician-taken samples (table 2). Cytology at the cutoff LSIL or worse showed the lowest pooled sensitivity (for CIN2 or worse), but the highest pooled specificity (for excluding CIN2 or worse; table 2).

The accuracy estimates for CIN2 or worse of HPV testing in self-samples and clinician-taken samples pooled from four studies that included high-risk women were similar to those in primary screening (table 2). However, the accuracy of cytology and of HPV testing for CIN3 or worse varied from the pooled estimates in primary screening, but data in the high-risk testing group were based on only one study.

The pooled sensitivities for CIN2 or worse derived from follow-up studies were not substantially different from those seen in primary screening. HPV testing of

	Number of studies (number of test-device combinations)	Relative sensitivity	Relative specificity
Test			
HC2	18	0·85 (0·81–0·90)*	0·96 (0·93–0·98)*
PCR GP5+/6+	5	0·95 (0·89–1·01)	1·11 (0·95–1·29)
CareHPV (at RLU≥0.5)	1	0·90 (0·79–1·04)	0·98 (0·95–1·00)
CareHPV (at RLU≥1)	1	0·86 (0·73–1·03)	1·00 (0·98–1·02)
PCR-SPF10	2	0·96 (0·89–1·02)	1·10 (0·85–1·41)
Abbott Real Time hrHPV Test	1	1·00 (0·75–1·34)	1·07 (0·65–1·78)
Cervista	1	0·76 (0·70–0·83)*	0·95 (0·94–0·96)*
APTIMA	1	0·64 (0·46–0·90)*	0·99 (0·98–1·01)
DNAchip	1	1·03 (0·89–1·19)	0·88 (0·55–1·42)
Modified GP5+/6+ PCR with Luminex reading	1	0·96 (0·75–1·24)	0·94 (0·67–1·33)
Linear Array	1	0·79 (0·54–1·16)	1·00 (0·89–1·12)
MALDI-TOF	1	1·00 (0·95–1·05)	0·98 (0·97–0·99)*
Other nonGP5+/6+ PCR	7	0·82 (0·66–1·01)	1·02 (0·97–1·07)
Collection device for self-sampling			
Brush	18 (24)	0·89 (0·83–0·94)*	0·98 (0·97–0·99)*
Lavage	5 (6)	0·94 (0·85–1·03)	0·95 (0·68–1·34)
Swab	10	0·86 (0·80–0·92)*	0·95 (0·90–1·01)
Tampon	1	0·71 (0·62–0·83)*	1·01 (1·00–1·02)*
Income status of country			
Low-income and middle-income countries	19 (23)	0·85 (0·79–0·91)*	0·97 (0·95–0·99)*
High-income countries	14 (17)	0·94 (0·90–0·97)*	0·99 (0·93–1·06)

RLU=relative light units. *Statistically significantly different from unity (1 excluded from 95% CI).

Table 4: **Subgroup meta-analysis of the relative sensitivity and specificity of human papillomavirus (self-sampled *vs* clinician sampled) testing for cervical intraepithelial neoplasia of grade 2 or worse, by covariate**

self-samples showed a higher pooled estimate (84% [78–89]), but the difference was not significant (p=0·23). However, the pooled specificity in follow-up settings was substantially and significantly (p<0·0001) lower than the specificity in screening (table 2).

Table 3 shows the pooled relative accuracy measures of HPV testing on self-samples versus the comparator tests on clinician-taken samples. From studies reporting results of multiple tests, we included the combination that yielded the highest relative sensitivity for CIN2 or worse in the meta-analysis under the condition that for a given study the HPV assay always had to be the same in self-samples and clinician-taken samples. The variation of the relative sensitivity and specificity of HPV testing on self-samples versus clinician-taken samples for detection of CIN2 or worse of all the included studies, grouped by clinical setting, is shown in figure 3. We identified substantial heterogeneity in all studies (p<0·001 for all studies). However, we detected no significant heterogeneity in relative sensitivity by clinical setting (p=0·345; data not shown). The variation in relative specificity by setting was also small but statistically significant (p=0·001; data not shown) due to large numbers. Because of the restricted inter-setting variation in the relative accuracy, we did meta-analytical pooling across all clinical settings combined.

