UK National Screening Committee

Cervical cancer screening - HPV self sampling

External review of key questions relating to screening programme implementation for the UK National Screening Committee (UK NSC)

Version: Second Draft

Version date: 14th March 2017

Costello Medical Consulting Ltd.

March 2017

The UK NSC advises Ministers and the NHS in all 4 UK countries about all aspects of screening policy. Its policies are reviewed on a 3 yearly cycle. Current policies can be found in the policy database at http://legacy.screening.nhs.uk/screening-recommendations.php and the policy review process is described in detail at https://www.gov.uk/guidance/evidence-and-recommendations-nhs-population-screening#evidence-review-process

Template v2.0, August 2016

1.1 Objectives

1.1.1 Use of self-sampling to improve screening uptake

Cervical screening coverage has fallen in recent years, especially amongst younger women. Since 2004, uptake has dipped below 80% and 5-year coverage among women aged 25 to 29 is now below two-thirds.[19] The NCSP would like to pilot the use of self-sampling as a means of improving screening uptake.

This review was part of a larger piece of work and the sections relevant to self sampling have been extracted from the larger document.

Question 4 aims to establish whether self-collected specimens are of comparable accuracy to clinician-collected specimens, and **Question 5** investigates whether inviting unscreened women to return a self-collected specimen increases overall uptake of screening.

1.1.2 Identified evidence

The specific questions addressed in this review are shown in Table 1 below, along with the relationship of each question to the UK NSC's Screening Criteria and the number of studies that were identified as providing relevant evidence for each question.

Table 1 Key questions for the evidence summary, and relationship to UK NSC screening criteria

Criterion		Key questions	# studies included
THE TEST			
4	There should be a simple, safe, precise and validated screening test.	Question 4: What is the accuracy of HPV testing in self-collected specimens?	1 systematic literature review (SLR) 1 primary study
6	The test, from sample collection to delivery of results, should be acceptable to the target population.	Question 5: Does self-collection of vaginal specimens increase uptake of cervical screening?	1 SLR 10 primary studies

1.2 Methods

The current review was conducted by Costello Medical Consulting, in collaboration with the UK National Screening Committee. Database searches were conducted on 20th October 2016.

2 Synthesis of evidence

2.1 Overall results

Database searches yielded 204 results, of which 12 records were judged to be relevant to this review. An additional record was pre-specified for inclusion at the start of the review so 13 articles were ultimately included.

A study-level summary of data extracted from each included publication is presented in Appendix 4. Results of the quality assessments are also presented in Appendix 4.

2.2 Use of self-sampling to improve screening uptake

2.2.1 Use of self-sampling to improve screening uptake

2.2.1.1 Question 4 – What is the accuracy of HPV testing in self-collected specimens?

Criterion 4 of the UK NSC Screening Criteria states that: 'There should be a simple, safe, precise and validated screening test.'

This review looked for prospective studies which directly compared the accuracy of HPV testing on clinician-collected or self-collected samples.

2.2.1.1.1 Description of the evidence

One systematic review and meta-analysis, Arbyn 2014,[35] and a single additional study[36] have been identified which assess the accuracy of HPV testing on self-collected samples.

The systematic review and meta-analysis examined 36 studies, with a combined population of 154,556 women, considering the comparative accuracy of self-collected and clinician-collected self-sampling for HPV testing. The review considered the accuracy of screening in 3 population groups: women in a 'healthy screening population' attending for cervical cancer screening, high-risk women, and women in a 'follow-up' population who had been referred for colposcopy. The population meeting the eligibility criteria for this review is the 'healthy screening population'. This group was assessed in 16 of the 36 studies included in Arbyn 2014, with some outcomes having been reported separately for this population.[35]

The additional primary study identified in this review, Stanczuk 2016, assessed the performance of a polymerase chain reaction (PCR)-based assay for the detection of HPV in self-collected vaginal and urine samples. The study investigated a cohort of 5,318 women attending routine screening in a primary care setting.[36]

2.2.1.1.2 Quality assessment

An assessment of the methodological quality of the SLR using the AMSTAR checklist demonstrated overall good quality, with only one of the 11 checklist questions not addressed.[35]

Arbyn 2014 assessed the quality of its studies using the QUADAS-2 checklist and reported overall moderate to good quality in its studies (Table 2).[35] An assessment of the quality of the Stanczuk 2016 publication,[36] using the same checklist, determined a low risk of bias and low concerns of applicability relating to participant selection, index tests and reference standards in the study. The risk of bias with regards to participant flow was assessed as being high because not all participants had available results for each test, and not all participants received colposcopy (Table 2).

Table 2 Quality assessment (QUADAS-2) of studies included in Arbyn 2014 and Stanczuk 2016

Study reference	Arbyn 2014[35]	Stanczuk 2016[36]
PARTICIPANT SELECTION		
Risk of bias	Moderate	Low
Concern about applicability		Low
INDEX TESTS		
Risk of bias	Low	Low
Concern about applicability		Low
REFERENCE STANDARD		
Risk of bias	Low	Low
Concern about applicability		Low
PARTICIPANT FLOW		
Risk of bias	Moderate	High

Participant selection

A low risk of bias with regards to participant selection is reported for 53% of all included studies in the SLR and meta-analysis, while a medium risk of bias is reported for 44%. There are concerns regarding the applicability of this SLR to this review question due to the inclusion of participants in a 'high-risk screening population' and women who had been called for follow-up. These populations do not fit the eligibility criteria for this review; however, absolute accuracy values are reported separately for each group, including for women in the healthy screening population of interest to this review. Furthermore, while some outcomes are reported as pooled results across all included studies, it was demonstrated that the variability in results across studies in different populations was very low, which reduces concerns about the risk of bias in the meta-analysis.[35]

A low risk of bias for participant selection was determined for Stanczuk 2016. The study assessed women attending routine screening in Scotland, excluding only women who had previously been diagnosed with CIN2+, which is considered to be an appropriate exclusion. The screening population in Scotland is slightly younger than in England, with 97% of participants aged between 20 and 59 years old, however, the risk of bias associated with this is judged to be very low.[36]

Index test

No studies in the SLR were assessed to have a high risk of bias with regards to the reporting or execution of index tests. The approach was considered adequate in 72% of studies and in 28% it was unclear.[35]

The screening in Stanczuk 2016 was undertaken in accordance with the UK Cervical Cancer Screening Programme and, therefore, is applicable to this review. The publication does not report the threshold for a positive HPV result, however, this is not considered a serious quality concern and generally the execution of index test was considered appropriate, resulting in an overall low risk of bias.[36]

Reference standard

The quality of test verification with a reference standard is reported in Arbyn 2014 to be good in 89% of studies, moderate in 8% of studies and possibly problematic in one study. The SLR required studies to have used either colposcopy, considered to be the gold standard, or biopsy as the reference standard and to assess either CIN2+ or CIN3+ as the target abnormality. These eligibility criteria are aligned with those applied in this review, and as a result, the studies identified are highly applicable.[35]

The overall risk of bias associated with the reference standard was assessed to be low in Stanczuk 2016. The study uses colposcopy as the reference standard and detects both CIN2+ and CIN3+. The publication does not report whether the reference standard results are interpreted without knowledge of the index tests results, which has potential to cause bias in test verification. However, in line with the approach taken in Arbyn 2014 with regards to unclear blinding, this domain was judged to have overall low risk of bias.[36]

Participant flow

A moderate risk of bias associated with participant flow was reported in Arbyn 2014 on assessment of all included studies in the SLR. The delay between self-sampling, clinician-sampling and verification was determined to be short (<6 months) in 69% of included studies, unreported in 9% and long in 6% of studies. Partial verification was avoided in 78% of studies and differential verification avoided in all but one study.[35] Arbyn 2014 noted that when the delay between tests was not reported the sensitivity was significantly lower than when it was clearly reported.[35]

The recall time for colposcopy referral was not reported in Stanczuk 2016, and additionally, not all patients enrolled in the trial received a reference standard or were included in analyses. Due to these concerns the study was determined to have a high risk of bias associated with participant flow.[36]

2.2.1.1.3 Results

Accuracy of testing on self-collected and clinician-collected samples were reported as pooled results from the Arbyn 2014 meta-analysis,[35] and additionally in Stanczuk 2016.[36] Study-level details of these results are presented in Appendix 4, Table 15. A summary of the results is presented below (Table 3).

Table 3 Summary of accuracy results from the Arbyn 2014 meta-analysis and Stanczuk 2016

	Arbyn 2014[a, 35]				Stanczuk 2016[36]		
		Clinician-collected			Self-collected		
	Self-collected	HPV	Cytology ASC-US+	Cytology LSIL+	Vaginal sample, n=5208	Urine sample, n=5003	Clinician-collected, n=5299
CIN2+							
Studies, n	16	16	12	8	NA		
Sensitivity (95% CI)	76% (69 to 82)	91% (87 to 94)	83% (75 to 89)	71% (66 to 76)	94.6% (90.7 to 98.5)	63.1% (54.6 to 71.1)	97.7% (95.0 to 100)
Specificity (95% CI)	86% (83 to 89)	88% (85 to 91)	91% (87 to 94)	97% (97 to 98)	85.4% (84.4 to 86.3)	59.8% (89.0 to 90.7)	87.3% (86.4 to 88.2)
LR+	NR	NR	NR	NR	6.48	1.57	7.69
LR-	NR	NR	NR	NR	0.054	0.62	0.026
CIN3+							
Studies, n	8	8	6	5	NA		
Sensitivity (95% CI)	84% (72 to 92)	95% (91 to 97)	91% (85 to 95)	78% (72 to 85)	95.8% (91.1 to 100)	50.7% (39.1 to 62.3)	98.6% (95.9 to 100)
Specificity (95% CI)	87% (84 to 90)	89% (87 to 92)	89% (86 to 91)	97% (96 to 97)	84.8% (83.8 to 85.8)	89.7% (88.8 to 90.5)	86.4% (85.5 to 87.3)
LR+	NR	NR	NR	NR	6.30	4.92	7.25
LR-	NR	NR	NR	NR	0.045	0.49	0.016

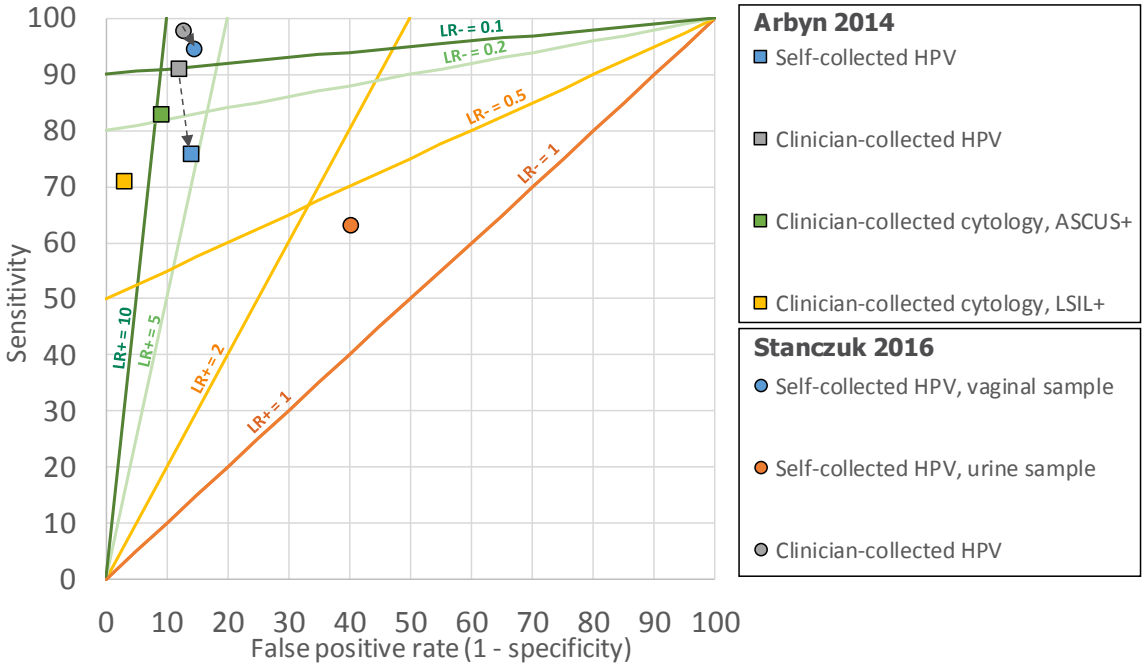
ASC-US, Atypical Cells of Undetermined Significance; CIN, Cervical Intraepithelial Neoplasia; LSIL, Low grade Squamous Intraepithelial Neoplasia

[a] Pooled values across studies in the systematic review enrolling participants considered part of a healthy screening population

Accuracy

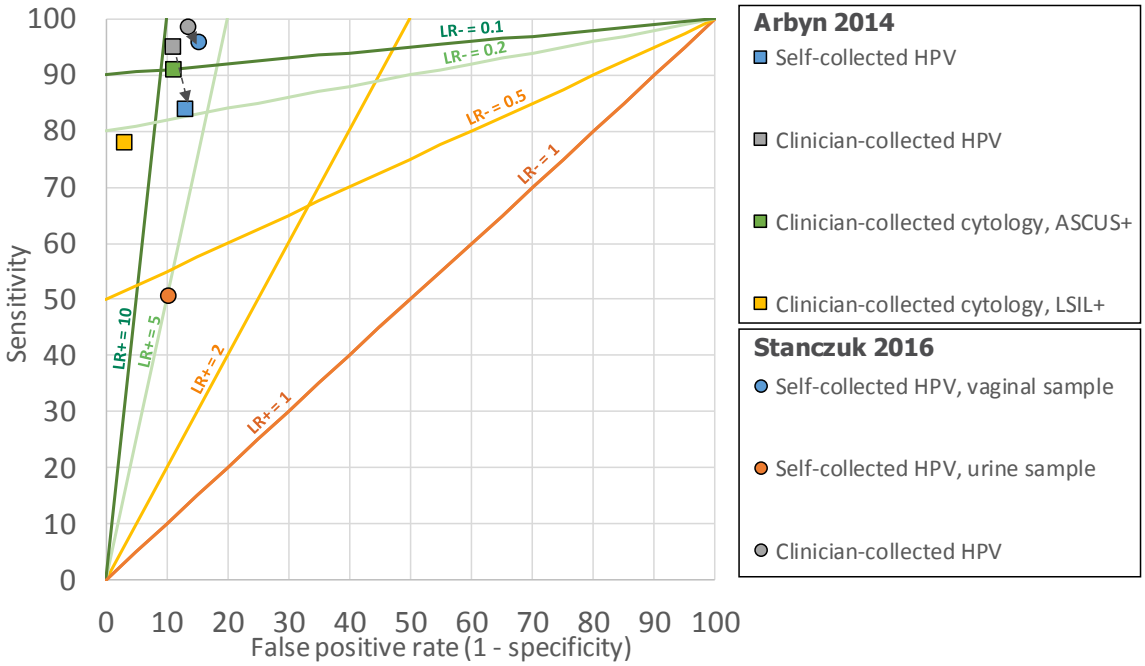
The following figures present the accuracy for self-collected and clinician-collected HPV samples pooled across the studies considering a healthy screening population in the Arbyn 2014 meta-analysis.[35] The figures additionally present the absolute accuracy of HPV testing on the 2 self-collection methods and clinician-collected samples determined in Stanczuk 2016.[36] The accuracy is presented separately for the detection of CIN2+ (Figure 1) and CIN3+ (Figure 2). The pooled clinician-collected cytology results reported in Arbyn 2014 are presented alongside the self-collected HPV test results to allow comparison to current clinical practice.[35]

Figure 1 Accuracy of screening methods for the detection of CIN2+



ASC-US, Atypical Cells of Undetermined Significance; CIN, Cervical Intraepithelial Neoplasia; HPV, Human Papilloma Virus; LSIL, Low grade Squamous Intraepithelial Neoplasia

Figure 2 Accuracy of screening methods for the detection of CIN3+



ASC-US, Atypical Cells of Undetermined Significance; CIN, Cervical Intraepithelial Neoplasia; HPV, Human Papilloma Virus; LSIL, Low grade Squamous Intraepithelial Neoplasia

Relative accuracy

The relative sensitivity and specificity results for detection of CIN2+ (Figure 3 and Figure 4) and CIN3+ (Figure 5 and Figure 6) with HPV testing are presented below. As discussed previously, the results from Arbyn 2014 were pooled across all 36 included studies, including those in high-risk and follow-up populations. A small variability in the results between different groups was demonstrated, and as a result, a low risk of bias is associated with this meta-analysis.[35] The results for Stanczuk 2016 present the accuracies of the 2 self-collection methods relative to the clinician-collected sample.[36]