HPV testing on self-samples was less sensitive and less specific than cytology with ASC-US or worse as a cut-off on clinician-taken samples with respect to detection of CIN2 or worse (table 3, appendix). However, for the detection of CIN3 or worse, HPV testing on self-samples was as sensitive as ASC-US or worse cytology on clinician specimen (table 3, appendix).

HPV testing on self-samples was more sensitive but less specific than cytology using LSIL or worse as a cut-off for CIN2 or worse and CIN3 or worse (table 3, appendix).

HPV testing on self-samples was, on average, less sensitive than HPV testing on clinician-taken samples, for both CIN2 or worse (ratio 0·88) and CIN3 or worse (ratio 0·89; table 3, appendix). Additionally, on self-samples, HPV testing was less specific than on clinician-taken samples in excluding CIN2 or worse (table 3, figure 3).

The variation in the absolute sensitivity and specificity of HPV testing on self-samples to detect CIN2 or worse in primary screening, according to study covariates is shown in the appendix. We saw a statistically significantly higher sensitivity when PCR with MALDI-TOF was used to identify HPV in self-samples than when the Hybrid Capture 2 (HC2) assay (ratio sensitivity of HC2 in self-samples/sensitivity of MALDI-TOF in self-samples for CIN2 or worse, 1·25 [95% CI 1·12–1·39]). The APTIMA test showed a significantly higher specificity than HC2 (ratio 1·04 [1·02–1·07]). Issues of study design or quality of reporting did not alter the clinical accuracy at the exception of one QUADAS item: when the delay between tests was not clearly reported, the sensitivity was significantly lower than when it was clearly reported. We detected no evidence of an effect of involvement of manufacturers of tests or sampling devices on the accuracy estimates.

HPV testing on self-samples was significantly less sensitive and also less specific than such testing on clinician-taken samples to detect CIN2 or worse when HC2 or Cervista were used (table 4). APTIMA testing was less sensitive but not less specific in self-samples versus clinician-taken samples. The sensitivity was similar in both types of samples when testing with GP5+/6+ PCR, SPF10 PCR, Abbott Real Time hrHPV test, DNAchip, modified GP5+/6+ PCR with Luminex reading, or MALDI-TOF. HPV testing on self-samples was less specific than testing on clinician-taken samples when HC2, Cervista, or MALDI-TOF were used. Covariate analyses for sampling device, type of reference test, or completeness of verification generally showed lower sensitivity of HPV testing on self-samples versus such testing on clinician-taken samples (table 4). This finding can be explained by the use of HC2 or other less sensitive HPV assays in most studies (table 5 for device, data not shown for other covariates).

Table 5 shows relative sensitivity and specificity values for tests clinically validated for primary cervical cancer screening on clinician-collected cervical scrapes.[7,62] When

HC2 was used, self-sampling with any device showed lower sensitivity with differences being statistically significant for brushes, swabs, and tampons. With GP5+/6+ PCR, the sensitivity of HPV testing on self-samples collected with a brush or lavage device was similar to that seen with a clinician-collected brush sample (table 5). Also, in the only available study,[59] where the Abbott RT HPV PCR, a GP5+/6+ derivative, was applied, no difference was seen between testing on a self-collected lavage sample and a clinician-collected brush sample.

No differences by country status (high-income *vs* low-income or middle-income) could be discerned in the relative sensitivity. However, the relative specificity was significantly lower than unity in middle-income or low-income countries but not in the only screening study done in a high-income country (table 4).[42]

The effective size funnel plot and associated regression test did not show statistically significant small-study effects in the absolute accuracy of HPV testing on self-samples, or in that of any of the comparator tests on clinician-taken samples (appendix).

We detected no statistically significant asymmetry in the relative specificity of HPV testing in self-sample versus comparator tests (p always >0·10; appendix). However, we detected a statistically significant small-study effect in the relative sensitivity of HPV testing on self-samples versus HPV testing (for CIN2 or worse and CIN3 or worse) or cytology at cut-off ASC-US (for CIN3 or worse) on clinician-taken samples (appendix).