Figure 3 Relative sensitivity of screening methods for detection of CIN2+

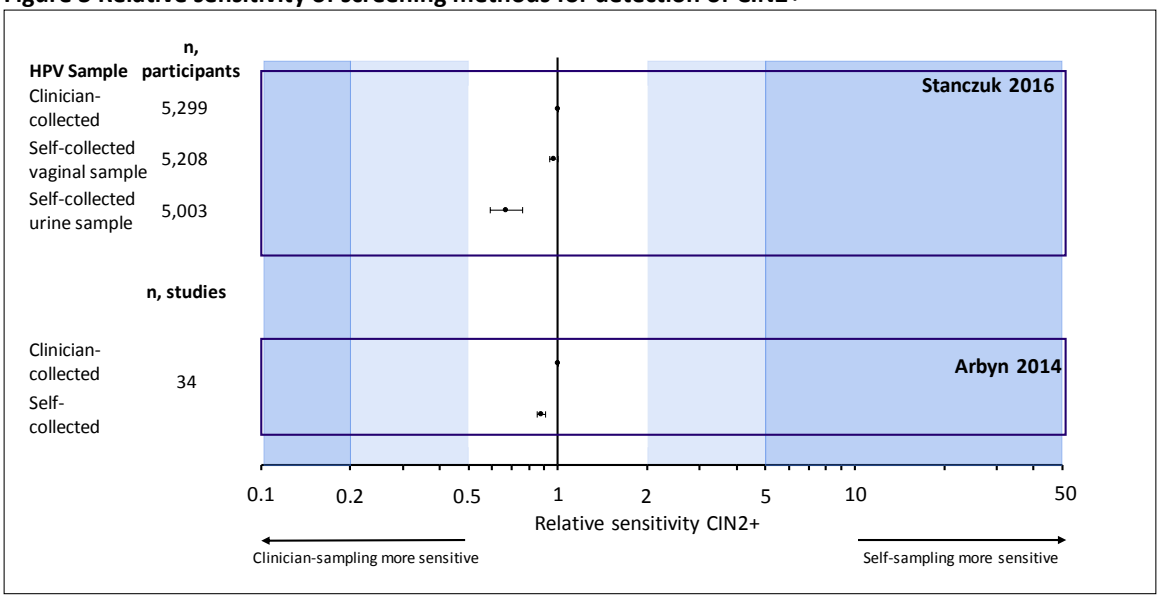


Figure 4 Relative specificity of screening methods for detection of CIN2+

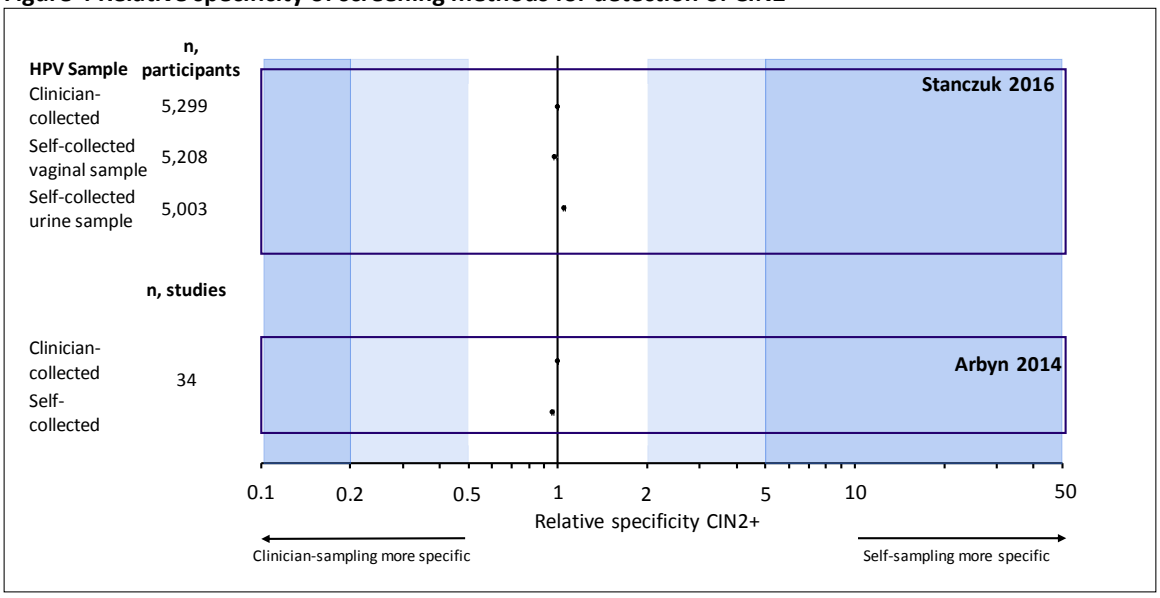


Figure 5 Relative sensitivity of screening methods for detection of CIN3+

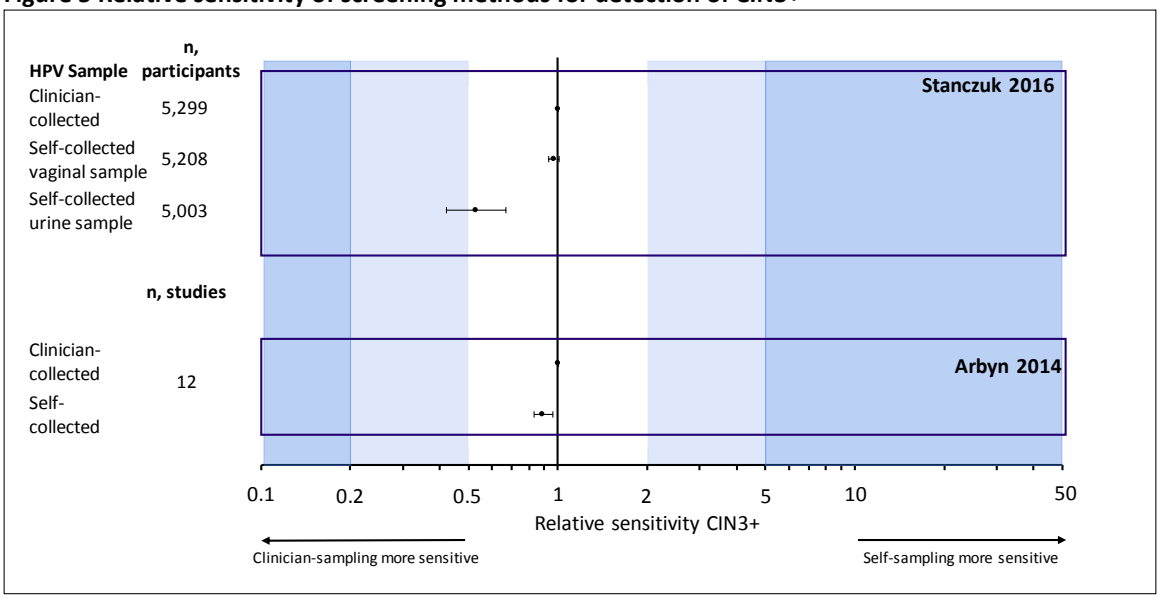
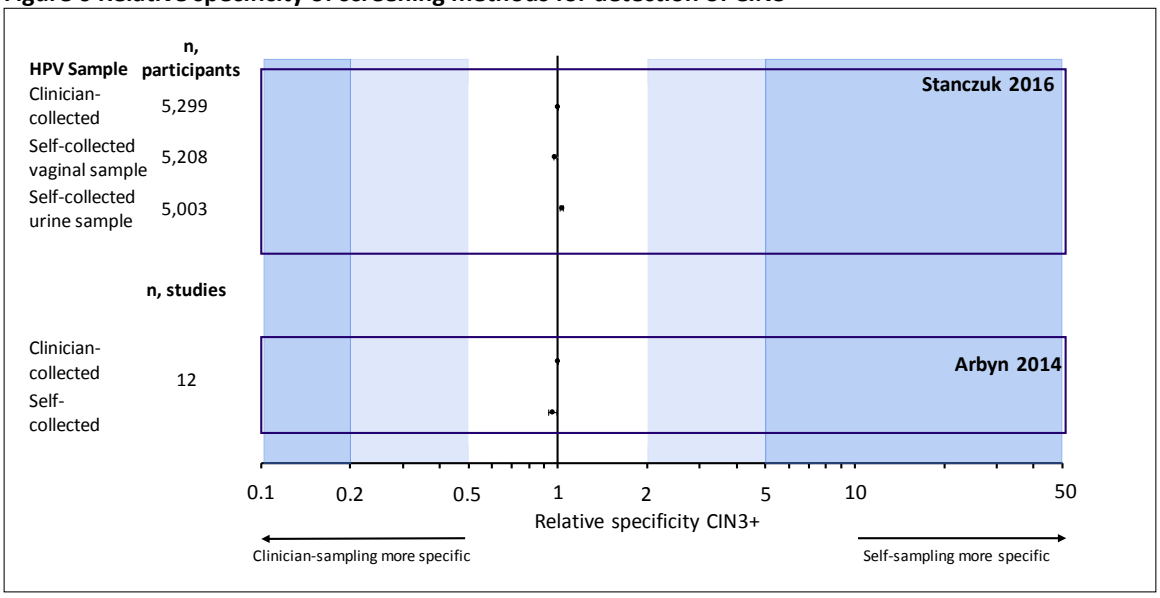


Figure 6 Relative specificity of screening methods for detection of CIN3+



The pooled relative accuracy results presented in the meta-analysis demonstrate that the sensitivity and specificity of HPV testing for detection of CIN2+ and CIN3+ on self-collected samples compared to clinician-collected samples are statistically significantly lower, but the difference is small. The pooled absolute sensitivity reported in Arbyn 2014 for the detection of CIN2+ was 76%, with results from individual studies ranging from 51% to 93%, within studies in a healthy screening population. The meta-analysis additionally reported a pooled absolute specificity of 86%, ranging from 67% to 93%. For the detection of CIN3+ the pooled results demonstrated a sensitivity of 84%, ranging from 63% to 94%, and a specificity of 87%.[35] While broad ranges in accuracy results are presented, Arbyn 2014 reported that issues relating to study design or reporting quality have not altered the accuracy of results in the meta-analysis. A single exception is that significantly lower sensitivity and specificity values are reported when

the delay between tests is not clearly reported. Furthermore, the variation between study results can partially be rationalised by heterogeneity in the HPV test used.

The absolute sensitivity values reported in Stanczuk 2016 for HPV testing on self-collected vaginal samples are within the range reported for similar studies in Arbyn 2014, with sensitivity for detection of CIN2+ and CIN3+ of 94.6% (95% CI 90.7% to 98.5%) and 95.8% (95% CI 91.1% to 100%) respectively. The absolute specificity observed was in agreement with the values from the meta-analysis. HPV testing on self-collected vaginal samples for the detection of both CIN2+ and CIN3+ was considered as sensitive and as specific as the testing on clinician-collected samples.[36] The relatively high sensitivity values may be explained by the use of a PCR detection method in Stanczuk 2016, which is known to be more analytically sensitive than other tests, such as the Hybrid Capture 2 (HC2) HPV test which was used in the majority of studies identified in Arbyn 2014.[37]

Overall, the accuracy results reported in Arbyn 2014 should be considered at low risk of bias. The SLR was well conducted, and identified studies which are highly relevant to this review in terms of study population and reference testing.[35] The results are also supported by an additional study identified in this review.[36]

Further points to note are considerations made in Arbyn 2014 regarding self-sampling devices and HPV test methods. No significant differences were observed between the different self-sampling devices used across included studies. The assay used in the vast majority of studies identified in the SLR was HC2. Using HC2 as a reference, generally no significant differences were observed to other HPV tests. There were just 2 exceptions: a significantly higher sensitivity using a 'MALDI-TOF' test; and significantly higher specificity using an APTIMA HPV test.[35]

2.2.1.1.4 Evidence summary

This review considers the accuracy of HPV testing on self-collected samples based on evidence from a good quality SLR[35] which has been demonstrated to have a good applicability to this review and results from one further primary study.[36]

Evidence across these publications suggests that the accuracy of HPV testing in self-collected samples is lower than in clinician-collected samples, but not substantially lower, and could be considered an appropriate alternative to clinician-sampling for women not attending primary screening.[35, 36] The high relative accuracy results reported in Stanczuk 2016 on self-collected vaginal samples highlight the potential for optimisation of testing by varying the HPV assay,[36] a factor which could warrant further investigation.

Analysis of evidence relevant to criterion 4 – Accuracy of HPV testing on self-collected specimens

Quantity: Overall the evidence identified comprises one SLR[35] and one additional primary study in a cohort of 5,318 women.[36] The SLR reports 36 primary studies, 16 of which are of high relevance to this review. The total number of studies represent a reasonable evidence base to consider the accuracy of HPV testing on self-collected samples.

Quality: An assessment of the methodological quality of the SLR demonstrated that it had overall good quality, with only one question in the AMSTAR checklist not having been addressed.[35] The SLR reported the quality of its included studies as moderate to good and an assessment of the quality of Stanczuk 2016 showed it to have a generally low risk of bias to this review question.[36] The only exception was that not all patients enrolled in the study received a reference standard or were included in the analyses, potentially biasing the accuracy results.

Applicability: The eligibility criteria for the inclusion of studies in Arbyn 2014 were generally well aligned to those applied in this review.[35] One concern regarding the study populations included in the SLR has been considered. Studies examining both women who are considered at high-risk, or who had been invited to follow-up cytology were included, while the population of interest in this review is a healthy screening population. However, some outcomes were presented separately for a low-risk population, and it was demonstrated that the variability between the populations was low, leading to a low risk of bias in results pooled across all included studies. Overall, the results from the SLR are considered applicable to this review. An assessment of applicability of Stanczuk 2016 using the QUADAS-2 checklist also demonstrated a low risk of bias.[36]

Consistency: A relatively broad range of absolute accuracy values for the detection of CIN2+ and CIN3+ in self-collected samples were reported in Arbyn 2014. However, to an extent, this result can be rationalised by the difference in HPV test methodology in the included studies and it was demonstrated that, with one exception, issues of study quality did not impact on accuracy.[35] The absolute accuracy values reported in Stanczuk 2016 fall within this range and support the consistency of the result presented from the meta-analysis.[36]

Conclusion

The publications identified in this review present a relatively broad range of accuracy results of HPV testing on self-collected samples. However, this is partially explained by heterogeneity in the study methodology and it is concluded that, while the accuracy of testing on self-collected samples in lower than on clinician-collected samples, it is not substantially lower and can be considered a suitable alternative for women who do not attend for primary screening. An investigation into the most accurate HPV testing methods may be of use to optimise the accuracy of testing on self-collected samples and further validate the use of this method in a screening programme.

Summary: Criterion 4 **met** for self-sampling

2.2.1.2 Question 5 – Does self-collection of vaginal specimens increase uptake of cervical screening?

Criterion 6 of the UK NSC Screening Criteria states that: 'The test, from sample collection to delivery of results, should be acceptable to the target population.'

This review looked for studies reporting the uptake of self-sampling as a screening method for HPV in populations of under-screened women.

2.2.1.2.1 Description of the evidence

This review identified one SLR and meta-analysis,[38] and 10 additional primary studies,[39-48] all reporting self-sampling participation in 'under-screened' populations.

The SLR, Verdoodt 2015, identified 16 studies and evaluated whether offering a self-sampling kit could increase screening attendance in irregularly-screened or never-screened women, or women who did not respond to ≥1 invitation for conventional screening.[38] These populations align with the eligibility criteria in this review. Further criteria applied in Verdoodt 2015 were the exclusion of studies without a comparator arm, and studies with less than 1,000 participants;[38] as a result, the SLR would have excluded some studies which would have been eligible for inclusion in this review. However, these would have been small, non-comparative studies which would be unlikely to change the overall weight of the evidence.

2.2.1.2.2 Quality assessment

An assessment of the methodological quality of the SLR, using the AMSTAR checklist, demonstrated moderate quality, with 6 of 11 checklist questions addressed.[38] The points which were not addressed were mostly in relation to reporting and were not considered to be of great concern to the overall SLR quality.

Verdoodt 2015 assessed the quality of its included studies using the Cochrane tool for bias and reported overall moderate to high study quality (Table 4).[38] The studies were all conducted in 'under-screened' populations, which is the population of interest in this review question. The eligibility criteria with regards to age of participants in the studies were generally well aligned to the UK screening population, ranging from a minimum of 25 to 39 years old to a maximum of 50 to 69 years old. Of note in this quality assessment is that the exact time interval which was set as a threshold for 'participation' was not reported in a quarter of studies. A high risk of bias for 'selective reporting' was assigned to one study because women who had undertaken conventional screening in the clinic were removed entirely from the total number of women in the self-sampling arm, as opposed to presenting a per-protocol result (that is, the uptake of self-collected sampling amongst the entire population who were offered self-sampling). Women who opted out from the study were also removed entirely from analyses. Both omissions have potential to bias the result.

Table 4 Quality assessment (Cochrane tool for bias) of the studies included in Verdoodt 2015[38]

Risk of bias	Selection		Attrition	Reporting	
	Random sequence generation	Allocation concealment	Incomplete outcome data	Reporting of timelines	Selective reporting
Low	9	6	16	12	12
Moderate	7	10	0	4	3
High	0	0	0	0	1

The 10 additional primary studies identified in this review were assessed for quality using a modified Downs and Black checklist, the results from which are presented in Table 5.[39-48]

Table 5 Quality assessment (modified Downs and Black) of the primary studies included in this review

Question	Bosgraaf 2015[39]	Ducancelle 2015[40]	Duke 2015[41]	Enerly 2016[42]	Sultana 2016[43]	Tamalet 2016[44]	Verhoef 2014[45]	Virtanen 2015[46]	Kitchener 2016[47]	Racey 2016[48]
EXTERNAL VALIDITY										
Summary	Low	Low	Low	Low	Low	Low	Low	Low	Low	Low
CONFOUNDING										
Summary	Low	Low	High	Moderate	Low	Low	Moderate	Low	Moderate	High
POWER										
Summary	High	N/A	Unclear	Unclear	Unclear	N/A	N/A	N/A	Low	Unclear

Across all 10 identified studies, a low risk of bias is associated with 'external validity'.[39-48] To a certain extent, all studies assessed the population of interest to this review (women eligible for cervical cancer services but who do not participate in clinician led screening), although some were slightly less applicable, for example, Duke 2015 recruited from a generally 'under-screened' population instead of identifying individual women who had not responded to a screening invitation.[41] An additional concern is that none of the study populations were in age

ranges which align with UK cervical cancer screening eligibility,[39-48] however, this is unlikely to have a major impact on outcomes. Overall the studies are considered applicable.

The primary studies had an overall moderate risk of bias with respect to confounding. All studies included an accurate measure for participation outcome, but in 3 studies the number of samples returned but which had inadequate sample for testing were not reported.[41, 45, 47] The risk of bias relating to confounding was generally unclear as a result of under-reporting of baseline characteristics for the study populations. Among comparative studies, only 2 studies reported prognostic factors for participants: Sultana 2016 reported participant age and Racey 2016 additionally reported lifetime smoking history and number of sexual partners.[43, 48] As a result of the majority of studies not reporting baseline characteristics, the impact of these factors, or the requirement for adjustment to analyses to account for these, was unclear.

Many of the studies were non-comparative. For 4 comparative studies, power calculations were not reported,[41, 42, 48] or not reported for the outcome of interest.[43] One study reported power calculations but did not meet its prespecified required sample size: 16,500 women in each arm were required to give 80% power to detect a 1.4% difference in participation rates, however only 30,130 women were randomised.[39] Only a single study reported that it had adequate power to detect a meaningful difference between study arms.[47]

2.2.1.2.3 Results

Participation in self-sampling and clinician-collected sampling was reported in Verdoodt 2015 separately for studies utilising an intention to treat (ITT) or per-protocol analysis approach.[38]

The distinction between intention to treat (ITT) and per-protocol analyses is as follows:

- ITT results consider all participants who are screened including both those participating in self-sampling and those who attend for clinician-collected samples
- per-protocol analyses only consider the number of participants returning self-samples

Pooled analysis results from Verdoodt 2015 are presented in Table 6.[38]

Table 6 Verdoodt 2015 absolute participation results in ITT and per-protocol analyses[38]

		Absolute participation		Relative participation (95% CI)	Participation difference, % (95% CI)
	Studies, n	Self-sampling, % (95% CI)	Clinician-sampling, % (95% CI)		
Per-protocol					
Mail-to-all	13	20.7 (16.9 to 24.8) [range: 6.4 to 34.0]	10.3 (6.2 to 15.2)	2.06 (1.44 to 2.96)	9.9 (5.8 to 13.9)
Opt-in	3	9.7 (6.5 to 13.5)	12.2 (10.9 to 13.6)	0.72 (0.53 to 0.99)	-3.2 (-6.6 to 0.1)
ITT					
Mail-to-all	13	23.6 (20.2 to 27.3) [range: 10.2 to 39.0]	10.3 (6.2 to 15.2)	2.40 (1.73 to 3.33)	12.6 (9.3 to 15.9)
Opt-in	3	14.0 (8.0 to 21.4)	12.2 (10.9 to 13.6)	0.97 (0.65 to 1.46)	0.2 (-4.5 to 4.9)

ITT, intention to treat

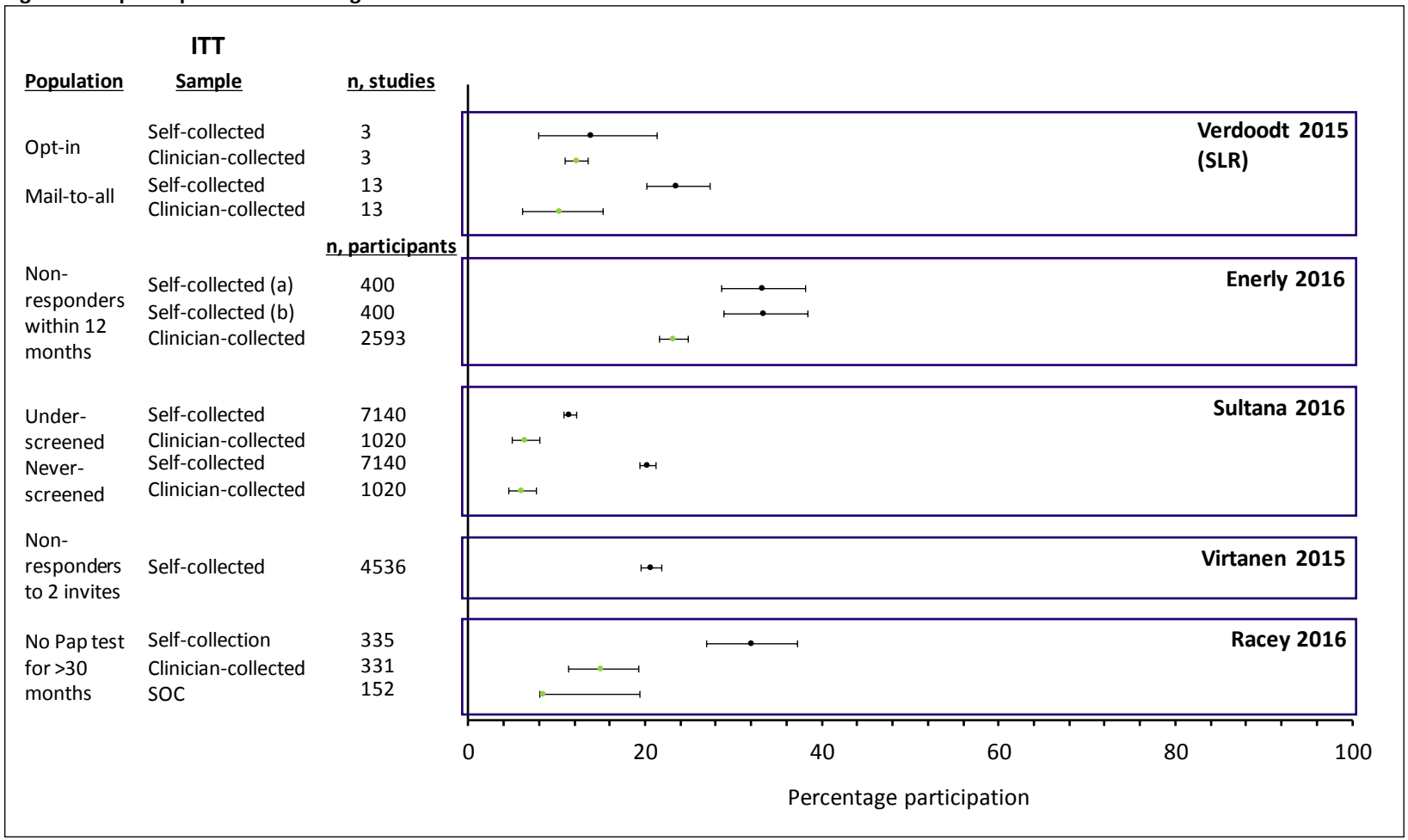
Verdoodt 2015 pools results separately for 'mail-to-all' or 'opt-in' study designs because the participation between these different distribution scenarios varied significantly. In mail-to-all studies, all participants were sent self-sampling kits directly to their home addresses, whereas in

opt-in study designs, women were sent an invitation to order a self-sampling kit by phone or mail, or alternatively to pick up a kit at a pharmacy. In mail-to-all studies, both per-protocol and ITT participation was significantly higher in the self-sampling arm than in the control arm. However, the results pooled across studies with an opt-in design did not demonstrate a significant difference between the self-collected and clinician-collected arms in either per-protocol or ITT analyses.[38] The results here highlight the impact of self-sample kit distribution on screening participation. The comparative ITT and per-protocol uptakes are discussed below.

Intention to treat participation

Four primary studies,[42, 43, 46, 48] in addition to the pooled analyses in Verdoodt 2015,[38] reported an ITT analysis. Results from these studies are presented in Figure 7.

Figure 7 ITT participation in screening



ITT, Intention to Treat; SOC, Standard of Care; SOC defined as opportunistic screening.

Self-collected (a): Evalyn Brush sampling device; Self collected (b): Delphi Screener sampling device

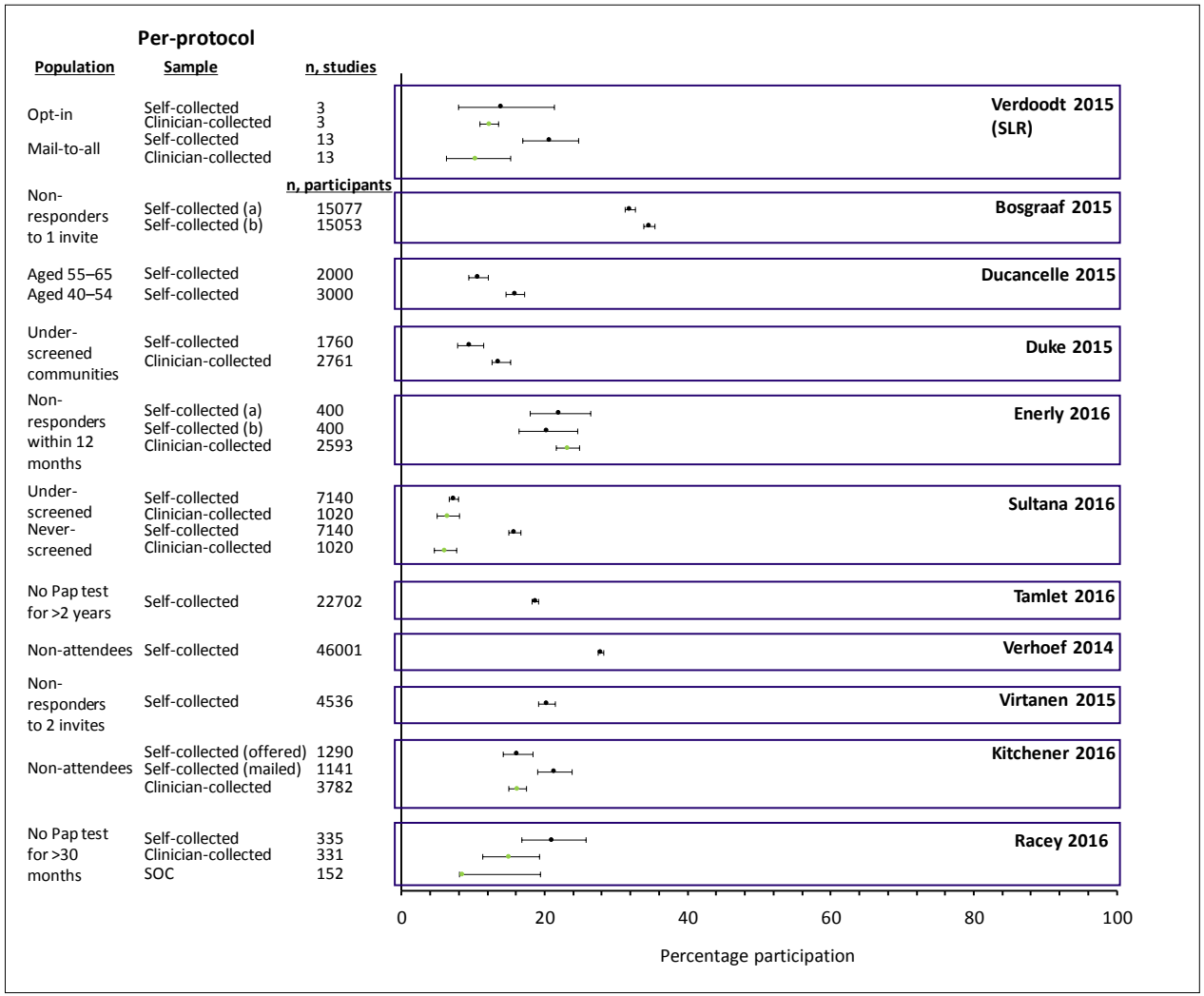
In all studies which reported participation in a control arm, the participation in screening was higher for participants offered self-collected sampling than for those who were offered clinician-sampling. Verdoodt 2015 reported an ITT uptake pooled across the mail-to-all studies of 23.6%, whereas in the comparable clinician-collected arm the uptake was 10.3%.[38] Absolute ITT participation in women offered self-collected sampling in the additional primary studies ranged from 11.5%[43] to 33.5%,[42] demonstrating a general agreement with the meta-analysis results.

A generally consistent result for screening uptake was reported across a well conducted SLR and 4 additional publications with no key quality concerns.[38, 42, 43, 46, 48] As a result, the results presented above are likely to be reliable.

Per-protocol participation

All 10 included primary studies,[39-48] in addition to Verdoodt 2015,[38] reported a per-protocol analysis. Results of these studies are presented in Figure 8.

Figure 8 Per-protocol participation in self-sampling



SOC: Standard of Care; SOC defined as opportunistic screening. Self-collected (a): Evalyn Brush sampling device; Self collected (b): Delphi Screener sampling device

Per-protocol participation in the self-collection arm, defined as the proportion of women returning self-sampling kits, was higher in a number of studies when compared to the clinician-collected participation, however, this was not always the case. The trend is not as apparent as in ITT analyses.

Verdoodt 2015 reported a per-protocol participation in the self-collection arm, pooled across mail-to-all studies, of 20.7%, whereas uptake in the clinician-collected arm was 10.3%. On consideration of opt-in study designs, the opposite trend was demonstrated and 9.7% of women in the self-collection arm returned samples, whereas participation in the clinician-collected arm was 12.2%.[38] Results from the self-collection arms of the 10 additional studies considering per-protocol participation ranged from 7.3%[43] to 34.6%.[39] These values are generally consistent. On consideration of the moderate to good quality of the primary studies, and particularly the low risk of bias associated with external validity, the results from the primary studies support the accuracy of the values reported in the meta-analysis.

As discussed previously, the SLR demonstrated an impact of self-sample kit distribution method on the uptake of self-collected sampling.[38] This outcome is further supported in the results presented here. Duke 2015 reports a per-protocol self-collected sampling uptake of 9.5% which is low in the overall range presented.[41] This result can be rationalised by the 'opt-in' methodology of the study and aligns closely with the 'opt-in' result from the meta-analysis. Furthermore, Kitchener 2016 assesses the difference in participation if self-sampling kits are sent to participants (analogous to a mail-to-all study design) or if kits are offered (opt-in design). Participation results are 21.3% and 16.2% respectively which adds further confidence in the SLR results.[47]

Intention to treat vs per-protocol

On consideration of the ITT and per-protocol participation results in Verdoodt 2015,[38] it is clear that sending a self-sampling kit can act as a prompt to encourage women to be screened and that some subsequently choose to attend clinician-sampling as opposed to returning a self-collected sample. In Verdoodt 2015, 20.7% of women returned self-sampling kits, whereas 23.6% of women attended any form of screening for cervical cancer.[38]

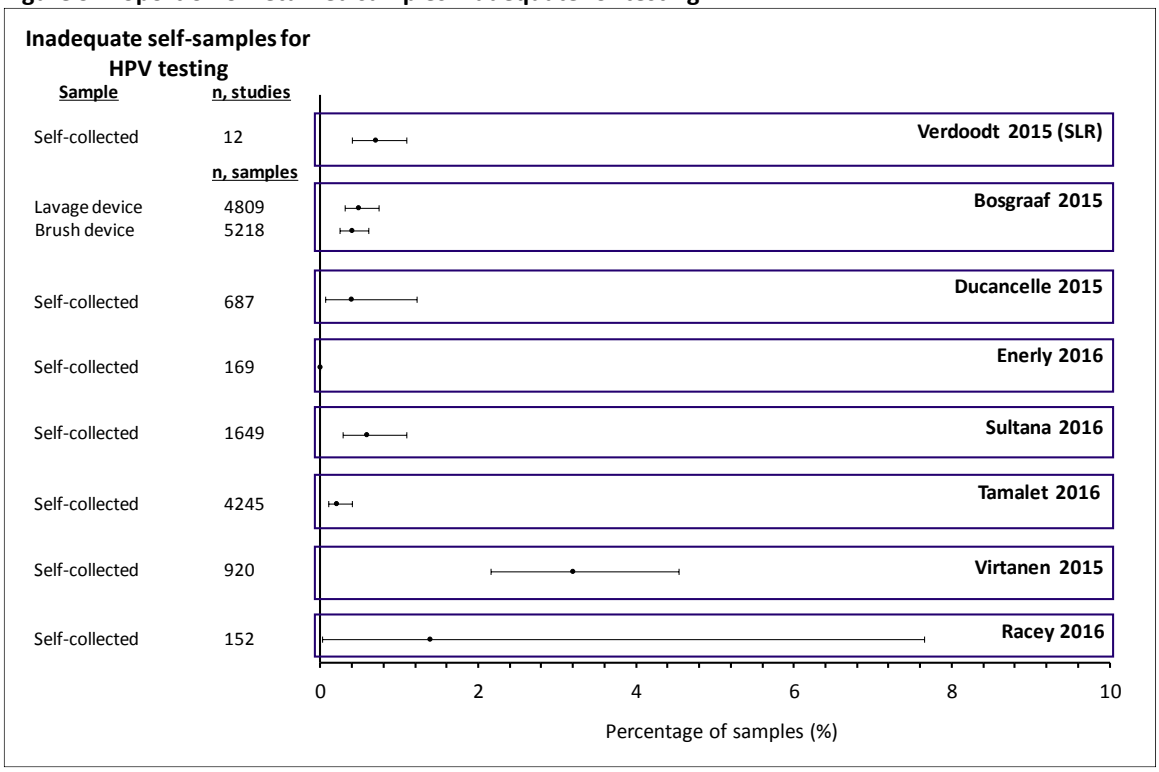
This trend was demonstrated to a greater extent in 3 of the additional primary studies reporting an ITT analysis.[42, 43, 48] Most notably, Enerly 2016 reported a screening participation in women offered a self-sampling kit of 33.4%; this value comprised 21.1% of women returning a self-collected sample and 12.3% attending for cytology at a clinic.[42] Similar results were seen in the other 2 studies;[43, 48]among the women who participated in screening after self-sampling was offered, approximately one-quarter to one-third chose instead to have clinician-collected sampling . In contrast, Virtanen 2015 reported a screening participation rate of 20.7%, of which 98% returned a self-collected sample.[46] This inconsistency between study results leads to uncertainty in the proportion of women who are offered self-sampling but instead attend for clinician-collected sampling.

An investigation into the potential impact this effect could have on the cost-effectiveness of a screening strategy would be beneficial. While this is a positive outcome with regards to increasing screening coverage, there is potential for an economic impact if a high proportion of women leave self-sampling kits unused and instead choose to visit a clinician. Additionally, it should be considered whether alternative approaches could increase screening uptake in a more cost-effective way.

Inadequate samples

The proportion of self-collected samples returned which were inadequate for HPV testing was presented in Verdoodt 2015,[38] pooled across 12 of its included studies, and was additionally presented in 8 of the primary studies identified in this review.[39, 40, 42-44, 46, 48] The proportions of inadequate samples from each study is presented in Figure 9.

Figure 9 Proportion of returned samples inadequate for testing



The pooled proportion of unsatisfactory samples reported in Verdoodt 2015 was 0.7% (95% CI 0.4% to 1.1%).[38] Similar results were presented in the additional primary studies.[39, 40, 42-44, 46, 48] The proportion of inadequate self-collected samples returned was 0[42] to 3.2%[46] across the studies, with the upper end of the 95% CI reaching 7.66% in one small study.[48] This supports the evidence in the SLR that concern associated with insufficient samples for HPV testing is low.[38]

2.2.1.2.4 Evidence summary

The evidence identified for this review question comprised one SLR and meta-analysis,[38] and 10 further primary studies reporting the uptake of self-collected sampling for HPV testing.[39-48] These publications have been demonstrated to be of moderate to good quality, and support the conclusion that offering previously under-screened women the option of self-collected sampling leads to a moderate increase in the overall screening uptake in this population. The increase in screening uptake is less pronounced when considered as an absolute difference to women in the clinician-collected arms of studies. However, since these studies are undertaken in populations of women who have previously not attended for primary screening, even a low increase in uptake could be considered successful, if a cost-effective strategy can be determined. Further investigation would be beneficial to determine if this is feasible. The evidence also suggests that the use of self-collection devices to collect samples for HPV sampling can be considered adequate, with only a very low proportion of insufficient samples reported.[38-48]

The SLR results demonstrate that participation in self-sampling is higher when sampling kits are mailed directly to eligible women, as opposed to inviting women to order or collect a kit.[38] This conclusion is supported further in primary studies and the method of distribution of self-sampling kits should be a key consideration in the design of a self-collected screening strategy.[41, 47] However, distributing kits to all eligible women when only approximately 1 in 5 will be returned could be costly and an assessment of the economic impact of this methodology could help ensure this is the most cost-effective method to increase uptake of screening.

Evidence presented here demonstrates that ITT analyses show generally improved participation rates when compared to per-protocol analyses.[38, 42, 43, 46, 48] This result could also affect the cost-effectiveness of a self-collection screening programme due to the cost of distributing self-sampling kits which remain unused, in addition to the higher cost of women subsequently attending for clinician-collected sampling, and would also warrant further investigation.

A final key consideration relating to this review question is the potential impact on primary screening uptake if women are aware that non-attendance will result in being sent a self-sampling kit, which might be considered more convenient than attending for clinician-based screening. Ensuring that implementation of a self-collection screening strategy is effective in reaching under-screened women, without encouraging women who would normally attend screening to miss clinician appointments, would be of great importance. An assessment of the impact that a subsequent 'self-sampling opportunity' would have on initial screening uptake would be of value.

Analysis of evidence relevant to criterion 6 – Uptake of self-collected sampling

Quantity: One SLR and meta-analysis[38] evaluating participation in self-collected sampling is supported by 10 primary studies identified in this review.[39-48] The SLR reported results from 16 studies, considering a total of 163, 431 women.[38] The evidence base for this review question is therefore large.

Quality: The quality of the SLR was assessed to be moderate using the AMSTAR checklist, addressing 6 of the 11 checklist questions.[38] The SLR reported moderate to high quality of its identified studies and highlighted no key quality concerns relating to this review.[38] An assessment of the 10 primary studies identified in this review demonstrated an overall moderate to good quality, particularly with regards to external validity for which all publications had a low risk of bias.

It should be noted that in 3 studies the proportion of samples returned which were inadequate for HPV-testing was not reported.[41, 45, 47] Additionally, the risk of bias associated with confounding was unclear because the majority of studies did not report population prognostic factors. However, it is unlikely that this would bias results significantly when considering uptake.

Applicability: The studies included in the SLR investigated populations of women considered to be 'under-screened' which aligns with the population of interest for this review question. Verdoodt 2015 only identified studies for inclusion which had a comparator arm, which was not an eligibility criterion in this review and could potentially impact the applicability of the SLR result.[38] None of the primary studies identified in this review assessed populations in the exact age range applicable to the UK cervical cancer screening population; in most instances this is unlikely to bias results, however, some studies were conducted in only younger[47] or older populations.[40]

Consistency: Generally consistent results were reported for self-sampling participation in both ITT[38, 42, 43, 46, 48] and per-protocol analyses.[38-48] This strengthens the conclusion that offering self-collected sampling to previously under-screened women could increase overall screening uptake. Additionally, a consistently low proportion of samples which were insufficient for HPV testing was reported across the studies,[38-40, 42-44, 46, 48] demonstrating that this is an adequate collection method.

Conclusions

A generally consistent result for participation was reported in the included studies, for both ITT and per-protocol analyses. This demonstrates that offering previously under-screened women the option of self-collected HPV testing leads to a moderate increase in the overall uptake in screening. However, while this is not insubstantial, the increase in uptake when compared to women sent another invitation for clinician sampling in the same studies was relatively small. Overall this result would benefit from further exploration.

The results also demonstrate that the use of a self-collection device is an adequate method of sample collection, with only very low proportions of samples reported to be inadequate for HPV testing.

The potential impact of offering self-sampling on the uptake of more accurate, clinician-based screening should be assessed to ensure a self-sampling strategy would not impact negatively on the overall detection of HPV. Additionally, an investigation into the circumstances in which the strategy would be useful and the most appropriate method of self-sample kit distribution, with regards to optimising uptake and cost-effectiveness, would aid the design of a self-sampling screening strategy.

Summary: Criterion 6 **uncertain** for self-sampling

3 Review summary

3.1 Conclusions and implications for policy

Summary of findings

Overall, this review found that:

- There is good evidence that self-collected sampling for HPV testing has lower accuracy than clinician tested sampling, but the difference is small
- There is good evidence that offering self-sampling could moderately increase the uptake of screening among women who do not respond to invitations for clinician-based screening, however, further investigation into the optimisation of a self-sampling strategy is required

3.1.1 Use of self-sampling to improve cervical screening uptake

The evidence identified in this review considering the use of self-collected samples for HPV screening demonstrated that offering previously under-screened women the option of self-collected sampling moderately increases overall screening uptake in this population. Additionally, it has been demonstrated that the use of a self-collection device is an adequate method for sample collection and that testing on these samples has sufficient accuracy to be considered an appropriate alternative to clinician-collected sampling where required. The evidence from which these conclusions are drawn is of moderate to good quality and with a reasonable consistency in results. However, the practical application of a self-collection strategy may need further exploration to ensure successful implementation and cost-effectiveness.