	Number of studies	Relative sensitivity	Relative specificity
HC2			
Brush[31,39,40,44,55,61]	8	0·89 (0·82–0·98)†	0·97 (0·96–0·99)†
Lavage[59,60]	2	0·82 (0·65–1·02)	0·68 (0·35–1·33)
Swab[32-34,36,42,45,48]	7	0·82 (0·86–0·90)†	0·95 (0·89–1·01)
Tampon[53]	1	0·71 (0·62–0·83)†	1·01 (1·00–1·02)†
GP5+/6+ PCR			
Brush[52,54,57]	3	0·95 (0·86–1·04)	1·08 (0·93–1·25)
Lavage[16,37]	2	0·95 (0·85–1·06)	1·23 (0·74–2·05)
Abbott Real Time hrHPV Test			
Lavage[59]	1	1·00 (0·75–1·34)	1·07 (0·65–1·78)

*Restricted to three tests clinically validated for primary cervical cancer screening using samples taken by a clinician.[7] †Statistically significantly different from unity (95% CI excludes 1).

Table 5: **Subgroup meta-analyses of the relative sensitivity and specificity of human papillomavirus (self-sampled *vs* clinician sampled) testing* for cervical intraepithelial neoplasia of grade 2 or worse, by self-collection device**

Discussion

Our findings suggest that screening with an HPV test on self-sampled material can detect, on average, 76% of CIN2 or worse and 84% of CIN3 or worse. The pooled specificity to exclude CIN2 or worse was estimated at 86%. Because the absolute accuracy varied by clinical setting, these values include only primary screening studies. However, the variation of the relative sensitivity and specificity of HPV testing on self-samples compared with tests on clinician-taken samples was low across settings. Such low variation enabled us to pool screening, testing in high-risk populations, and follow-up studies in the meta-analyses of the relative accuracy. The pooled sensitivity of HPV testing on self-samples was statistically significantly lower than HPV testing on a clinician-taken sample. Furthermore, HPV testing on self-samples was also 4% less specific for exclusion of CIN2 or worse. Compared with cytological processing of a cervical cell specimen collected by a health professional and using ASC-US as a cutoff, HPV testing on self-samples was slightly less sensitive and clearly less specific than testing on clinician-taken samples. However, when the threshold for positive cytology was LSIL, HPV testing on a self-sample was 14% more sensitive in detecting CIN2 or worse and 19% more sensitive in detecting CIN3 or worse, but was 12% less specific in excluding CIN2 or worse.

We could discern no obvious collection device effects in the relative sensitivity of HPV testing on self-samples versus clinician-taken samples. Because the included studies did not address comparisons between different sampling devices, and because of the variability in study design and settings, we can draw no strong conclusions about the influence of devices for self-sampling on the accuracy estimates. However, our findings did show obvious test effects. Sensitivity for HPV testing on self-samples was lower than HPV testing on clinician-taken samples when HC2, APTIMA, or Cervista were used. For HC2, this lower sensitivity was consistent for all types of devices separately and could be pooled from 18 studies, whereas findings for Cervista and APTIMA were based on findings from only single studies. Moreover, we saw lower specificity on self-samples for HC2 and Cervista but not for APTIMA. PCR amplification with the GP5+/6+ primers (in five studies[16,37,52,54,57]) and the Abbott RT hrHPV Test (in one study[59]) showed similar sensitivity and specificity on self-samples versus clinician-taken samples. The MALDI-TOF assay was used in only one study,[51] but that study enrolled more than 8000 women attending cervical cancer screening among whom 141 women with CIN3 or worse and 233 women with CIN2 or worse were identified. In this large study, the investigators could assess non-inferiority of MALDI-TOF testing on self-tests versus MALDI-TOF testing on clinician samples, with high precision: relative sensitivity of 1·00 for CIN2 or worse and 1·01 for CIN3 or worse with a lower 95% CI bound greater than 0·90 (p for non-inferiority of <0·0001). MALDI-TOF testing on self-samples showed a small but statistically significant loss in specificity (ratio of 0·98). Some other PCR-based HPV assays also showed a sensitivity which was not statistically significantly lower in self-samples but the precision of equivalency was low (unity included in the 95% CI but lower bound <0·90).