It should be assessed whether awareness that self-sampling would subsequently be available if primary screening is missed reduces uptake of standard clinician-based screening. All of the studies identified in this review investigated women who had failed to respond to 1, or a number of, previous screening invitations without being aware of a later self-sampling option and, as a result, did not investigate the impact this could have on initial screening uptake. Given that the accuracy of testing on self-collected samples is lower than on clinician-collected samples, it is important that women are not discouraged from attending clinician-sampling in favour of self-sampling. A Dutch model assessed the impact of women switching from clinician-collected to self-collected screening on the cost-effectiveness of a screening programme. The model varied the accuracy of the self-sampling test, the increase in attendance after offering self-sampling and risk of women developing cervical cancer. Under all scenarios, switching of women from clinician-collected to self-sampling resulted in a decrease in QALYs gained. If self-sampling were to be implemented it would be important to communicate the relative advantages of clinician-collected screening to reduce the impact of switching.[52]

The circumstances in which the strategy should be used is an important consideration. An additional consideration which warrants further investigation is the method of distributing self-sampling kits, and the impact this has on both the uptake and cost-effectiveness of screening. Verdoodt 2015 demonstrated that participation in self-sampling is higher when sampling kits are mailed directly to eligible women, as opposed to inviting women to order or collect a kit.[38] However, while sending all under-screened women self-sampling kits may increase screening uptake, distributing kits to all eligible women when only a small proportion are returned could

have a sizeable economic impact on a screening programme. An assessment of the threshold return rate at which this methodology would be cost-effective should be made to help ensure that this is a cost-effective method to increase screening participation. Modelling could be proposed to evaluate this factor.

3.2 Limitations of this review

This rapid review was conducted in line with the UK NSC requirements for evidence summaries.[53] These requirements are mostly in line with published guidelines for systematic reviews, but allowing for some methodological compromises. Some specific limitations relating to this review are discussed below.

3.2.1 Included study designs

It should be noted that this review was only designed to look for primary evidence that directly addressed the questions being considered. This review did not aim to systematically identify alternative forms of evidence such as modelling studies. A more holistic view of the evidence, such as modelling using published data on test accuracy and natural history, might be sufficient to give confidence that longer screening intervals would be safe.

3.2.2 Included publication types

This review only included peer-reviewed journal publications, and excluded any literature that was not peer-reviewed such as congress presentations and government reports. This may have led to the exclusion of relevant evidence that has only been published in non-peer-reviewed formats. However, this is an accepted methodological adjustment for a rapid review, and is unlikely to miss any pivotal studies, which would likely be published in peer-reviewed journals.

3.2.3 Review methodology

Articles were reviewed by a single reviewer in the first instance. A second reviewer examined all included articles, 10% of excluded articles, and any articles where there was uncertainty about inclusion. Although a fully systematic review would require all articles to be reviewed by both reviewers, this pragmatic strategy should have ensured that any articles where the eligibility was unclear were reviewed twice.

4 Appendices

Appendix 1 – Search strategy

Electronic databases

The search strategy included searches of the databases shown in **Error! Reference source not found.** and Table 7. MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase were searched simultaneously.

Table 7 Summary of electronic database searches and dates - use of self-sampling to improve screening uptake

Database	Platform	Searched on date	Date range of search
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	19th October 2016	1974 to 2016 October 19
Embase	Ovid SP	19th October 2016	1974 to 2016 October 19
The Cochrane Library, including: • Cochrane Database of Systematic Reviews (CDSR) • Cochrane Central Register of Controlled Trials (CENTRAL) • Database of Abstracts of Reviews of Effects (DARE)	Wiley Online	20th October 2016	CENTRAL: Issue 9 of 12, September 2016 DARE: Issue 2 of 4, April 2015

Search terms

Search terms included combinations of free text and subject headings. Search terms for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase are shown in **Error! Reference source not found.** (Topic 1) and Table 8 (Topic 2), and search terms for the Cochrane Library databases are shown in **Error! Reference source not found.** (Topic 1) and Table 9 (Topic 2).

Table 8 Search terms for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase (searched simultaneously via Ovid SP) - use of self-sampling to improve screening uptake

Term group	#	Search terms	Results
Cervical cancer	1	exp Uterine cervical neoplasms/ or exp Uterine cervix cancer/ or exp Cervical intraepithelial neoplasia/ or uterine cervix carcinoma in situ/	166564
	2	(cervi$ adj3 (cancer$ or tumor$ or tumour$ or neoplas$ or dysplas$ or carcinoma$ or malignan$ or adenocarcinoma$ or choriocarcinoma$ or orteratoma$ or sarcoma$ or precancer$ or pre-cancer$)).tw.	151796
	3	(CIN or CIN1$ or CIN2$ or CIN3$ or "CIN 1$" or "CIN 2$" or "CIN 3$" or CINI$ or "CIN I$").tw.	23959
	4	or/1-3	213824
HPV	5	exp papillomavirus infections/ or exp papillomaviridae/ or exp human papilloma virus/ or exp papovavirus/	116054
	6	(human papillomavir$ or human papilloma vir$ or HPV$ or HR-HPV$ or hrHPV$).tw.	93535
	7	5 or 6	142016
Screening	8	exp mass screening/	310262

Term group	#	Search terms	Results
	9	exp Vaginal smears/ or human papillomavirus DNA tests/ or DNA probes, HPV/	58611
	10	(test$ or cotest$ or co-test$ or screen$).ti.	1074508
	11	case find$.tw.	9207
	12	((human papillomavir$ or human papilloma vir$ or HPV$ or HR-HPV$ or hrHPV$) adj2 (DNA or test$)).tw.	25002
	13	((vagina$ or cervi*) adj3 (smear$ or swab$ or scrap$ or test$ or sampl$)).tw.	37904
	14	exp Early detection of cancer/ or exp *Uterine cervical neoplasms/pc or exp *Uterine cervix cancer/pc or exp *Cervical intraepithelial neoplasia/pc or exp *Uterine cervix carcinoma in situ/pc	25327
	15	(annual adj (surveillance or review)).tw.	2464
	16	or/8-15	1350922
Self-collection	17	(self sampl$ or self collect$ or self care$ or self test$).tw.	34770
Outcome 1: Diagnostic test accuracy	18	exp "Sensitivity and Specificity"/ or *Diagnostic Accuracy/	763286
	19	(sensitiv$ or specific$ or accura$).tw.	8275328
	20	((pre-test or pretest) adj probability).tw.	4488
	21	((post-test or posttest) adj probability).tw.	1544
	22	predictive value$.tw.	204444
	23	likelihood ratio$.tw.	26869
	24	or/18-23	8647872
	25	limit 24 to yr=2013-2016	2056259
Outcome 2: Uptake/compliance	26	exp Patient acceptance of health care/ or exp Patient Attitude/	515638
	27	(satisf$ or dropout$ or drop out).tw.	624360
	28	(compliance or complie$ or comply$).tw.	263153
	29	(encourage$ or improve$ or improving or increas$ or promot$).tw.	14094754
	30	(uptake or particip$ or nonattend$ or non-attend$).tw.	2590532
	31	(accept$ or attend$ or attitude$ or utilisation or utilization).tw.	1696361
	32	(refus$ or respon$ or reluctan$ or nonrespon$).tw.	6484726
	33	or/26-32	20042442
	34	limit 33 to yr=2015-2016	2421932
Outcomes	35	25 or 34	3881443
Total	36	4 and 7 and 16 and 17 and 35	348
	37	exp animals/ not exp humans/	8792525
	38	("Journal: Conference Abstract" or comment or letter or case reports).pt.	6214134
	39	37 or 38	14681174
	40	36 not 39	301
	41	remove duplicates from 40	177

Table 9 Search terms for the Cochrane Library Databases (searched via the Wiley Online platform) - use of self-sampling to improve screening uptake

Term group	#	Search terms	Results
Cervical cancer	1	[mh "Uterine cervical neoplasms"] or [mh "Cervical intraepithelial neoplasia"]	1989
	2	(cervi* near/3 (cancer* or tumor* or tumour* or neoplas* or dysplas* or carcinoma* or malignan* or adenocarcinoma* or choriocarcinoma* or orteratoma* or sarcoma* or precancer* or pre-cancer*)):ti,ab,kw	3570
	3	(CIN or CIN1* or CIN2* or CIN3* or "CIN 1*" or "CIN 2*" or "CIN 3*" or CINI* or "CIN I*"):ti,ab,kw	1085
	4	{or #1-#3}	4198
HPV	5	[mh "papillomavirus infections"] or [mh papillomaviridae] or [mh "human papilloma virus"]	1218

Term group	#	Search terms	Results
	6	("human papillomavir*" or "human papilloma vir*" or HPV* or "HR-HPV*" or hrHPV*):ti,ab,kw	1593
	7	#5 or #6	1956
Screening	8	[mh "mass screening"]	5513
	9	[mh "Vaginal smears"] or [mh "test, hpv dna"]	798
	10	(test* or cotest* or co-test* or screen*):ti	23886
	11	"case find*":ti,ab,kw	215
	12	(("human papillomavir*" or "human papilloma vir*" or HPV* or "HR-HPV*" or hrHPV*) near/2 (DNA or test*)):ti,ab,kw	528
	13	((vagina* or cervi*) near/3 (smear* or swab* or scrap* or test* or sampl*)):ti,ab,kw	1650
	14	[mh "Early detection of cancer"] or [mh "Uterine cervical neoplasms"] or [mh "Cervical intraepithelial neoplasia"]	2727
	15	(annual next (surveillance or review)):ti,ab,kw	30
	16	{or #8-#15}	28112
Self-collection	17	("self sampl*" or "self collect*" or "self care*" or "self test*"):ti,ab,kw	5745
Outcome 1: Diagnostic test accuracy	18	[mh "Sensitivity and Specificity"] or [mh "Diagnostic Accuracy"]	18093
	19	(sensitiv* or specific* or accura*):ti,ab,kw	111264
	20	(("pre-test" or pretest) next probability):ti,ab,kw	86
	21	(("post-test" or posttest) next probability):ti,ab,kw	43
	22	"predictive value*":ti,ab,kw	11838
	23	"likelihood ratio*":ti,ab,kw	488
	24	{or #18-#24}	117737
	25	#24 Publication Year from 2013 to 2016	33032
Outcome 2: Uptake/compliance	26	[mh "Patient acceptance of health care"] or [mh "Patient Attitude"]	24278
	27	(satisf* or dropout* or "drop out*"):ti,ab,kw	37641
	28	(compliance or complie* or comply*):ti,ab,kw	26628
	29	(encourage* or improve* or improving or increas* or promot*):ti,ab,kw	381535
	30	(uptake or particip* or nonattend* or "non-attend*"):ti,ab,kw	128968
	31	(accept* or attend* or attitude* or utilisation or utilization):ti,ab,kw	59548
	32	(refus* or respon* or reluctan* or nonrespon* or "non-respon*"):ti,ab,kw	179564
	33	{or #26-#32}	535459
	34	#33 Publication Year from 2015 to 2016	53244
Outcomes	35	#25 or #34	75818
Total	36	#4 and #7 and #16 and #17 and #35 in Other Reviews and Trials	27

Appendix 2 – Study selection

Review process

The following review process was followed:

- Each abstract was reviewed against the inclusion/exclusion criteria by one reviewer. Where the applicability of the inclusion criteria was unclear, the article was included at this stage in order to ensure that all potentially relevant studies were captured. A second independent reviewer provided input in cases of uncertainty, and validated 20% of the first reviewer's screening decisions. Any disagreements were resolved by discussion until a consensus was met.
- Full-text articles required for the full-text review stage were acquired.
- Each full-text article was reviewed against the inclusion/exclusion criteria by one reviewer, who determined whether the article was relevant to one or more of the review questions. A second independent reviewer provided input in cases of uncertainty, and validated 20% of the first reviewer's screening decisions. Any disagreements were resolved by discussion until a consensus was met.

Eligibility criteria

Eligibility criteria for each question are presented in **Error! Reference source not found.** to Table 11 below. All search results for Topic 1 were reviewed against the eligibility criteria for Questions 1 to 3; all search results for Topic 2 were reviewed against the eligibility criteria for Questions 4 and 5.

For all topics, systematic reviews and meta-analyses were considered for inclusion in this review. If the scope of a systematic review or meta-analysis was very closely aligned to one of the questions in this review, it was included in this review in its own right. However, if the scope was not closely aligned to one of the questions in this review but some of the included articles were of interest, the reference list of the systematic review or meta-analysis was hand-searched. Any primary research articles that were identified as being relevant to this review were then included.

Table 10 Eligibility criteria for publications relating to question 4

Domain	Inclusion criteria	Exclusion criteria
Population	Women in the cervical cancer screening population	Studies that do not include women eligible for cervical cancer screening
Intervention (s)	HPV testing on a self-collected sample	Studies that do not include an HPV test on a self-collected sample
Comparator	HPV or cytology testing on a clinician-collected sample	Studies that do not include a comparator test on a clinician-collected sample
Reference Standard	Colposcopy or biopsy	Any other reference standard
Outcomes	Measures of screening accuracy, or sufficient data to calculate these: • Sensitivity • Specificity • Positive predictive value • Negative predictive value • Accuracy • Likelihood ratio	Outcomes not relating to the measures of screening accuracy of HPV testing on self-collected specimens
Study design and publication type	Peer-reviewed evidence derived from the following types of study: • RCTs	• Other study designs or publication types • Retrospective studies, case control studies or cross-sectional studies

Domain	Inclusion criteria	Exclusion criteria
	• Non-randomised, comparative interventional studies • Prospective cohort studies • Systematic reviews and meta-analyses of the above study types	• Conference abstracts or other publication types that have not been peer-reviewed
Language	English language	Non-English language

Table 11 Eligibility criteria for publications relating to question 5

Domain	Inclusion criteria	Exclusion criteria
Population	Women who are eligible for cervical cancer screening but do not participate in clinician-led screening services (note that these may be a sub-group of a larger study)	Studies that do not include women who are eligible for cervical cancer screening but do not participate in clinician-led screening services, or studies that do not report outcomes separately for this group
Intervention	Offer or invitation of HPV testing on a self-collected sample	Studies that do not include an offer or invitation of HPV self-sampling
Comparator	Offer or invitation of a clinician-collected sample, or no comparator	-
Outcomes	Measures of uptake, compliance or participation	Any other outcomes
Study design and publication type	Peer-reviewed evidence derived from the following types of study: • RCTs • Non-randomised, comparative interventional studies • Prospective cohort studies • Systematic reviews and meta-analyses of the above study types	• Other study designs or publication types • Retrospective studies, case control studies or cross-sectional studies • Conference abstracts or other publication types that have not been peer-reviewed
Language	English language	Non-English language

Appraisal for quality and risk of bias

The following tools were used to assess the quality and risk of bias of each study included in the review:

- **Systematic literature reviews:** Assessing the Methodological Quality of Systematic Reviews (AMSTAR) checklist[54]
- **Diagnostic accuracy studies**: Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool[55]
- **Interventional and observational studies**: Modified versions of the Downs and Black checklist[56]

Reviewers were provided with guidance criteria to ensure consistent applicability of the quality assessment checklists for primary studies; these criteria are detailed in **Error! Reference source not found.** (Modified Downs & Black, Topic 1), Table 12 (QUADAS-2, Topic 2, Question 4) and Table 13 (Modified Downs & Black, Topic 2, Question 5).

Table 12 Template quality assessment checklist for question 4

Question	Literature-recommended criteria	Guideline criteria for HPV self-testing
PATIENT SELECTION		
Was a consecutive or random sample of patients enrolled?	A study should ideally enrol all consecutive, or a random sample of, eligible patients with suspected disease – otherwise there is potential for bias. Studies that make inappropriate exclusions, e.g. excluding "difficult to diagnose" patients, may result in overoptimistic estimates of diagnostic accuracy	**Yes** if all eligible women (asymptomatic women within the screening age range) or a random sample of women within the study period were included **No** if participants were selected in a different way, eg. by referral or convenience sample

Question	Literature-recommended criteria	Guideline criteria for HPV self-testing
Was a case-control design avoided?	Studies enrolling patients with known disease and a control group without the condition may exaggerate diagnostic accuracy	**Yes** if the study was a prospective cohort study **No** if cases of colposcopy-confirmed HPV were matched to controls with negative colposcopy results
Did the study avoid inappropriate exclusions?	Exclusion of patients with "red flags" for the target condition, who may be easier to diagnose, may lead to underestimation of diagnostic accuracy	**Yes** if all patients were included, or if exclusions were appropriate (such as women outside the low-risk screening population) and unlikely to lead to bias **No** if any group within the screening population was systematically excluded
Could the selection of patients have introduced bias?	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias	Answered based on the previous questions in this domain
Is there concern that the included patients do not match the review question?	There may be concerns regarding applicability if patients included in the study differ, compared to those targeted by the review question, in terms of severity of the target condition, demographic features, presence of differential diagnosis or co-morbidity, setting of the study and previous testing protocols	**Low** if patients overall have a normal risk for HPV so are representative of the screening population; the population should be women who had not responded to invitations for clinical testing **High** if patients overall are not representative of the screening population
INDEX TESTS		
Were the index test results interpreted without knowledge of the reference standard?	This item is similar to "blinding" in intervention studies. Interpretation of index test results may be influenced by knowledge of the reference standard	**Yes** if self-testing results were interpreted before colposcopy was performed, or if the results were interpreted after colposcopy was performed but without knowledge of the colposcopy diagnosis **No** if the colposcopy diagnosis was known when self-testing results were interpreted
If a threshold was used, was it pre-specified?	Selecting the test threshold to optimise sensitivity and/or specificity may lead to overoptimistic estimates of test performance, which is likely to be poorer in an independent sample of patients in whom the same threshold is used	**Yes** when the threshold for a positive result on the HPV test was pre-specified, such as using the threshold specified by the manufacturer **No** when the threshold was not pre-specified
Could the conduct or interpretation of the index test have introduced bias?	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias	Answered based on the previous questions in this domain
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Variations in test technology, execution, or interpretation may affect estimates of its diagnostic accuracy. If index tests methods vary from those specified in the review question there may be concerns regarding applicability	**Low** when the conduct and interpretation of the self-testing kits were relevant to the UK care setting **High** if any aspect of the index test, including its conduct or interpretation, was substantially different from the UK care setting
REFERENCE STANDARD		
Is the reference standard likely to correctly classify the test condition?	Estimates of test accuracy are based on the assumption that the reference standard is 100% sensitive and specific. Disagreements between the reference standard and index test are assumed to result from incorrect classification by the index test	**Yes** if the diagnosis was confirmed by colposcopy **No** if any other reference standard was used (note that HPV testing or cytology on a clinician-collected sample is a relevant *comparator* in this review but is not the *reference standard*)

Question	Literature-recommended criteria	Guideline criteria for HPV self-testing
Were the reference standard results interpreted without knowledge of the results of the index test?	Potential for bias is related to the potential influence of prior knowledge on the interpretation of the reference standard	**Yes** if colposcopy results were interpreted before self-testing was performed, or if the results were interpreted after self-testing was performed but without knowledge of the self-testing diagnosis **No** if the self-testing diagnosis was known when the colposcopy results were interpreted
Could the reference standard, its conduct, or its interpretation have introduced bias?	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias	Answered based on the previous questions in this domain
Is there concern that the target condition as defined by the reference standard does not match the review question?	The reference standard may be free of bias but the target condition that it defines may differ from the target condition specified in the review question. For example, when defining urinary tract infection, the reference standard is generally based on specimen culture but the threshold above which a result is considered positive may vary	**Low** if the target condition is cervical abnormalities (CIN2+, CIN3+, invasive cervical cancer) **High** for any other target condition
PATIENT FLOW		
Was there an appropriate interval between the index test(s) and the reference standard?	Ideally results of the index test and reference standard are collected on the same patients at the same time. If there is a delay or if treatment is started between index test and reference standard, misclassification may occur due to recovery or deterioration of the condition. The length of interval leading to a high risk of bias will vary between conditions. A delay of a few days may not be a problem for chronic conditions, while for acute infectious diseases a short delay may be important	**Yes** if the self-testing was conducted within a week of the clinical tests, on average **No** if the self-testing was conducted more than a week before or after the clinical tests, on average
Did all patients receive a reference standard? Did patients receive the same reference standard?	Verification bias occurs when not all of the study group receive confirmation of the diagnosis by the same reference standard. If the results of the index test influence the decision on whether to perform the reference standard or which reference standard is used, estimated diagnostic accuracy may be biased	**Yes, Yes** if all screened patients had confirmation of their diagnosis, and all were diagnosed in the same manner (similarly trained staff, similar timing of diagnosis) **No, Yes** if not all patients had colposcopy or biopsy, but those who did had the same reference standard **No** if patients received different reference standards
Were all patients included in the analysis?	All patients who were recruited into the study should be included in the analysis. There is a potential for bias if the number of patients enrolled differs from the number of patients included in the 2x2 table of results, for example because patients lost to follow-up differ systematically from those who remain	**Yes** if all screened patients were included in the final analysis **No** if any screened patients were not included in the final analysis
Could the patient flow have introduced bias?	If all signalling questions for a domain are answered "yes" then risk of bias can be judged "low". If any signalling question is answered "no" this flags the potential for bias	Answered based on the previous questions in this domain

Table 13 Template quality assessment checklist for question 5

Question	Guideline criteria for question 5
REPORTING	

Question	Guideline criteria for question 5
Is the hypothesis/aim/objective of the study clearly described?	Section removed – relates to reporting rather than study quality
Are the main outcomes to be measured clearly described in the Introduction or Methods section?	
Are the characteristics of the patients included in the study clearly described?	
Are the intervention(s) of interest clearly described?	
Are the distributions of principal confounders in each group of subjects to be compared clearly described?	
Are the main findings of the study clearly described?	
Does the study provide estimates of the random variability in the data for the main outcomes?	
Have all important adverse events that may be a consequence of the intervention been reported?	
Have the characteristics of patients lost to follow-up been described?	
EXTERNAL VALIDITY	
Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	Answer should relate to the outcome measures of interest (adherence, compliance, uptake)
Modified question: Were the subjects asked to participate in the study representative of the population of interest for this review? Original question: Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	**Yes** only when the target population was either all women eligible for screening, or women who had not responded to invitations for clinical testing **No** if study was performed in only a certain subgroup of the population of interest
Were the staff, places, and facilities where the patients were treated, representative of the treatment the majority of patients receive?	Question removed – investigating new care settings, so not possible to be representative of the treatment the majority of patients currently receive
CONFOUNDING	
Was an attempt made to blind study subjects to the intervention they have received?	Question removed – not applicable to screening and surveillance methods
Was an attempt made to blind those measuring the main outcomes of the intervention?	Question removed – outcome is either returning a self-sampling kit (for intervention) or attendance at clinic (for comparator), so it would not be possible to blind the investigator measuring these outcomes
If any of the results of the study were based on "data dredging", was this made clear?	Question removed – only interested in adherence, compliance, uptake
In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients? Removed wording: "…or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?"	**Yes** if analyses were adjusted for different lengths of follow-up if necessary, or if length of follow-up was comparable between groups **No** if the length of follow-up was not comparable, and analyses were not adjusted
Were the statistical tests used to assess the main outcomes appropriate?	**Yes** if groups were compared appropriately using risk difference, risk ratios, odds ratios, unpaired t-tests or similar; for single-arm trials a paired t-test may be appropriate; other methods may also be appropriate if justified in the publication **No** if the statistical tests were not appropriate – to be determined on a case-by-case basis

Question	Guideline criteria for question 5
	NA for single arm studies
Was compliance with the intervention/s reliable?	Question removed – compliance is an outcome of interest
Were the main outcome measures used accurate (valid and reliable)?	Answer should relate to the outcome measures of interest to this review (adherence, compliance, uptake) **Yes** when uptake was measured in a valid and reliable way, and the proportion of usable samples returned from self-testing was reported **Unclear** when uptake was measured in a valid and reliable way, but the proportion of useable samples has not been reported **No** if uptake was not measured in a valid and reliable way, or it is not clear how many samples returned from self-testing were usable
Were the patients in different intervention groups (trials and cohort studies) recruited from the same population? Removed wording: "…or were the cases and controls (case-control studies)…"	**Yes** if patients from all intervention groups were recruited from the same population **No** if different intervention groups were recruited from different populations, such as different geographical location or different baseline characteristics NA for single arm studies
Question added: Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	**Yes** if baseline characteristics were similar between treatment groups, particularly age and proportion with HPV vaccination **No** if there were significant differences between the groups in either of the characteristics listed above NA for single arm studies
Were study subjects in different intervention groups (trials and cohort studies) recruited over the same period of time? Removed wording: "…or were the cases and controls (case-control studies)…"	**Yes** if patients from all intervention groups were recruited over the same period of time **No** if patients from different intervention groups were recruited at different times, such as historical control groups NA for single arm studies
Were study subjects randomised to intervention groups?	**Yes** if randomisation was performed using computer-generated random numbers or random number tables **Inadequate** if alternation, case record numbers, birth dates or week days were used to allocate patients to treatment arms **No** if no attempt was made at randomisation
Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	**Yes** if the allocation sequence was protected before and until assignment, using methods such as: centralised or pharmacy-controlled randomisation, serially-numbered identical containers, on-site computer-based system with a randomisation sequence that is not readable until allocation, or other approaches with robust methods to prevent foreknowledge of the allocation sequence **No** if inadequate methods of randomisation were used, or if random number lists could have been viewed before allocation, such as open random number lists or serially numbered envelopes NA in non-randomised studies
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	Answer should relate to the outcome measures of interest to this review **Yes** if analyses were adjusted for differences in key baseline characteristics, or if adjustment was not necessary **No** if adjustment was necessary but was not performed NA for single arm studies
Were losses of patients to follow-up taken into account?	Question removed – loss to follow-up is related to compliance, which is an outcome of interest
POWER	

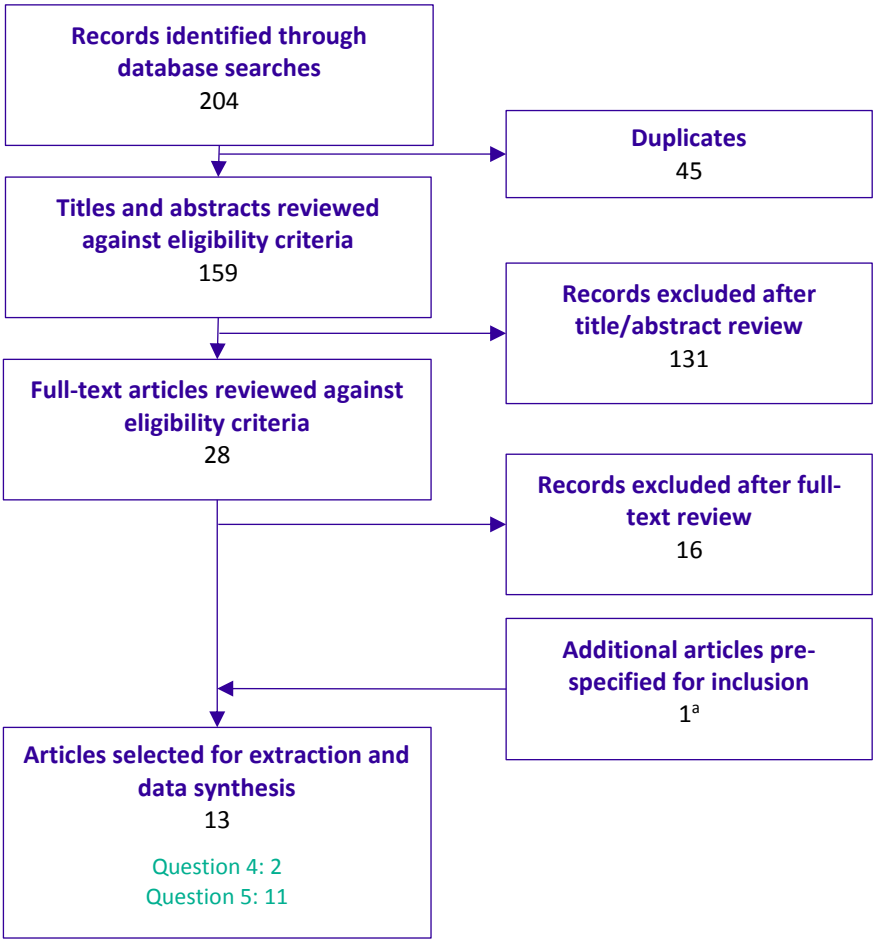
Question	Guideline criteria for question 5
Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	**Yes** if power calculations are reported and an adequate sample size was used **No** if power calculations are reported and an adequate sample size was not reached **Unclear** if power calculations are not reported (adequate sample sizes may be calculated for each outcome when a clinically important difference has been determined) NA for single arm studies

Appendix 3 – Included and excluded studies

PRISMA flowcharts

Figure 13 summarises the volume of publications included and excluded at each stage of the review.

Figure 10 Summary of publications included and excluded at each stage of the review



[a] Verdoodt 2015 was identified as a relevant SLR at the protocol stage and was therefore included in the evidence synthesis

Appendix 4 – Study-level synthesis of results

Screening and surveillance intervals

Table 14 Quality assessments for the SLRs informing question 4 and question 5

Question	Arbyn 2014[35]	Verdoodt 2015[38]
Was an 'a priori' design provided?	Yes	Unclear
Was there duplicate study selection and data extraction?	Yes for selection, Unclear for data extraction	Yes
Was a comprehensive literature search performed?	Yes	Yes
Was the status of publication (ie. Grey literature) used as an inclusion criterion?	Yes	No
Was a list of studies (included and excluded) provided?	Yes	No
Were the characteristics of the included studies provided?	Yes	Yes
Was the scientific quality of the included studies assessed and documented?	Yes	Yes
Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes	No
Were the methods used to combine the findings of studies appropriate?	Yes	Yes
Was the likelihood of publication bias assessed?	Yes	No
Was the conflict of interest included?	No	No

Accuracy of self-sampling

Details of studies relevant to question 4 are presented in Table 15, and full quality assessments are presented in Table 16.

Table 15 Studies relevant to question 4

Study	Study design	Population characteristics	Screening methods	Epidemiology
Arbyn 2014[35] (SLR)	Design SLR and meta-analysis Objective To assess whether HPV testing on self-collected samples is equivalent to HPV testing on samples	Eligible studies Intervention: A vaginal sample was self-taken by a woman followed by a sample taken by a clinician or RCT with self-sample in 1 arm and clinician sample in the other A high-risk HPV DNA or RNA test was done on both samples or clinician sample examined microscopically for presence of	Self-testing In studies in a healthy screening population the devices used for self-sampling were: • Swab in 7 studies (5 papers) • Brush in 6 studies • Tampon in 1 study • Spatula in 1 study To define test positivity of the HPV test, the	Screening test accuracy Comparison of self-testing and clinician-testing: Pooled absolute sensitivity and specificity

	Self-sampling	Clinician-sampling		
		HPV	Cytology ASC-US+	Cytology LSIL+
CIN2 or worse				
Studies, n	16	16	12	8

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Epidemiology</th></tr>
<tr>
<td></td>
<td>collected by clinicians

<u>Dates</u>
Papers published between Jan 1, 1990 and June 3, 2013</td>
<td>cytological epithelial lesions

Presence or absence of CIN2+ verified by colposcopy or biopsy in all enrolled women with at least 1 positive test

<u>Patient recruitment:</u>
• Those attending routine cervical cancer screening (population of interest)
• High risk women
• Those referred to colposcopy because of previous positive screening results

<u>Data analysis</u>
The pooled absolute sensitivity and specificity of the tests were estimated jointly with metandi, a procedure in STATA, based on a bivariate model for the logit transformations of sensitivity and specificity, taking the intrinsic correlation between true positive and false positive rates and the variability between studies into account.

The relative sensitivity and specificity of HPV testing on self-samples compared with cytology HPV testing on clinician-taken samples using metadas, a SAS macro for the meta-analysis of diagnostic accuracy studies that allows the inclusion of type of test as a covariate, making comparison of tests possible.

<u>SLR results</u>
Sample size and demographics:

Data from 36 studies (reported in 34 papers) which altogether enrolled 154,556 women.

16 studies (14 papers) were in the population of interest for this review (primary screening of generally healthy</td>
<td>cut-off proposed by the manufacturer was accepted.

The reference standard (as per the inclusion criteria for the SLR) was colposcopy or biopsy in all enrolled women with at least 1 positive test.

<u>Clinician-testing</u>
34 of the selected studies used HPV testing on a clinician taken sample as the comparator. Additionally, the clinician taken samples were examined cytologically in 20 reports. 18 of these 20 studies included both cytology and HPV testing on the clinician-taken sample.

To define test positivity of the HPV test, the cutoff proposed by the manufacturer was accepted.

For cytological tests, 2 cutoffs were considered:

• Atypical squamous cells of undetermined significance (ASC-US) or worse
• Low-grade squamous intraepithelial lesions (LSIL) or worse
The reference standard (as per the inclusion criteria for the SLR) was colposcopy or biopsy in all enrolled women with at least 1 positive test.</td>
<td>
<table>
<tr><td>Sensitivity, % (95% CI) [range]</td><td>76 (69 to 82) [51% to 93%]</td><td>91 (87 to 94) [NR]</td><td>83 (75 to 89) [NR]</td><td>71 (66 to 76) [NR]</td></tr>
<tr><td>Specificity, % (95% CI) [range]</td><td>86 (83 to 89) [67% to 93%]</td><td>88 (85 to 91) [NR]</td><td>91 (87 to 94) [NR]</td><td>97 (97 to 98) [NR]</td></tr>
<tr><td colspan="5">CIN3 or worse</td></tr>
<tr><td>Studies, n</td><td>8</td><td>8</td><td>6</td><td>5</td></tr>
<tr><td>Sensitivity, % (95% CI) [range]</td><td>84 (72 to 92) [63% to 94%]</td><td>95 (91 to 97) [NR]</td><td>91 (85 to 95) [NR]</td><td>78 (72 to 85) [NR]</td></tr>
<tr><td>Specificity, % (95% CI)</td><td>87 (84 to 90)</td><td>89 (87 to 92)</td><td>89 (86 to 91)</td><td>97 (96 to 97)</td></tr>
</table>
Relative accuracy of HPV self-samples vs clinician-taken samples in all included studies^
<table>
<tr><th>CIN grade</th><th>Studies n</th><th>Relative sensitivity (95% CI)</th><th>Relative specificity (95% CI)</th></tr>
<tr><td colspan="4">HPV on self-samples vs HPV on clinician samples</td></tr>
<tr><td>CIN2 or worse</td><td>34</td><td>0.88 (0.85 to 0.91)*</td><td>0.96 (0.95 to 0.97)*</td></tr>
<tr><td>CIN3 or worse</td><td>12</td><td>0.89 (0.83 to 0.96)*</td><td>0.96 (0.93 to 0.99)*</td></tr>
<tr><td colspan="4">HPV on self-samples vs cytology (ASC-CU+) on clinician samples</td></tr>
<tr><td>CIN2 or worse</td><td>19</td><td>0.95 (0.91 to 0.99)*</td><td>0.92 (0.90 to 0.94)*</td></tr>
<tr><td>CIN3 or worse</td><td>6</td><td>0.99 (0.94 to 1.06)</td><td>0.98 (0.97 to 0.99)*</td></tr>
<tr><td colspan="4">HPV on self-samples vs cytology (LSIL+) on clinician samples</td></tr>
<tr><td>CIN2 or worse</td><td>11</td><td>1.14 (1.07 to 1.21)*</td><td>0.88 (0.86 to 0.90)*</td></tr>
<tr><td>CIN3 or worse</td><td>6</td><td>1.19 (1.09 to 1.29)*</td><td>0.90 (0.87 to 0.94)*</td></tr>
</table>
</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Epidemiology</th></tr>
<tr>
<td></td>
<td></td>
<td>
women)

Most common exclusion criteria for women in studies were pregnancy (in 10 studies), hysterectomy (in 7 studies), prior pelvic radiation (7 studies) and recent history of screening (in 7 studies)

<u>Study quality assessments</u>
Methodological quality of all included studies was assessed by QUADAS checklist and was overall moderate to good.

<u>Risk of bias in all included studies</u>
<table>
<tr><th>Risk</th><th>Studies n (%), n=36</th></tr>
<tr><td colspan="2">Enrolment of patients</td></tr>
<tr><td>Low</td><td>19 (53%)</td></tr>
<tr><td>Moderate</td><td>16 (44%)</td></tr>
<tr><td>High</td><td>1 (3%)</td></tr>
<tr><td colspan="2">Reporting and execution of index and comparator test</td></tr>
<tr><td>Adequate</td><td>26 (72%)</td></tr>
<tr><td>Unclear</td><td>10 (28%)</td></tr>
<tr><td>High</td><td>0 (0)</td></tr>
<tr><td colspan="2">Quality of the verification with a reference standard</td></tr>
<tr><td>Good</td><td>32 (89%)</td></tr>
<tr><td>Moderate</td><td>3 (8%)</td></tr>
<tr><td>Possibly problematic</td><td>1 (3%)</td></tr>
<tr><td colspan="2">Delay between self-sampling, clinician sampling and verification with reference standard</td></tr>
<tr><td>Short (<6 months)</td><td>25 (69%)</td></tr>
<tr><td>Long</td><td>9 (25%)</td></tr>
<tr><td>Unreported</td><td>2 (6%)</td></tr>
<tr><td colspan="2">Partial verification</td></tr>
<tr><td>Avoided</td><td>28 (78%)</td></tr>
<tr><td>Present</td><td>8 (22%)</td></tr>
<tr><td colspan="2">Differential verification</td></tr>
</table>
</td>
<td></td>
<td>^Studies pooled across settings (healthy screening population, high-risk patients and patients at follow up) due to low variation in relative sensitivity and specificity between self-collected and clinician collected samples

*Statistically significant</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Epidemiology</th></tr>
<tr><td></td><td></td><td>
<table>
<tr><td>Absent</td><td>35 (97%)</td></tr>
<tr><td>Present</td><td>1 (3%)</td></tr>
<tr><td colspan="2">Withdrawal of patients explained appropriately</td></tr>
<tr><td>Yes</td><td>25 (69%)</td></tr>
<tr><td>No</td><td>9 (25%)</td></tr>
<tr><td colspan="2">Poor reporting of uninterpretable assessed tests</td></tr>
<tr><td>Yes</td><td>20 (56%)</td></tr>
<tr><td>No</td><td>16 (44%)</td></tr>
<tr><td colspan="2">Poor reporting of uninterpretable reference standard</td></tr>
<tr><td>Yes</td><td>22 (61%)</td></tr>
<tr><td>No</td><td>12 (39%)</td></tr>
</table>
</td><td></td><td></td></tr>
<tr><td>Stanczuk 2016[36]

(PaVDaG study)</td><td><u>Design</u>
Prospective cohort study

<u>Objective</u>
To assess the performance of a high-risk human papillomavirus (hrHPV) PCR-based assay to detect CIN2+ in self-collected vaginal and urine samples

<u>Dates</u>
April 2013 to July 2014

<u>Country</u>
Scotland

<u>Setting</u>
Primary care</td><td><u>Patient recruitment</u>
All women, other than those previously diagnosed with CIN2+, attending routine screening in primary care were invited to consent to the study

<u>Data collection</u>
Participants first provided a random void urine sample, then self-collected a vaginal sample prior to a routine cervical sample being collected by a clinician.

Participants with high grade abnormalities were referred for colposcopy and women with borderline changes or low-grade cytology were recalled for repeat cytology after 6 months. Women with 2 low-grade or three borderline smears were referred for colposcopy.

<u>Sample size and demographics</u>
5,318 women

Mean age: 41.3 years (17 to 76)

Median age: 46 (<20=7, >59=145)</td><td><u>Self-testing</u>
Urine collected in universal containers, 6 ml was mixed with 3 ml of Roche PCR media.

Self-collected vaginal samples were obtained using cobas PCR female swab sample packets, women were advised to follow printed instructions, swabs were immediately immersed in tubes containing Roche PCR media.

Samples were tested with the cobas 4800 DNA HPV test using the standard procedure.

<u>Clinician-testing</u>
Cervical LBC samples were clinician collected using a Rovers Cervex-Brush and suspended in 20 ml of ThinPrep solution. 3 ml of this sample was aliquoted for HPV testing.