Overall, HPV testing on self-samples was equally sensitive compared with cytology at cutoff ASC-US to detect CIN3 or worse but less sensitive for CIN2 or worse. However, this lower pooled sensitivity for CIN2 or worse was driven by one study with very low sensitivity of GPMY09/11 PCR amplification (49%) on self-samples taken with an endocervical brush in three colposcopy clinics in the USA, Peru, and Mexico.[15] Omission of these outlying findings resulted in a relative sensitivity of 0·99 (95% CI 0·96–1·03) and a relative specificity of 0·97 (0·94–1·00). When positive cytology on clinician-taken samples was defined as LSIL or worse, HPV testing on self-samples was more sensitive in the detection of CIN2 or worse and CIN3 or worse. However, this finding was affected by the outlying high relative sensitivity noted in a large community-based randomised trial,[47] done in Mexico, which compared HC2 on self-samples taken with a conical brush with a conventional Pap smear prepared by a health professional (relative sensitivity of 3·41 for CIN2 or worse and 3·20 for CIN3 or worse).

Although we often judged study design and quality to be non-optimal or insufficiently documented, QUADAS items did not explain the heterogeneity seen in the absolute or relative accuracy estimates (appendix).

In meta-analyses assessing test performance, the pooled parameters are sensitivity and specificity rather than predictive values, which are strongly affected by the background prevalence of disease.[63] However, for a given region or country, compared with sensitivity and specificity values, predictive values are more informative for clinical practice and decision making. We have therefore applied the pooled absolute accuracy estimates of tests on three screening situations with low (0·25%), medium (0·5%), and high (2%; seen in some high-income countries) prevalence of CIN3 or worse to compute post-test probabilities of disease (appendix), in which we regard a risk lower than 1% after a negative test as reassuring, whereas a risk greater than 10% after a positive test would indicate a referral as being appropriate. A negative screening test always yielded a low post-test risk (<1%) of underlying CIN3 or worse but the future risk, considered over a period of 5 years, exceeded 1% after a negative HC2 test on self-sample and after a negative cytology result on a clinician sample in high-risk populations. The cross-sectional positive predictive values on self-samples and clinician-taken samples were usually lower than 10% in low-risk and medium-risk populations, indicating the need for a triage test before referral to colposcopy for diagnostic work-up or treatment. In high-risk populations, the post-test probability of CIN3 or worse was higher than 10% when using cytology or hrHPV testing on clinician samples on a self-sample and also for MALDI-TOF, but not for HC2, on a self-sample.

In terms of the biological plausibility of our findings, the lower clinical sensitivity of HPV testing with HC2, Cervista, or APTIMA on self-samples can be explained by lower loads of high-risk HPV DNA from virus particles in the vagina that are beneath the detection threshold of respective assays, but which can be detected by more analytically sensitive PCR tests.[51,64] Additionally, the lower specificity of HC2, Cervista, and MALDI-TOF might be partly attributable to the increased presence of cross-reacting low-risk HPV types in the vagina but also to presence of high-risk HPV particles in the vagina that have not caused CIN2 or worse.[64]

Our meta-analysis included 36 studies that, together, enrolled more than 150 000 women, potentially allowing for robust estimates with use of statistical procedures recommended by the Cochrane Collaboration. However, only 14 studies took place in a typical screening situation, where self-sampling is of interest. Nonetheless, the relative accuracy comparing testing in self-samples versus clinician-taken samples did not vary substantially by clinical setting, which justified pooling all the studies, and which enabled analysis of the effect of multiple sources of heterogeneity. However, the assessment of the effect of covariates was restricted by paucity of detail in reported aggregated data and sometimes yielded findings on categories based on few or only one study.

Our meta-analysis assessed test accuracies but did not assess the effect of self-sampling on the incidence of cervical cancer (or precancer) in a population. However, if HPV testing on a specimen collected by a woman is analysed with an assay showing similar accuracy on self-samples and clinician samples, it would be likely to result also in reduced incidence of cervical cancer as already shown in randomised trials in which samples were taken by clinicians.[7] The use of an imperfect gold standard based on colposcopy and colposcopy-targeted biopsies could have generated bias. Nonetheless, our main conclusion that HPV testing on self-samples samples was less sensitive than HPV testing on clinician-taken samples was seen for all types of reference testing, including taking random cervical biopsies.