Samples were tested with the cobas 4800 DNA HPV test using the standard procedure.

negative cervical cytology results did not have underlying cervical precancer.

<u>Reference standard</u></td><td>Screening test accuracy

Comparison of self-testing (vaginal and urine samples) and clinician-collected cervical samples
<table>
<tr><th></th><th>Self-collected vaginal samples n=5208</th><th>Self-collected urine samples n=5003</th><th>Clinician collected cervical samples n=5299</th></tr>
<tr><td colspan="4">CIN2 or worse</td></tr>
<tr><td>Sensitivity (95% CI)</td><td>94.6% (90.7 to 98.5)</td><td>63.1% (54.6 to 71.7)</td><td>97.7% (95.0 to 100)</td></tr>
<tr><td>Specificity (95% CI)</td><td>85.4% (84.4 to 86.3)</td><td>59.8% (89.0 to 90.7)</td><td>87.3% (86.4 to 88.2)</td></tr>
<tr><td>LR+</td><td>6.48</td><td>1.57</td><td>7.69</td></tr>
<tr><td>LR-</td><td>0.054</td><td>0.62</td><td>0.026</td></tr>
<tr><td>Relative sensitivity (95% CI)</td><td>0.97 (0.94 to 1.00), p 0.1250</td><td>0.67 (0.59 to 0.76), p<0.0001</td><td>Reference</td></tr>
<tr><td>Relative specificity (95% CI)</td><td>0.98 (0.97 to 0.99), p<0.0001</td><td>1.05 (1.04 to 1.06), p<0.0001</td><td>Reference</td></tr>
<tr><td colspan="4">CIN3 or worse</td></tr>
<tr><td>Sensitivity</td><td>95.8% (91.1</td><td>50.7% (39.1</td><td>98.6% (95.9</td></tr>
</table>
</td></tr>
</table>

Study	Study design	Population characteristics	Screening methods	Epidemiology
		Of women aged ≤23 years, 66% (354/533) had been vaccinated with at least 2 doses of bivalent HPV vaccine	The reference standard for both tests was colposcopy. In the absence of a reference standard result it was assumed that women with no history of CIN2+ and 2 previous consecutively negative cervical cytology results did not have underlying cervical precancer.	

	to 100)	to 62.3)	to 100)
Specificity	84.8% (83.8 to 85.8)	89.7% (88.8 to 90.5)	86.4% (85.5 to 87.3)
LR+	6.30	4.92	7.25
LR-	0.045	0.49	0.016
Relative sensitivity (95% CI)	0.97 (0.93 to 1.01), p 0.5000	0.53 (0.42 to 0.67), p<0.0001	Reference
Relative specificity (95% CI)	0.98 (0.97 to 0.99), p<0.0001	1.03 (1.02 to 1.04) , p<0.0001	Reference

Table 16 Quality assessment of studies relevant to question 4

	Stanczuk 2016[36]
PARTICIPANT SELECTION	
Was a consecutive or random sample of patients enrolled?	Yes
Was a case-control design avoided?	Yes
Did the study avoid inappropriate exclusions?	Yes
Could the selection of patients have introduced bias?	Low
Is there concern that the included patients do not match the review question?	No
INDEX TEST	
Were the index test results interpreted without knowledge of the reference standard?	Yes
If a threshold was used, was it pre-specified?	Unclear
Could the conduct or interpretation of the index test have introduced bias?	Low
Is there concern that the index test, its conduct, or interpretation differ from the review question?	No
REFERENCE STANDARD	
Is the reference standard likely to correctly classify the test condition?	Yes
Were the reference standard results interpreted without knowledge of the results of the index test?	Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?	Low
Is there concern that the target condition as defined by the reference standard does not match the review question?	No
PARTICIPANT FLOW	
Was there an appropriate interval between the index test(s) and the reference standard?	Unclear
Did all patients receive a reference standard?	No

	Stanczuk 2016[36]
Did patients receive the same reference standard?	Yes
Were all patients included in the analysis?	No
Could the patient flow have introduced bias?	High

Uptake of self-sampling

Details of studies relevant to question 5 are presented in Table 17, and full quality assessments are presented in Table 18.

Table 17 Studies relevant to question 5

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td>Verdoodt 2015[38] (SLR)</td>
<td><u>Design</u>
SLR and Meta-analysis

<u>Objective</u>
To evaluate if offering a kit for self-sampling (at home) could increase screening attendance, compared to sending reminder letters for a Pap smear or HPV test on a sample collected by a clinician (at the clinic)

<u>Dates</u>
Up to 12th February 2015</td>
<td><u>Eligible studies</u>
Population:
• Study population involved irregularly* or never-screened women, or women who did not respond to ≥1 invitation for conventional cervical cancer screening (collectively ‘under-screened women)
• A minimum of 1000 women were included in the study
*Women were considered to be irregularly screened if their last screening exceeded the locally defined screening interval

Intervention:
• Women in the interventional group (self-sampling arm) were invited to collect a self-sample for hrHPV testing
• Women in the control group were invited to undergo conventional cytology screening and/or hrHPV testing on a sample taken by a clinician
Outcome:
• The participation in the self-sampling arm was documented

<u>Data analysis</u>
Per protocol and ITT analyses performed, the latter includes data on women who were invited to perform a self-sample but instead opted to have a Pap smear test</td>
<td><u>Self-testing</u>
Kits were distributed by the following methods:
• Kit mailed directly to the home address of all women (Mail-to-all)
• Women were sent an invitation to order a self-sampling kit (Opt-in)
• Women were approached at their home and offered a self-sampling kit (Door-to-door)

<u>Clinician-testing</u>
In 12 studies, women in the control arm were invited for cytology.

In 2 studies there were arms for cytology and HPV testing.

In 2 studies only hrHPV testing was performed in the control arm.</td>
<td><u>Screening test uptake</u>
Comparison of self-testing and clinician-testing

*Low resource setting, women with limited access to health care as
<table>
<tr><th></th><th></th><th colspan="2">Participation</th></tr>
<tr><th></th><th>Studies, n</th><th>Self-sampling, % (95% CI)</th><th>Clinician-sampling, % (95% CI)</th></tr>
<tr><td colspan="4">Per-protocol</td></tr>
<tr><td>Mail-to-all</td><td>13</td><td>20.7
(16.9 to 24.8)
[range: 6.4% to 34.0%]</td><td>10.3
(6.2 to 15.2)</td></tr>
<tr><td>Opt-in</td><td>3</td><td>9.7
(6.5 to 13.5)</td><td>12.2
(10.9 to 13.6)</td></tr>
<tr><td>Door-to-door*</td><td>2</td><td>91.3
(65.8 to 100)</td><td>54.1
(0.9 to 100)</td></tr>
<tr><td colspan="4">ITT</td></tr>
<tr><td>Mail-to-all</td><td>13</td><td>23.6
(20.2 to 27.3)
[range: 10.2% to 39.0%]</td><td>10.3
(6.2 to 15.2)</td></tr>
<tr><td>Opt-in</td><td>3</td><td>14.0
(8.0 to 21.4)</td><td>12.2
(10.9 to 13.6)</td></tr>
<tr><td>Door-to-door*</td><td>2</td><td>92.4
(71.3 to 100)</td><td>54.1
(0.9 to 100)</td></tr>
</table>
opposed to having previously resistant to screening

Relative participation and participation difference in self-sampling vs</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td></td>
<td></td>
<td>
taken by a clinician.

Pooled proportions were calculated be a random effects model using metaprop, a statistical procedure for meta-analysis of binomial data. Relative rates and absolute differences were assessed by applying random effects models using metan.

<u>SLR results</u>
163,431 women in 16 studies included in the meta-analysis (2 studies in low resource settings)

134,262 women in 14 studies in high-resource settings

<u>Baseline characteristics</u>
'Under-screened':

Women who did not respond to 1 invitation for regular screening or an invitation and a reminder were eligible: 10 studies

Women who persistently did not respond to invitations for regular screening (more than 2 rounds of screening) were eligible: 4 studies

Women with limited access to health services in low-resource settings: 2 studies

<u>Quality assessment</u>
Study quality evaluated moderate to high, based on the Cochrane tool for bias

<table>
<tr><th rowspan="2">Risk of bias</th><th colspan="2">Selection</th><th>Attrition</th><th colspan="2">Reporting</th></tr>
<tr><th>S1</th><th>S2</th><th>A</th><th>R1</th><th>R2</th></tr>
<tr><td>Bais 2007</td><td>L</td><td>M</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Gok 2010</td><td>L</td><td>M</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Gorgi Rossi</td><td>L</td><td>L</td><td>L</td><td>L</td><td>L</td></tr>
</table>
</td>
<td></td>
<td>
control arm
^Two studies had 2 control arms (15 comparisons)

<table>
<tr><th></th><th>Studies, n</th><th>Relative participation, (95% CI)</th><th>Participation difference, % (95% CI)</th></tr>
<tr><td colspan="4">Per-protocol</td></tr>
<tr><td>Mail-to-all</td><td>13^</td><td>2.06
(1.44 to 2.96)</td><td>9.9
(5.8 to 13.9)</td></tr>
<tr><td>Opt-in</td><td>3†</td><td>0.72
(0.53 to 0.99)</td><td>-3.2 (-6.6 to 0.1)</td></tr>
<tr><td>Door-to-door</td><td>2</td><td>2.17
(0.33 to 14.13)</td><td>36.0
(-16.6 to 88.5)</td></tr>
<tr><td colspan="4">ITT</td></tr>
<tr><td>Mail-to-all</td><td>13^</td><td>2.40
(1.73 to 3.33)</td><td>12.6
(9.3 to 15.9)</td></tr>
<tr><td>Opt-in</td><td>3†</td><td>0.97
(0.65 to 1.46)</td><td>0.2
(-4.5 to 4.9)</td></tr>
<tr><td>Door-to-door</td><td>2</td><td>2.21
(0.32 to 15.48)</td><td>37.5
(-17.7 to 92.8)</td></tr>
</table>
†Two studies had 2 control arms (5 comparisons)

<u>Unsatisfactory tests</u>
Data on sample adequacy in the self-sampling are was reported in 12 studies.

The pooled proportion of unsatisfactory samples: 0.7% (95% CI 0.4 to 1.1%)
</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td></td>
<td></td>
<td>
<table>
<tr><td>2011</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Lazcano-Ponce 2011</td><td>M</td><td>L</td><td>L</td><td>M</td><td>M</td></tr>
<tr><td>Piana 2011</td><td>L</td><td>M</td><td>L</td><td>M</td><td>H</td></tr>
<tr><td>Szarewski 2011</td><td>M</td><td>M</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Virtanen 2011</td><td>L</td><td>M</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Wikstrom 2011</td><td>M</td><td>M</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Gok 2012</td><td>L</td><td>M</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Darlin 2013</td><td>M</td><td>M</td><td>L</td><td>M</td><td>M</td></tr>
<tr><td>Sancho-Garnier 2013</td><td>M</td><td>M</td><td>L</td><td>L</td><td>M</td></tr>
<tr><td>Broberg 2014</td><td>M</td><td>M</td><td>L</td><td>M</td><td>L</td></tr>
<tr><td>Cadman 2014</td><td>L</td><td>L</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Haguenoer 2014</td><td>L</td><td>L</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Arrossi 2015</td><td>M</td><td>L</td><td>L</td><td>L</td><td>L</td></tr>
<tr><td>Giorgi Rossi 2015</td><td>L</td><td>L</td><td>L</td><td>L</td><td>L</td></tr>
</table>
S1: Random sequence generation, S2: Allocation concealment, A: Incomplete outcome data, R1: Reporting of timelines, R2: Selective reporting

L: Low risk, M: Medium risk, H: high risk
</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosgraaf 2015[39]

PROHTECT-3B</td>
<td><u>Design</u>
RCT

<u>Objective</u>
To assess the participation rate of 2 self-sampling methods and physician taken-smear, to compare performance of the sampling methods for detection of</td>
<td><u>Patient recruitment</u>
Women who did not respond to an invitation for a cervical smear in 2008, and living in the regions of North Holland Flevoland, Utrecht and Gelderland were invited to participate. All eligible women could opt-out.

Exclusion criteria:

• Previous hysterectomy
• Previous abnormal cytological test result within the last 2 years
• Current pregnancy

<u>Data collection</u></td>
<td><u>Self-testing</u>
Women were randomised (1:1) to receive either a brush device (Evalyn Brush) or a lavage device (second generation Delphi Screener) which was provided in a self-sampling kit with an explanatory letter, an informed consent form, user instructions, a questionnaire and a return envelope.</td>
<td><u>Screening test uptake</u>
Comparison of brush device and lavage device
<table>
<tr><th></th><th>Brush device, n=15,077</th><th>Lavage device, n=15,053</th><th>Eligible population, n=33,279</th></tr>
<tr><td>Participation rate, n (%, 95% CI]</td><td>5,218 (34.6, 33.9 to 35.4)</td><td>4,809 (31.9, 31.2 to 32.7)</td><td>10,027 (30.1)</td></tr>
<tr><td>Absolute difference, % (95% CI)</td><td colspan="2">2.7 (1.8 to 4.2)</td><td></td></tr>
</table>
Also reports participation rate in age ranges</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr><td></td><td>hrHPV and CIN2+ or CIN3+ and to investigate acceptability and user friendliness of the devices

<u>Dates</u>
Oct 2011 to Feb 2012

<u>Country</u>
Netherlands

<u>Setting</u>
Home testing - National Screening Programme</td><td>Women were sent self-sampling kits which they returned by post, all women who submitted self-samples between October 2011 and December 2012 were counted as self-sampling responders

<u>Sample size and demographics</u>
35,477 women were invited to take part, of these 33,279 were eligible, 5,347 (15.1%) opted out (3,149 of whom were eligible)

30,130 women randomised</td><td></td><td><u>Inadequate samples</u>
23/5,218 (0.4%) [95% CI 0.25 to 0.61] brush samples were inadequate for evaluation

24/4809 (0.5%) [95% CI 0.32 to 0.74] lavage samples were inadequate for evaluation</td></tr>
<tr><td>Ducancelle 2015[40]

CapU study</td><td><u>Design</u>
Prospective cohort study

<u>Objective</u>
To evaluate the participation rate of urinary HPV testing

<u>Dates</u>
July 2010 to Jan 2013

<u>Country</u>
France

<u>Setting</u>
Home testing</td><td><u>Patient recruitment</u>
Women aged 40 to 65 who had not responded to previous invitations and reminders for pap smears

<u>Exclusion criteria</u>
Previous hysterectomy

<u>Data collection</u>
Women accepting to participate returned a sample by mail to the Angers University Hospital Virology Laboratory, response rate determined on reception of informed consent forms and urine samples

<u>Sample size and demographics</u>
5,000 women aged 40 to 65

3,000 in a 40 to 54 years age group and 2,000 in a 55 to 65 years age group</td><td><u>Self-testing</u>
Women received an invitation letter with a urinary HPV DNA testing information note, a letter of consent a sterile container, a procedure protocol, a survey on the motives for refusal of the smear, a bubble envelope and a prepaid return envelope</td><td><u>Screening test uptake</u>
13.7% overall participation rate
<table><tr><th></th><th>40 to 54 years of age, n=3000</th><th>55 to 65 years of age, n=2000</th><th>p-value</th></tr><tr><td>Uptake, n (%)</td><td>512 (17)</td><td>259 (12.9)</td><td></td></tr><tr><td>Participation rate, n (%) [95% CI]</td><td>479 (15.9) [14.61 to 17.26]</td><td>208 (10.7) [9.38 to 12.14]</td><td><0.001</td></tr></table>Uptake represents number who returned samples, participation rate is a measure of the samples which were eligible for inclusion, women were considered ineligible due to prior hysterectomy or for refusal to participate in study.

<u>Inadequate samples</u>
• 40 to 54 years of age: 2 invalid tests (from eligible population)
• 55 to 65 years of age: 1 invalid test (from eligible population)
In total 3/687 (0.4%) [95% CI 0.07 to 1.22] samples could not be</td></tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr><td></td><td></td><td></td><td></td><td>analysed</td></tr>
<tr><td>Duke 2015[41]</td><td><u>Design</u>
Community based cohort study, case control

<u>Objective</u>
To determine whether offering self-collected HPV testing screening increased cervical cancer screening rates in rural communities

<u>Dates</u>
2010 to 2011

<u>Country</u>
Canada

<u>Setting</u>
Rural communities in the Canadian province of Newfoundland and Labrador</td><td><u>Patient recruitment</u>
All eligible women living in rural communities in the Canadian province of Newfoundland and Labrador

<u>Exclusion criteria</u>
Pregnancy

<u>Data collection</u>
In community A women were considered responders if they returned a self-collection kit, in community B women were considered responders if they presented for Pap smear and agreed to be part of the study; providing information about themselves and their screening history.

<u>Sample size and demographics</u>
1,760 women in community A, 2,761 women in community B and 1,536 women in community C at the end of the study</td><td><u>Self-testing</u>
Women in community A were given the opportunity of being screened for HPV infection through vaginal self-screening. A self-collection kit containing a Dacron swab, collection tube, instructions with explanatory pictures, consent forms and a participant questionnaire were available at public locations or a research nurse was available to drop off kits at a women's home or work. A kit was also available at the end of an educational presentation on cervical cancer screening.

<u>Clinician-testing</u>
In community B a campaign was used to raise awareness on the importance of regular screening through pap smear.

Women in community C received no intervention</td><td><u>Screening test uptake</u>
Return rate of self-collection kits

• 168/837 (20.1)
Comparison of response to self-testing and Pap smear invitation in eligible women

<table><tr><th></th><th>Community A: self-testing, n=1760</th><th>Community B: Pap smear, n=2761</th></tr><tr><td>Response rate, n (%) [95% CI]</td><td>168 (9.5) [7.80 to 11.43]</td><td>374 (13.5) [12.63 to 15.25]</td></tr><tr><td>Proportion of under- or unscreened participants, n (%)</td><td>26 (15.5)</td><td>52 (13.9)</td></tr></table>
Change in cervical cancer screening rates, 2008/2009 compared to 2010/2011

<table><tr><th></th><th>Community A</th><th>Community B</th><th>Community C</th></tr><tr><td>Cervical cancer screening rate 2008/2009, n/N[a] (%)</td><td>1,020/1,928 (52.9)</td><td>1,484/2,833 (52.4)</td><td>1,098/1,524 (72.0)</td></tr><tr><td>Cervical cancer screening rate 2010/2011, n/N[a] (%)</td><td>1,187/1,760 (67.4)</td><td>1,529/2,761 (55.3)</td><td>1,236/1,536 (80.5)</td></tr><tr><td>Change in rate, %</td><td>+15.2</td><td>+2.9</td><td>+8.5</td></tr><tr><td>p-value for change</td><td><0.001</td><td>0.07</td><td><0.01</td></tr><tr><td>p-value for difference in change</td><td>Reference</td><td><0.001</td><td>0.193</td></tr></table>
[a]Denominator (eligible population of women age 30 to 69) for the 2008/2009 period based on 2006 census data and for the 2010/2011 period is based on the 2011 census data

Cervical cancer screening rates for Women in Community A for 2010/2011 was determined as the number of women who had a pap smear and the number who did self-collection but did not have a pap smear</td></tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td>Enerly 2016[42]

SESAM study</td>
<td><u>Design</u>
RCT

<u>Objective</u>
To demonstrate the effects of self-sampling among women who do not attend the NCCSP, in particular:
• Impact of the self-sampling on screening attendance and coverage
• The performance of 2 different self-sampling devices for hrHPV testing
• Women's experience of the 2 self-sampling devices used

<u>Dates</u>
April/May 2013

<u>Country</u>
Norway

<u>Setting</u>
Home testing in National Screening Programme</td>
<td><u>Patient recruitment</u>
Non-attenders to the Norwegian Cervical Cancer Screening Programme (NCCSP), defined as a woman aged 26-69 years without any cytology, HPV or histology result recorded in the NCCSP registries within 12 months of the first reminder, identified in Oslo in April/May 2013

<u>Data collection</u>
Screening attendance was defined as either returning a self-sampling device and/or having a cervical smear taken by a clinician between April 2013 and the end of 2013

<u>Sample size and demographics</u>
3,393 women

800 assigned to the 'intervention group' (300 each from the age groups 26 to 34 and 35 to 49 years and 200 from the age group 50 to 69 years), 729 women were successfully contacted and consented to their inclusion in the trial

2,593 served as the control group</td>
<td><u>Self-testing</u>
Patients selected for the intervention group were sent an information letter inviting them to participate in the study. Those participating the study were randomized and sent 1 of 2 self-sampling devices along with user instructions, an informed consent form, a pre-paid return envelope and a questionnaire.