Our findings draw attention to the need for well-designed studies to assess the accuracy of combinations of HPV assays and self-sampling devices. A colposcopy clinic where both a self-sample and clinician-taken sample are obtained, followed by verification by colposcopy and biopsies, offers a valid design in which the absolute sensitivity, as well as the relative sensitivity and specificity, can be assessed without bias on a rather small study sample and which will be also relevant for a primary screening setting. The absolute specificity assessed in a colposcopy clinic is irrelevant for primary screening. Population screening trials can address questions of screening coverage, follow-up compliance of screen-positive women, and positive predictive value for high-grade CIN, but might not enable assessment of absolute specificity. However, the approximate absolute specificity can be assessed without any bias and without need for verifying women with negative screening tests.[65] Particular attention should be given to the assessment of HPV testing on self-samples in low-income and middle

income countries that do not have the infrastructure for cytopathological assessment and have a substantial burden of cervical cancer—in such countries, self-sampling could enable good screening coverage. Longitudinal follow-up studies are also needed. Experts and methodologists should work out well-designed protocols for further studies and define minimal accuracy thresholds that have to be fulfilled to accept a given combination of test and self-collection device in screening as done for screening tests on clinician-taken samples.[62]

Findings from randomised trials published over the past 6 years have shown that HPV screening followed by cytology triage results in a lower incidence of cancer than with cytology screening.[7,66] Because cytology triage is not possible on self-samples, an adequate molecular reflex test should be developed to triage women with HPV. Hyper-methylation of some viral or human genes involved in carcinogenesis have shown promising accuracy profiles and could be applicable on self-samples, but needs further validation.[67,68]

In a screening programme using an HPV assay based on signal-amplification or RNA detection, samples taken by a medical professional rather than self-samples should be chosen, given the superior clinical sensitivity and specificity on clinician-taken samples. In a cytology-based or HPV-based screening programme, HPV-testing on a self-sample can be suggested as an additional strategy to reach women not participating in the regular screening programme. HPV testing on self-samples can also be considered in areas lacking high-quality cytopathology laboratories and where self-sampling could achieve good attendance. Before making a decision on whether to introduce strategies that use self-sampling, adequate transport of samples and communication—with women with a positive self-test being referred to further management—should be assured. In the future, some HPV tests that amplify viral DNA sequences from self-samples might reach similar accuracy as from clinician samples. Such tests might be preferred if validated, available, affordable, and feasible. However, before deciding on HPV screening using self-samples instead of clinician-collected samples, a careful pilot study should assess the feasibility, the clinical accuracy of the combination of the proposed test and the self-collection device, as well as the costs, logistics, and population compliance.

Contributors

MA, PJFS, F-HZ, and AA designed the study, formulated the clinical question, and identified the PICOS components. SM, CB, RB, FV, and MA did the literature searches. MA designed the data extraction form. MA, ES, PS, VMJV, LD, and FV extracted the data. MA and FV did the statistical analyses. MA wrote the paper. FV, PS, VMJV, ES, LD, SM, CB, RB, F-HZ, PH, and AA critically revised subsequent drafts.

Conflicts of interest

PJFS has received speaker's fees from Roche, Abbott, Qiagen, and Gen-Probe, and is shareholder of Self-screen. PH received research grants or lecture fees from Abbott, Hologic, and Roche. MA, FV, VMJV, ES, LD, SM, CB, RB, F-HZ, and AA declare that they have no conflicts of interest.

Acknowledgments

This study was funded by the seventh Framework Program of DG Research of the European Commission (PREHDICT Project [grant number 242061] and CoheaHr Project [603019], coordinated by the VU University Medical Center of Amsterdam, the Netherlands), the Belgian Foundation against Cancer (Brussels, Belgium), the International Agency for Research on Cancer (IARC, Lyon, France), and the German Guideline Program in Oncology (Hannover, Germany). We thank J L Belinson and F-H Zhao for the contribution of additional data and V Liausson for high-quality eps files for figures 2 and 3.

References

1 Cogliano V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, and the WHO International Agency for Research on Cancer. Carcinogenicity of human papillomaviruses. *Lancet Oncol* 2005; **6:** 204.

2 IARC Monograph Working Group. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol 90: Human Papillomaviruses. Lyon: IARC Press, 2007.

3 Arbyn M, Sasieni P, Meijer CJ, Clavel C, Koliopoulos G, Dillner J. Chapter 9: clinical applications of HPV testing: a summary of meta-analyses. *Vaccine* 2006; **24** (suppl 3)**:** S3, 78–89.

4 European Commission. European Guidelines for Quality Assurance in Cervical Cancer Screening. Luxembourg: Office for Official Publications of the European Communities; 2008.