The self-sampling devices used in the study:
• Lavage based sampler: Delphi Screener™
• Dry brush sampler: Evalyn Brush

<u>Clinician-testing</u>
Followed according to the established procedures of the NCCSP; if no cytology result is recorded within 12 months of the initial reminder letter women were sent a second reminder, each woman is responsible for scheduling her own screening appointment</td>
<td><u>Screening test uptake</u>
Comparison of intervention and control group
<table>
<tr><th></th><th colspan="3">Intervention group</th><th>Control group</th></tr>
<tr><th></th><th>Delphi Screener n=400</th><th>Evalyn Brush n=400</th><th>Total n=800</th><th>Total n=2,593</th></tr>
<tr><td>Self-testing participants, n (%) [95% CI]</td><td>81 (20.3) [16.47 to 24.58]</td><td>88 (22.0) [18.04 to 26.38]</td><td>169 (21.1)</td><td></td></tr>
<tr><td>Cytology participants, n (%)</td><td>53 (13.3)</td><td>45 (11.3)</td><td>98 (12.3)</td><td></td></tr>
<tr><td>Total participants, n (%) [95% CI]</td><td>134 (33.5) [28.89 to 38.36]</td><td>133 (33.3) [28.70 to 38.15]</td><td>267 (33.4)</td><td>601 (23.2) [21.59 to 24.87]</td></tr>
</table>
ITT population

Attendance rates in the intervention and control groups are also reported by age (26 to 34, 35 to 49 and 50 to 69 years)

Total participation intervention/control arm relative risk: 1.44 (95% CI 1.28 to 1.62)

<u>Useable samples</u>
All 169 devices returned contained sufficient biological material for HPV testing</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td>Kitchener 2016[47]

STRATEGIC</td>
<td><u>Design</u>
RCT

<u>Objective</u>
To evaluate the clinical effectiveness of a range of interventions in:
• All women receiving their first invitation for cervical screening
• Those who had not attended by 6 months
To evaluate the cost effectiveness of these interventions and to study preferences for cervical screening among non-attenders.

<u>Dates</u>
April 2012 to June 2014 in north-west England

Oct 2012 to Dec 2014 in north-east Scotland

<u>Country</u>
UK (England and Scotland)

<u>Setting</u>
Home testing and</td>
<td><u>Patient recruitment</u>
Non-attenders to screening in Phase 1 of the trial during which women were sent their first routine invitation to attend cervical screening. Non-attenders were women who had no record of cytology test 6 months after their test date.

Women in Greater Manchester aged 24.75 years, women in Grampian aged 20 years.

<u>Data collection</u>
Data on uptake were obtained from the screening agency (Lancashire and South Cumbria Agency) in Greater Manchester and from the research team in Grampian, primary time point for uptake was 12 months following standard invitations

<u>Sample size and demographics</u>
10,126 women randomised (from 258 practices)
• 1,141 women received an unrequested self-sample kit
• 1,290 women received a letter offering a self-sample kit
• 3,782 women from 97 practices served as controls.
Patients were also randomised to be offered a nurse navigator (n=1,007), a timed appointment (n=1,629) or the option of a nurse navigator or self-sample kit (n=1,277)

<u>Baseline characteristics</u>
<table><tr><th>Characteristic</th><th>Women in Grampian, n=2,608</th></tr><tr><td colspan="2">Vaccination status</td></tr><tr><td>None</td><td>708</td></tr><tr><td>Incomplete</td><td>149</td></tr><tr><td>Full</td><td>1,724</td></tr><tr><td>Missing</td><td>27</td></tr></table></td>
<td><u>Self-testing</u>
There were two HPV self-sampling interventions:
• A letter offering the opportunity to request a self-sample kit
• An unrequested self-sample kit sent directly to the home
The self-sample kit comprised either a Delphi lavage device or The Rovers® Evalyn-Brush, an information sheet, a consent form and packaging to return the sample

<u>Clinician-testing</u>
Patients in the control arm were sent their first routine invitation for screening and received no further intervention</td>
<td><u>Screening test uptake</u>
Comparison of self-sampling kit sent, self-sampling kit offered and control groups (clinician sampling)
<table><tr><th>Intervention</th><th>Attendance, n (%) [95% CI]</th><th>OR[a] (95% CI)</th></tr><tr><td colspan="3">12 month follow up</td></tr><tr><td>Control, n=3,782</td><td>613 (16.2) [15.04 to 17.41]</td><td>Reference</td></tr><tr><td>Self-sampling sent, n=1,141</td><td>243 (21.3) [18.96 to 23.79]</td><td>1.512 (1.197 to 1.910), p=0.001</td></tr><tr><td>Self-sampling offered, n=1290</td><td>209 (16.2) [14.23 to 18.33]</td><td>1.074 (0.871 to 1.325), p=0.505</td></tr><tr><td colspan="3">18 month follow up</td></tr><tr><td>Control, n=3,782</td><td>27.1 (1026)</td><td>Reference</td></tr><tr><td>Self-sampling sent, n=1,141</td><td>30.0 (342)</td><td>1.286 (1.056 to 1.567), p=0.012</td></tr><tr><td>Self-sampling offered, n=1,290</td><td>25.8 (333)</td><td>1.056 (0.884 to 1.262), p=0.548</td></tr></table>[a] Adjusted OR associated with the change in odds of attendance occurring with intervention compared with control, adjusted for practice attendance rate and Primary Care Trust region

Attendance based on location, Greater Manchester or Grampian, is also reported

Type of screening undergone by participants
<table><tr><th></th><th colspan="4">Type of screen</th></tr><tr><th></th><th colspan="2">Single</th><th colspan="2">Both</th></tr><tr><th></th><th>HPV only</th><th>Cytology only</th><th>HPV first</th><th>Cytology first</th></tr><tr><td colspan="5">12 month follow up</td></tr><tr><td>Control, n=613</td><td>1</td><td>612</td><td>-</td><td>-</td></tr><tr><td>Self-sampling sent, n=243</td><td>52</td><td>158</td><td>32</td><td>1</td></tr><tr><td>Self-</td><td>12</td><td>190</td><td>7</td><td>-</td></tr></table></td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr><td></td><td>primary care</td><td></td><td></td><td>
<table>
<tr><td>sampling offered, n=209</td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="5">18 month follow up</td></tr>
<tr><td>Control, n=1,026</td><td>1</td><td>1025</td><td>-</td><td>-</td></tr>
<tr><td>Self-sampling sent, n=342</td><td>59</td><td>248</td><td>34</td><td>1</td></tr>
<tr><td>Self-sampling offered, n=333</td><td>12</td><td>314</td><td>7</td><td>-</td></tr>
</table>
ITT population</td></tr>
<tr><td>Racey 2016[48]</td><td><u>Design</u>
RCT

<u>Objective</u>
To determine if cervical cancer screening uptake would increase among under-screened women living in rural Ontario, Canada, if at home self-collected sampling for HPV testing was offered as a primary cervical cancer screening modality, compared to invited Pap testing to routine opportunistic screening</td><td><u>Patient recruitment</u>
Women (aged 30 to 70 years) were identified as being under-screened/overdue for screening through their electronic medical record system, this was defined as not having had a pap test recorded in the preceding 30 months

<u>Exclusion criteria:</u>
• Residing in a long-term care facility
• Medical history of hysterectomy
• Any other medical contraindication
• Invalid mailing address
• Inactivated medical chart
Final eligibility was determined post-randomisation

<u>Data collection</u>
All women who participated in the self-collected HPV test had their results recorded in their medical chart.

Pap test completion was recorded from the medical charts at the end of the study period for eligible women in the study</td><td><u>Self-testing</u>
Women were sent a study information letter informing them about the study and giving them the option to opt-out 2 weeks before the self-collection kit was sent.

The self-collection kit contained a vaginal swab, collection tube, annotated pictorial instructions, a questionnaire, an information sheet on cervical cancer and HPV and a return envelope.

A reminder phone call was placed to non-responders 1 month after self-collection kits were sent

<u>Clinician-testing</u>
Women in the Pap testing arm were sent an invitation letter that asked for them to call their doctor and book an</td><td><u>Screening test uptake</u>
Comparison of self-testing, pap testing and opportunistic screening
<table>
<tr><th colspan="2"></th><th>Screened, n (%) [95% CI)</th></tr>
<tr><td rowspan="3">Self-collection arm, n=335</td><td>Self-collection sample</td><td>70 (21) [16.76 to 25.76]</td></tr>
<tr><td>Pap testing</td><td>37 (11)</td></tr>
<tr><td>Total</td><td>107 (32) [27.03 to 37.29]</td></tr>
<tr><td colspan="2">Pap invitation arm, n=331</td><td>51 (15.4) [11.33 to 19.31]</td></tr>
<tr><td colspan="2">Standard of care, n=152</td><td>13 (8.6) [8.10 to 19.41]</td></tr>
</table>
• Women in self-collection arm were 3.7 (95% CI 2.2 to 6.4) times more likely to undergo screening compared with the standard of care arm
• Women in Pap test arm were 1.8 (95% CI 1.0 to 3.2) times more likely to screen compared to women in the standard of care arm
• Women in self-collection arm were 2.1 (95% CI 1.5 to 2.8) times more likely to undergo screening compared with women in the Pap test arm (p=0.097)</td></tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td></td>
<td><u>Dates</u>
October 2012 to July 2013

<u>Country</u>
Canada

<u>Setting</u>
Rural community where low rates of cervical cancer screening have been observed, in partnership with primary care</td>
<td>A modified ITT analysis was used for all post-randomisation eligible women to calculate the RR for each arm.

<u>Sample size and demographics</u>
964 women identified as under-screened and randomised:

• 400 to self-collection
• 400 to Pap invitation
• 164 to standard of care opportunistic screening- women seeking cervical cancer screening through their own initiative, with or without prompting from a healthcare provider
After adjusting for eligibility:

• 818 eligible women
• 335 received a self-collected HPV testing kit
• 331 received a reminder letter
• 152 received standard of care opportunistic screening
No women contacted the clinic to opt-out

<u>Baseline characteristics</u>
<table>
<tr><th>Characteristic</th><th>Self-collection arm</th><th>Pap test arm</th></tr>
<tr><td>Mean age, years (95% CI)</td><td>53.6 (51.2 to 56.0)</td><td>50.5 (46.0 to 55.0)</td></tr>
<tr><td>Age, years, n (%)</td><td>n=76</td><td>n=24</td></tr>
<tr><td>30 to 39</td><td>7 (9.2)</td><td>3 (13.0)</td></tr>
<tr><td>40 to 49</td><td>22 (29.0)</td><td>8 (34.8)</td></tr>
<tr><td>50 to 59</td><td>18 (23.7)</td><td>9 (39.1)</td></tr>
<tr><td>60+</td><td>29 (38.2)</td><td>4 (16.7)</td></tr>
<tr><td>Screening history, n (%)</td><td>n=76</td><td>n=23</td></tr>
<tr><td>Prior Pap test, yes</td><td>75 (98.7)</td><td>23 (100)</td></tr>
<tr><td>3 years or more since last Pap test</td><td>47 (62.7)</td><td>14 (60.9)</td></tr>
</table>
</td>
<td>appointment, in addition to an information sheet HPV and cervical cancer screening. Women who did not respond within 1 month were called by the clinic to follow-up and book an appointment if possible, a change in the protocol during the trial led to only 20% of the women in the Pap invitation arm receiving a follow up call due to a shortage in resources.

Women in the opportunistic screening arm were not contacted during the study period.</td>
<td>Underpowered exploratory sub analysis (per protocol):

• Uptake of self-collected sampling for HPV testing vs the standard of care arm: RR= 2.4 (95% CI 1.4 to 4.3), significantly higher
• Uptake of self-collected sampling for HPV testing vs Pap test arm: RR= 1.4 (95% CI 0.98 to 1.9), no significant difference

<u>Adequate samples</u>
1/70 (1.4%) [95% CI 0.03 to 7.66] samples were not β-globin positive, which demonstrates a high DNA sample quality</td>
</tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr><td></td><td></td><td>
<table>
<tr><td><2 years since last Pap test</td><td>24 (32.0)</td><td>5 (21.7)</td></tr>
<tr><td>Do not remember</td><td>4 (5.3)</td><td>4 (17.4)</td></tr>
</table>
</td><td></td><td></td></tr>
<tr><td>Sultana 2016[43]

iPap</td><td><u>Design</u>
RCT

<u>Objective</u>
To determine if HPV self-sampling could increase participation in the Australian cervical cancer screening program

<u>Dates</u>
March to July 2014

<u>Country</u>
Australia

<u>Setting</u>
Home testing and testing in primary care, data from Pap test registers</td><td><u>Patient recruitment</u>
Women who were residents of Victoria and were identified through the Victorian Cervical Cytology Register as never-screened (women on the electoral role but for whom no match was found on the registry) or under-screened (not screened in the past 5 years)

<u>Eligibility criteria:</u>
• Aged 30 to 69 years
• Not pregnant
• Not had a hysterectomy

<u>Data collection</u>
Primary outcome was participation in screening at three and 6 months after initial letters were mailed, indicated by returning a self-sampling swab or having a Pap test, women who had Pap tests after randomisation were identified by performing a semi-automated match of the trial database with Registry records of Pap tests conducted in 2014

<u>Sample size and demographics</u>
8,160 women

7,140 in the self-sampling arm and 1,020 in the Pap test arm

<u>Baseline characteristics</u>
<table>
<tr><th></th><th colspan="2">Apparently never screened</th><th colspan="2">Apparently under-screened</th></tr>
<tr><th></th><th>S, n= 7,140</th><th>P, n= 1,020</th><th>S, n= 7,140</th><th>P, n= 1,020</th></tr>
<tr><td colspan="5">Age, years n (%)</td></tr>
<tr><td>30 to 39</td><td>1,950 (27.3)</td><td>276 (27.1)</td><td>2,334 (32.7)</td><td>323 (31.7)</td></tr>
</table>
</td><td><u>Self-testing</u>
Women randomised to self-sampling arm were sent a pre-invitation letter, informing them that they would be receiving a kit and giving the opportunity to withdraw from the trial. Participants were then sent a package containing an information brochure, a nylon-tipped flocked swab enclosed in a dry plastic tube within a re-sealable plastic bag, an instruction sheet, a personal information form and a postage paid envelope

<u>Clinician-testing</u>
Women received a single invitation letter (never screened population) or a standard reminder letter (under-screened population) to have a pap test, this included a pap test brochure, a personal information form and a postage paid envelope</td><td><u>Screening test uptake</u>
Comparison of response within 6 months to self-testing and clinician-testing invitations in never-screened and under-screened population
<table>
<tr><th></th><th colspan="3">Self-sampling arm, n=7,140</th><th>Pap test arm, n=1,020</th><th rowspan="2">Absolute difference (95% CI)</th></tr>
<tr><th></th><th>Self-sampling</th><th>Pap test</th><th>Total</th><th>Pap test</th></tr>
<tr><td>Never screened, n (%) [95% CI]</td><td>1,131 (15.8) [14.96 to 16.67]</td><td>321 (4.5)</td><td>1,452 (20.3) [19.37 to 21.25]</td><td>61 (6) [4.62 to 7.64]</td><td>14.4% (12.6 to 16.1, p<0.001)</td></tr>
<tr><td>Under-screened, n (%) [95% CI]</td><td>518 (7.3) [6.71 to 7.93]</td><td>300 (4.2)</td><td>818 (11.5) [10.77 to 12.26]</td><td>65 (6.4) [4.98 to 8.08]</td><td>5.1% (3.4 to 6.8, p<0.001)</td></tr>
</table>
A sensitivity analysis was performed for the never-screened population to account for women who were determined ineligible for the trial post-randomisation. This was either women who were found to have had a prior Pap test or women who were found to have had a prior hysterectomy. The results from these analyses determined participation rates of 14.2% for the self-sampling arm and 4.2% for the Pap test arm, with an absolute difference of 10%.

Difference in participation between arms stratified by age, socioeconomic status and time from last Pap test also reported.

<u>Unsatisfactory tests</u>
9 (0.6%) [95% CI 0.29 to 1.10] of the returned samples were found to be unsatisfactory</td></tr>
</table>

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr><td></td><td></td><td>
<table>
<tr><td>40 to 49</td><td>1,342 (18.8)</td><td>176 (17.3)</td><td>2,351 (32.9)</td><td>358 (35.1)</td></tr>
<tr><td>50 to 59</td><td>1,453 (20.4)</td><td>198 (19.4)</td><td>1,453 (20.4)</td><td>207 (20.3)</td></tr>
<tr><td>60 to 69</td><td>2,395 (33.5)</td><td>370 (36.3)</td><td>1,002 (14.0)</td><td>132 (12.9)</td></tr>
</table>
Baseline characteristics of socioeconomic status and area remoteness also reported
</td><td></td><td></td></tr>
<tr><td>Tamalet 2016[44]</td><td><u>Design</u>
Prospective cohort

<u>Objective</u>
To describe high risk-HPV types in 35 to 69-year-old women from low socioeconomic groups not attending regular cytological screening in Marseille, France

<u>Dates</u>
2011 to 2012

<u>Country</u>
France

<u>Setting</u>
Home testing</td><td><u>Patient recruitment</u>
Women aged 35 to 69 years living in the Northern districts of Marseille were identified in the National Insurance Registry as not having had a Pap smear for more than 2 years. Women were informed by mail the importance of regular screening and that they would receive a vaginal self-sampling kit at home in the next month

<u>Data collection</u>
Women who returned a self-sample were considered participants.

<u>Sample size and demographics</u>
27,000 women initially contacted, 22,702 were sent self-sampling HPV tests after elimination of women refusing tests or not living at the mailing address

<u>Baseline characteristics</u>
<table>
<tr><th>Characteristic</th><th>Eligible population</th></tr>
<tr><td colspan="2">Age, years</td></tr>
<tr><td>35 to 39, n</td><td>4,395</td></tr>
<tr><td>40 to 44, n</td><td>4,211</td></tr>
<tr><td>45 to 49, n</td><td>4,044</td></tr>
<tr><td>50 to 54, n</td><td>3,488</td></tr>
<tr><td>55 to 59, n</td><td>2,886</td></tr>
<tr><td>60 to 64, n</td><td>1,600</td></tr>
<tr><td>65 to 69, n</td><td>2,078</td></tr>
</table>
</td><td><u>Self-testing</u>
Women were sent a self-sampling HPV test with instructions and a response envelope. Vaginal cells and secretions were collected using flocked swabs (MAST diagnostics) and subsequently placed in Abbott transport medium. The swab was placed in a tube and then sent in the mail to a Virology laboratory participating in the study.</td><td><u>Screening test uptake</u>
4,245/22,702 (18.7%) [95% CI 18.19 to 19.21] women performed self-sampling

Participation is also reported by age groups

<u>Unsatisfactory tests</u>
9/4,245 (0.21%) [95% CI 0.10 to 0.40] samples were excluded due to low cellularity</td></tr>
</table>

Study	Study design	Population characteristics	Screening methods	Results
Verhoef 2014[45]	Design RCT Objective To investigate if direct DNA methylation-based molecular triage on self-sampled cervicovaginal specimens was non-inferior to cytology triage on additional physician-collected cervical samples in the detection of CIN2+ in women who did not attend cervical screening programmes Dates 1 Nov 2010 to 31 Dec 2011 Country Netherlands Setting Home testing, centres: VU University Medical Centre, Radboud University Nijmegen Medical Centre, the screening organisations Mid-West and East	Patient recruitment Women registered as non-attendees in 2007 in the databases of screening organisations Eligibility criteria: • Aged 33 to 63 years • Living in Noord-Holland Flevoland, Utrecht and Gelderland • No hysterectomy • No history of CIN2+ • No abnormal cytology in the preceding 2 years Data collection Women who returned samples and informed consent forms were considered responders Sample size and demographics 46,001 women invited, 38,913 sent self-sampling devices	Self-testing Non-attendees were sent a letter allowing them to opt out of the trial, and those who did not opt out subsequently received a self-sampling lavage device (Delphi screener), an explanation letter, an informed consent form, an instruction form, a collection tube, a seal bag and a free return envelope. Women were asked to return their self-sampled material, together with a signed consent form to the laboratory for hrHPV testing.	Screening test uptake 12,819/38,913 (*32.9%*) self-sampling devices returned 12,819/46,001 (*27.9%*) [95% CI *27.49 to 28.*31] women invited to take part in study

<table>
<tr><th>Study</th><th>Study design</th><th>Population characteristics</th><th>Screening methods</th><th>Results</th></tr>
<tr>
<td>Virtanen 2015[46]</td>
<td><u>Design</u>
Prospective cohort

<u>Objective</u>
To study the effect of reminder letters (1st reminder) and self-sampling tests (2nd reminder) as means to increase attendance within the routine cervical cancer screening programme

<u>Dates</u>
2011-2012

(11 municipalities took part both years, 11 only in 2011 and 9 only in 2012)

<u>Country</u>
Finland

<u>Setting</u>
Home testing</td>
<td><u>Patient recruitment</u>
Women were identified for screening from the Population Register based on their age and home municipality, and all with address information available are invited to screening by personal letters. Non-attendees received a second invitation (1st reminder) within the same year, however in 2012, women were not sent a reminder letter if they cancelled their given appointment. As a second reminder letter, a self-sampling test was sent out to non-attendees. Prior to mailing the device, the possibility was introduced in an invitation letter with an opt out option.