5 Wright TC Jr, Massad LS, Dunton CJ, Spitzer M, Wilkinson EJ, Solomon D, and the 2006 American Society for Colposcopy and Cervical Pathology-sponsored Consensus Conference. 2006 consensus guidelines for the management of women with cervical intraepithelial neoplasia or adenocarcinoma in situ. *Am J Obstet Gynecol* 2007; **197:** 340–45.

6 Jordan J, Arbyn M, Martin-Hirsch P, et al. European guidelines for quality assurance in cervical cancer screening: recommendations for clinical management of abnormal cervical cytology, part 1. *Cytopathology* 2008; **19:** 342–54.

7 Arbyn M, Ronco G, Anttila A, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. *Vaccine* 2012; **30** (suppl 5)**:** F88–999.

8 Saslow D, Solomon D, Lawson HW, et al, and the ACS-ASCCP-ASCP Cervical Cancer Guideline Committee. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin* 2012; **62:** 147–72.

9 Arbyn M, de Sanjosé S, Saraiya M, et al. EUROGIN 2011 roadmap on prevention and treatment of HPV-related disease. *Int J Cancer* 2012; **131:** 1969–82.

10 Snijders PJ, Verhoef VM, Arbyn M, et al. High-risk HPV testing on self-sampled versus clinician-collected specimens: a review on the clinical accuracy and impact on population attendance in cervical cancer screening. *Int J Cancer* 2013; **132:** 2223–36.

11 Racey CS, Withrow DR, Gesink D. Self-collected HPV testing improves participation in cervical cancer screening: a systematic review and meta-analysis. *Can J Public Health* 2013; **104:** e159–66.

12 Ogilvie GS, Patrick DM, Schulzer M, et al. Diagnostic accuracy of self collected vaginal specimens for human papillomavirus compared to clinician collected human papillomavirus specimens: a meta-analysis. *Sex Transm Infect* 2005; **81:** 207–12.

13 Petignat P, Faltin DL, Bruchim I, Tramèr MR, Franco EL, Coutlée F. Are self-collected samples comparable to physician-collected cervical specimens for human papillomavirus DNA testing? A systematic review and meta-analysis. *Gynecol Oncol* 2007; **105:** 530–35.

14 Stewart DE, Gagliardi A, Johnston M, et al, and the HPV Self-collection Guidelines Panel. Self-collected samples for testing of oncogenic human papillomavirus: a systematic review. *J Obstet Gynaecol Can* 2007; **29:** 817–28.

15 Garcia F, Barker B, Santos C, et al. Cross-sectional study of patient- and physician-collected cervical cytology and human papillomavirus. *Obstet Gynecol* 2003; **102:** 266–72.

16 Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, et al. Primary screening for high risk HPV by home obtained cervicovaginal lavage is an alternative screening tool for unscreened women. *J Clin Pathol* 2002; **55:** 435–39.

17 Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; **151:** 264–69.

18 Diagnostic Test Accuracy Working Group. Handbook for Diagnostic Test Accuracy Reviews 2011. http://srdta.cochrane.org/handbook-dta-reviews (accessed Dec 1, 2013).

19 Whiting PF, Rutjes AW, Westwood ME, et al, and the QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155:** 529–36.

20 Solomon D, Davey D, Kurman R, et al, and the Forum Group Members, and the Bethesda 2001 Workshop. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 2002; **287:** 2114–19.

21 Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007; **8:** 239–51.

22 Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *Stata J* 2009; **9:** 211–29.

23 Takwoingi Y. Diagnostic Test AccuracyWorking Group. METADAS: A SAS macro for meta-analysis of diagnostic accuracy studies. http://srdta.cochrane.org/sites/srdta.cochrane.org/files/uploads/MetaDAS%20Quick%20Reference%20v1.3%20May%202012.pdf (accessed Dec 1, 2013).

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

25 Harris R, Bradburm M, Deeks J, et al. Metan: fixed- and random-effects meta-analysis. *Stata J* 2008; **8:** 3–26.

26 Tobias A. Assessing the influence of a single study in the meta-analysis estimate. *Stata Tech Bull* 1998; **8:** 108–10.

27 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol* 2005; **58:** 882–93.

28 Harbord RM, Egger M, Sterne JA. A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Stat Med* 2006; **25:** 3443–57.

29 Harbord RM, Harris RJ, Sterne JAC. Updated tests for small-study effects in meta-analyses. *Stata J* 2009; **9:** 197–210.

30 Morrison EAB, Goldberg GL, Hagan RJ, Kadish AS, Burk RD. Self-administered home cervicovaginal lavage: a novel tool for the clinical-epidemiologic investigation of genital human papillomavirus infections. *Am J Obstet Gynecol* 1992; **167:** 104–07.

31 Hillemanns P, Kimmig R, Hüttemann U, Dannecker C, Thaler CJ. Screening for cervical neoplasia by self-assessment for human papillomavirus DNA. *Lancet* 1999; **354:** 1970.

32 Sellors JW, Lorincz AT, Mahony JB, et al. Comparison of self-collected vaginal, vulvar and urine samples with physician-collected cervical samples for human papillomavirus testing to detect high-grade squamous intraepithelial lesions. *CMAJ* 2000; **163:** 513–18.

33 Wright TC Jr, Denny L, Kuhn L, Pollack A, Lorincz A. HPV DNA testing of self-collected vaginal samples compared with cytologic screening to detect cervical cancer. *JAMA* 2000; **283:** 81–86.

34 Belinson J, Qiao YL, Pretorius R, et al. Shanxi Province Cervical Cancer Screening Study: a cross-sectional comparative trial of multiple techniques to detect cervical neoplasia. *Gynecol Oncol* 2001; **83:** 439–44.

35 Lorenzato FR, Singer A, Ho L, et al. Human papillomavirus detection for cervical cancer prevention with polymerase chain reaction in self-collected samples. *Am J Obstet Gynecol* 2002; **186:** 962–68.

36 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.

37 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.

38 Daponte A, Pournaras S, Mademtzis I, et al. Evaluation of HPV 16 PCR detection in self- compared with clinician-collected samples in women referred for colposcopy. *Gynecol Oncol* 2006; **103:** 463–66.

39 Girianelli VR, Thuler LC, Szklo M, et al. Comparison of human papillomavirus DNA tests, liquid-based cytology and conventional cytology for the early detection of cervix uteri cancer. *Eur J Cancer Prev* 2006; **15:** 504–10.

40 Holanda F Jr, Castelo A, Veras TM, de Almeida FM, Lins MZ, Dores GB. Primary screening for cervical cancer through self sampling. *Int J Gynaecol Obstet* 2006; **95:** 179–84.

41 Seo SS, Song YS, Kim JW, Park NH, Kang SB, Lee HP. Good correlation of HPV DNA test between self-collected vaginal and clinician-collected cervical samples by the oligonucleotide microarray. *Gynecol Oncol* 2006; **102:** 67–73.

42 Szarewski A, Cadman L, Mallett S, et al. Human papillomavirus testing by self-sampling: assessment of accuracy in an unsupervised clinical setting. *J Med Screen* 2007; **14:** 34–42.

43 Qiao YL, Sellors JW, Eder PS, et al. A new HPV-DNA test for cervical-cancer screening in developing regions: a cross-sectional study of clinical accuracy in rural China. *Lancet Oncol* 2008; **9:** 929–36.

44 Bhatla N, Dar L, Patro AR, et al. Can human papillomavirus DNA testing of self-collected vaginal samples compare with physician-collected cervical samples and cytology for cervical cancer screening in developing countries? *Cancer Epidemiol* 2009; **33:** 446–50.

45 Balasubramanian A, Kulasingam SL, Baer A, et al. Accuracy and cost-effectiveness of cervical cancer screening by high-risk human papillomavirus DNA testing of self-collected vaginal samples. *J Low Genit Tract Dis* 2010; **14:** 185–95.

46 Gustavsson I, Sanner K, Lindell M, et al. Type-specific detection of high-risk human papillomavirus (HPV) in self-sampled cervicovaginal cells applied to FTA elute cartridge. *J Clin Virol* 2011; **51:** 255–58.

47 Lazcano-Ponce E, Lorincz AT, Cruz-Valdez A, et al. Self-collection of vaginal specimens for human papillomavirus testing in cervical cancer prevention (MARCH): a community-based randomised controlled trial. *Lancet* 2011; **378:** 1868–73.

48 Taylor S, Wang C, Wright TC, Denny L, Kuhn L. A comparison of human papillomavirus testing of clinician-collected and self-collected samples during follow-up after screen-and-treat. *Int J Cancer* 2011; **129:** 879–86.

49 Twu NF, Yen MS, Lau HY, Chen YJ, Yu BK, Lin CY. Type-specific human papillomavirus DNA testing with the genotyping array: a comparison of cervical and vaginal sampling. *Eur J Obstet Gynecol Reprod Biol* 2011; **156:** 96–100.

50 Wikström I, Lindell M, Sanner K, Wilander E. Self-sampling and HPV testing or ordinary Pap-smear in women not regularly attending screening: a randomised study. *Br J Cancer* 2011; **105:** 337–39.

51 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.

52 Dijkstra MG, Heideman DA, van Kemenade FJ, et al. Brush-based self-sampling in combination with GP5+/6+-PCR-based hrHPV testing: high concordance with physician-taken cervical scrapes for HPV genotyping and detection of high-grade CIN. *J Clin Virol* 2012; **54:** 147–51.

53 Longatto-Filho A, Naud P, Derchain SF, et al. Performance characteristics of Pap test, VIA, VILI, HR-HPV testing, cervicography, and colposcopy in diagnosis of significant cervical pathology. *Virchows Arch* 2012; **460:** 577–85.

54 van Baars R, Bosgraaf RP, ter Harmsel BW, et al Dry storage and transport of a cervicovaginal self-sample by use of the Evalyn Brush, providing reliable human papillomavirus detection combined with comfort for women. *J Clin Microbiol* 2012; **50:** 3937–43.

55 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.

56 Darlin L, Borgfeldt C, Forslund O, Hénic E, Dillner J, Kannisto P. Vaginal self-sampling without preservative for human papillomavirus testing shows good sensitivity. *J Clin Virol* 2013; **56:** 52–56.

57 Geraets DT, van Baars R, Alonso I, et al. Clinical evaluation of high-risk HPV detection on self-samples using the indicating FTA-elute solid-carrier cartridge. *J Clin Virol* 2013; **57:** 125–29.

58 Guan Y, Gravitt PE, Howard R, et al. Agreement for HPV genotyping detection between self-collected specimens on a FTA cartridge and clinician-collected specimens. *J Virol Methods* 2013; **189:** 167–71.

59 Jentschke M, Soergel P, Hillemanns P. Evaluation of a multiplex real time PCR assay for the detection of human papillomavirus infections on self-collected cervicovaginal lavage samples. *J Virol Methods* 2013; **193:** 131–34.

60 Jentschke M, Lange V, Soergel P, Hillemanns P. Enzyme-linked immunosorbent assay for p16(INK4a) – a new triage test for the detection of cervical intraepithelial neoplasia? *Acta Obstet Gynecol Scand* 2013; **92:** 160–64.

61 Nieves L, Enerson CL, Belinson S, et al. Primary cervical cancer screening and triage using an mRNA human papillomavirus assay and visual inspection. *Int J Gynecol Cancer* 2013; **23:** 513–18.

62 Meijer CJLM, 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.

63 Diagnostic Test Accuracy Working Group. Handbook for diagnostic test accuracy reviews. Birmingham: Cochrane Collaboration, DTA Working Group, 2013. http://srdta.cochrane.org/handbook-dta-reviews (accessed Jan 2, 2014).

64 Belinson JL, Hu S, Niyazi M, et al. Prevalence of type-specific human papillomavirus in endocervical, upper and lower vaginal, perineal and vaginal self-collected specimens: implications for vaginal self-collection. *Int J Cancer* 2010; **127:** 1151–57.

65 Arbyn M, Ronco G, Cuzick J, Wentzensen N, Castle PE. How to evaluate emerging technologies in cervical cancer screening? *Int J Cancer* 2009; **125:** 2489–96.

66 Ronco G, Giorgi-Rossi P, Carozzi F, et al, and the New Technologies for Cervical Cancer screening (NTCC) Working Group. Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. *Lancet Oncol* 2010; **11:** 249–57.

67 Snellenberg S, De Strooper LM, Hesselink AT, et al. Development of a multiplex methylation-specific PCR as candidate triage test for women with an HPV-positive cervical scrape. *BMC Cancer* 2012; **12:** 551.

68 Wentzensen N, Sun C, Ghosh A, et al. Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J Natl Cancer Inst* 2012; **104:** 1738–49.