<u>Data collection</u>
Women were considered attenders by returning a self-taken sample of by coming to the clinic for a Pap smear

<u>Sample size and demographics</u>
31,053 women identified for screening, of whom 30,827 received an initial invitation to screening.

4,536 invited to obtain self-sampling kit, of whom 3,836 received the kit

<u>Characteristics of women invited to self-sampling</u>
<table>
<tr><th>Characteristic</th><th>Women invited to self-sampling, n=4536</th></tr>
<tr><td colspan="2">Age, years</td></tr>
<tr><td>30 to 34, n</td><td>994</td></tr>
<tr><td>35 to 39, n</td><td>753</td></tr>
<tr><td>40 to 44, n</td><td>528</td></tr>
<tr><td>45 to 49, n</td><td>585</td></tr>
<tr><td>50 to 54, n</td><td>535</td></tr>
<tr><td>55 to 59, n</td><td>562</td></tr>
<tr><td>60 to 64, n</td><td>579</td></tr>
</table>
Characteristics of mother tongue, municipality type, education level, marital status, geographical location are also reported</td>
<td><u>Self-testing</u>
The sample taking was done by the Delphi Scanner (lavage device). Samples were sent to the screening laboratory in a test-tube in the regular mail.</td>
<td><u>Screening test uptake</u>
939/4,536 (20.7%) [95% CI 19.53 to 21.91] women took part in screening after invitation to receive a self-sampling kit

920 (20.3%) [95% CI 19.14 to 21.50] returned a self-sampling kit

19 (0.4%) attended a Pap-smear

Increase in total participation rate to 82.2% (95% CI: 81.8 to 82.7) from 79.2% (95% CI 78.8 to 79.7) after 1st reminder

Screening attendance by age-group, mother tongue, municipality type, education level, marital status and geographical location also reported

<u>Unsatisfactory tests</u>
30/920 (3.2%) [95% CI 2.16 to 4.55] of the originally returned samples were not considered adequate.

(Only samples which produced a visible pellet after centrifugation at 1500 rpm were considered adequate)</td>
</tr>
</table>

Study	Study design	Population characteristics	Screening methods	Results
		Characteristics of self-sampling participants Time from previous Pap-smear, years / Self-sampling participants, n=939 <5, n (%) 533 (56.8) 5 to 9, n (%) 157 (16.7) ≥10 years, n (%) 72 (7.7) Never, n (%) 40 (4.3) No information, n (%) 137 (14.6)		

Table 18 Quality assessment of studies relevant to question 5

Question	Bosgraaf 2015[39]	Ducancelle 2015[40]	Duke 2015[41]	Enerly 2016[42]	Sultana 2016[43]	Tamalet 2016[44]	Verhoef 2014[45]	Virtanen 2015[46]	Kitchener 2016[47]	Racey 2016[48]
EXTERNAL VALIDITY										
Have actual probability values been reported (e.g. 0.035 rather than <0.05) for the main outcomes except where the probability value is less than 0.001?	No	Yes	Yes	No	Yes	No	No	No	Yes	Yes
Were the subjects asked to participate in the study representative of the population of interest for this review?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CONFOUNDING										
In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients?	Yes	N/A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Were the statistical tests used to assess the main outcomes appropriate?	Yes	N/A	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes
Were the main outcome measures used accurate (valid and reliable)?	Yes	Yes	Unclear	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes

Question	Bosgraaf 2015[39]	Ducancelle 2015[40]	Duke 2015[41]	Enerly 2016[42]	Sultana 2016[43]	Tamalet 2016[44]	Verhoef 2014[45]	Virtanen 2015[46]	Kitchener 2016[47]	Racey 2016[48]
Were the patients in different intervention groups (trials and cohort studies) recruited from the same population?	Yes	N/A	No	Yes	Yes	N/A	N/A	N/A	Yes	Yes
Were the groups similar at the outset of the study in terms of prognostic factors, for example, severity of disease?	Unclear	N/A	Unclear	Unclear	Yes	N/A	N/A	N/A	Unclear	Yes
Were study subjects in different intervention groups (trials and cohort studies) recruited over the same period of time?	Yes	N/A	Yes	Yes	Yes	N/A	N/A	N/A	Yes	Yes
Were study subjects randomised to intervention groups?	Yes	No	No	Inadequate	Yes	No	No	No	Yes	Yes
Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	Yes	N/A	N/A	Unclear	Yes	N/A	N/A	N/A	Yes	Unclear
Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	Yes	N/A	No	Unclear	Yes	N/A	N/A	N/A	Unclear	Yes
POWER										
Did the study have sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?	No	N/A	Unclear	Unclear	Unclear	N/A	N/A	N/A	Yes	Unclear

Appendix 5 – Explanation of screening test accuracy graphs

[To be completed after finalisation of graph design and development of wording for 'Examples and Explanations' document]

Lines represent values of sensitivity and specificity that give the same LR+ or LR-

Figure 11 Positive likelihood ratios

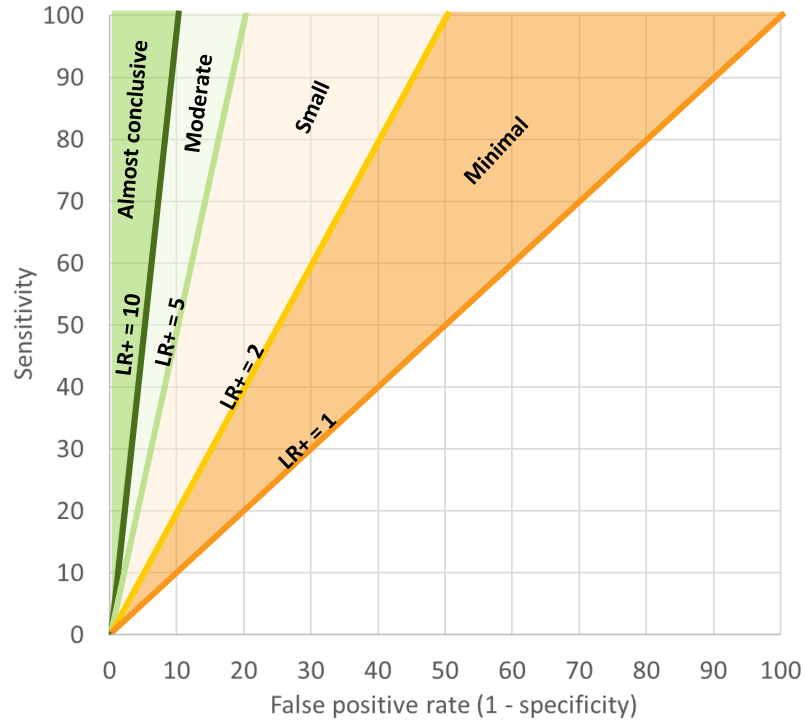
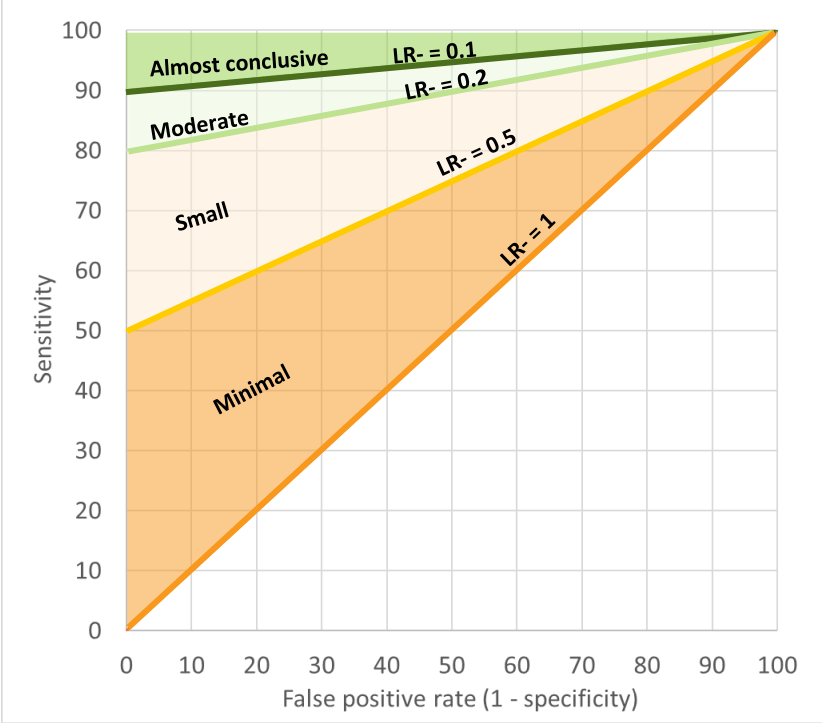


Figure 12 Negative likelihood ratios



5 References

1. Ferlay J, Soerjomataram I, Ervik M, et al. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr. Last accessed 4th March 2017. 2013.
2. Cancer Research UK. Cervical cancer mortality statistics. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/mortality. Last accessed 5th March 2017., 2014.
3. Nayar R, Wilbur DC. The Pap test and bethesda 2014. Acta cytologica 2015;59:121-132.
4. Cancer Research UK. Treatment if you have abnormal cervical cells. Available from: http://www.cancerresearchuk.org/about-cancer/cervical-cancer/treatment-for-abnormal-cervical-cells/treatment. Last accessed 5th March 2017., 2014.
5. Franco EL, Villa LL, Sobrinho JP, et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women from a high-risk area for cervical cancer. Journal of Infectious Diseases 1999;180:1415-1423.
6. Cancer Research UK. Cervical cancer statistics. Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer. Last accessed 5th March 2017., 2014.
7. Walboomers JM, Jacobs MV, Manos MM, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. The Journal of pathology 1999;189:12-19.
8. Chesson HW, Dunne EF, Hariri S, et al. The estimated lifetime probability of acquiring human papillomavirus in the United States. Sexually transmitted diseases 2014;41:660-664.
9. Burchell AN, Winer RL, de Sanjose S, et al. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. Vaccine 2006;24 Suppl 3:S3/52-61.
10. Anderson L, O'Rorke M, Jamison J, et al. Prevalence of human papillomavirus in women attending cervical screening in the UK and Ireland: new data from northern Ireland and a systematic review and meta-analysis. J Med Virol 2013;85:295-308.
11. Hibbitts S, Tristram A, Beer H, et al. UK population based study to predict impact of HPV vaccination. J Clin Virol 2014;59:109-14.
12. Ho GY, Bierman R, Beardsley L, et al. Natural history of cervicovaginal papillomavirus infection in young women. N Engl J Med 1998;338:423-8.
13. Baseman JG, Koutsky LA. The epidemiology of human papillomavirus infections. Journal of clinical virology 2005;32:16-24.
14. Richardson H, Kelsall G, Tellier P, et al. The natural history of type-specific human papillomavirus infections in female university students. Cancer Epidemiology and Prevention Biomarkers 2003;12:485-490.
15. Public Health England. Cervical screening: programme overview, https://www.gov.uk/guidance/cervical-screening-programme-overview. Last accessed: 4th March 2017., 2015.
16. Kitchener HC. Report to the National Screening Committee. Available from: https://legacyscreening.phe.org.uk/policydb_download.php?doc=555. Last accessed 4th March 2017. 2015.
17. UK National Screening Committee. UK NSC HPV recommendation. Available from: https://legacyscreening.phe.org.uk/policydb_download.php?doc=594. Last accessed 4th March 2017. 2016.

18. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine 2012;30 Suppl 5:F88-99.
19. Health and Social Care Information Centre. Cervical Screening Programme, England. Statistics for 2014-15. Available from: http://content.digital.nhs.uk/catalogue/PUB18932/nhs-cervical-stat-eng-2014-15-rep.pdf. Last accessed 14th March 2017. 2015.
20. Dillner J, Rebolj M, Birembaut P, et al. Long term predictive values of cytology and human papillomavirus testing in cervical cancer screening: Joint European cohort study. Bmj 2008;337:969-972.
21. Kitchener HC, Almonte M, Thomson C, et al. HPV testing in combination with liquid-based cytology in primary cervical screening (ARTISTIC): a randomised controlled trial. The lancet oncology 2009;10:672-682.
22. Cook DA, Mei W, Smith LW, et al. Comparison of the Roche cobas® 4800 and Digene Hybrid Capture® 2 HPV tests for primary cervical cancer screening in the HPV FOCAL trial. BMC cancer 2015;15:968.
23. Ronco G, Giorgi-Rossi P, Carozzi F, et al. Human papillomavirus testing and liquid-based cytology in primary screening of women younger than 35 years: results at recruitment for a randomised controlled trial. The lancet oncology 2006;7:547-555.
24. Anttila A, Kotaniemi-Talonen L, Leinonen M, et al. Rate of cervical cancer, severe intraepithelial neoplasia, and adenocarcinoma in situ in primary HPV DNA screening with cytology triage: randomised study within organised screening programme. BMJ (Clinical research ed.). Volume 340, 2010:c1804.
25. Malila N, Leinonen M, Kotaniemi-Talonen L, et al. The HPV test has similar sensitivity but more overdiagnosis than the Pap test - A randomised health services study on cervical cancer screening in Finland. International Journal of Cancer 2013;132:2141-2147.
26. Leinonen MK, Nieminen P, Lonnberg S, et al. Detection rates of precancerous and cancerous cervical lesions within one screening round of primary human papillomavirus DNA testing: Prospective randomised trial in Finland. BMJ (Online) 2012;345 (7886) (no pagination).
27. Veijalainen O, Kares S, Kujala P, et al. Human papillomavirus test with cytology triage in organized screening for cervical cancer. Acta Obstetricia et Gynecologica Scandinavica 2016;95:1220-1227.
28. Del Mistro A, Frayle H, Ferro A, et al. Cervical cancer screening by high risk HPV testing in routine practice: results at one year recall of high risk HPV-positive and cytology-negative women. Journal of medical screening 2014;21:30-37.
29. Zorzi M, Del Mistro A, Farruggio A, et al. Use of a high-risk human papillomavirus DNA test as the primary test in a cervical cancer screening programme: A population-based cohort study. BJOG: An International Journal of Obstetrics and Gynaecology 2013;120:1260-1268.
30. Zorzi M, Frayle H, Rizzi M, et al. A 3-year interval is too short for re-screening HPV negative women: a population-based cohort study. Bjog 2017.
31. Gustinucci D, Rossi PG, Cesarini E, et al. Use of Cytology, E6/E7 mRNA, and p16INK4a–Ki-67 to Define the Management of Human Papillomavirus (HPV)–Positive Women in Cervical Cancer Screening. American journal of clinical pathology 2016;145:35-45.
32. 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.

33. Naucler P, Ryd W, Törnberg S, et al. Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening. Journal of the National Cancer Institute 2009;101:88-99.
34. Rijkaart D, Berkhof J, Van Kemenade F, et al. HPV DNA testing in population-based cervical screening (VUSA-Screen study): results and implications. British journal of cancer 2012;106:975-981.
35. Arbyn M, Verdoodt F, Snijders PJ, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. The lancet oncology 2014;15:172-183.
36. Stanczuk G, Baxter G, Currie H, et al. Clinical validation of hrHPV testing on vaginal and urine self-samples in primary cervical screening (cross-sectional results from the Papillomavirus Dumfries and Galloway-PaVDaG study). BMJ Open 2016;6:e010660.
37. 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.
38. Verdoodt F, Jentschke M, Hillemanns P, et al. 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. European journal of cancer 2015;51:2375-2385.
39. Bosgraaf RP, Verhoef VMJ, Massuger LFAG, et al. Comparative performance of novel self-sampling methods in detecting high-risk human papillomavirus in 30,130 women not attending cervical screening. International Journal of Cancer 2015;136:646-655.
40. Ducancelle A, Reiser J, Pivert A, et al. Home-based urinary HPV DNA testing in women who do not attend cervical cancer screening clinics. Journal of Infection 2015;71:377-384.
41. Duke P, Godwin M, Ratnam S, et al. Effect of vaginal self-sampling on cervical cancer screening rates: A community-based study in Newfoundland. BMC Women's Health 2015;15 (1) (no pagination).
42. Enerly E, Bonde J, Schee K, et al. Self-Sampling for Human Papillomavirus Testing among Non-Attenders Increases Attendance to the Norwegian Cervical Cancer Screening Programme. PLoS ONE [Electronic Resource] 2016;11:e0151978.
43. Sultana F, English DR, Simpson JA, et al. Home-based HPV self-sampling improves participation by never-screened and under-screened women: Results from a large randomized trial (iPap) in Australia. International Journal of Cancer 2016;139:281-290.
44. Tamalet C, Halfon P, Retraite LL, et al. Genotyping and follow-up of HR-HPV types detected by self-sampling in women from low socioeconomic groups not participating in regular cervical cancer screening in France. Journal of Clinical Virology 2016;78:102-107.
45. Verhoef VMJ, Bosgraaf RP, Van Kemenade FJ, et al. Triage by methylation-marker testing versus cytology in women who test HPV-positive on self-collected cervicovaginal specimens (PROHTECT-3): A randomised controlled non-inferiority trial. The Lancet Oncology 2014;15:315-322.
46. Virtanen A, Anttila A, Luostarinen T, et al. Improving cervical cancer screening attendance in Finland. International Journal of Cancer 2015;136:E677-E684.
47. Kitchener HC, Gittins M, Rivero-Arias O, et al. A cluster randomised trial of strategies to increase cervical screening uptake at first invitation (STRATEGIC). Health Technology Assessment (Winchester, England) 2016;20:1-138.
48. Racey CS, Gesink DC, Burchell AN, et al. Randomized Intervention of Self-Collected Sampling for Human Papillomavirus Testing in Under-Screened Rural Women: Uptake of Screening and Acceptability. Journal of Women's Health 2016;25:489-497.

49. Berkhof J, Coupe VM, Bogaards JA, et al. The health and economic effects of HPV DNA screening in The Netherlands. Int J Cancer 2010;127:2147-58.
50. Naber SK, Matthijsse SM, Rozemeijer K, et al. Cervical Cancer Screening in Partly HPV Vaccinated Cohorts - A Cost-Effectiveness Analysis. PLoS One 2016;11:e0145548.
51. Ronco G, Biggeri A, Confortini M, et al. [Health technology assessment report: HPV DNA based primary screening for cervical cancer precursors]. Epidemiol Prev 2012;36:e1-72.
52. Rozemeijer K, de Kok IM, Naber SK, et al. Offering Self-Sampling to Non-Attendees of Organized Primary HPV Screening: When Do Harms Outweigh the Benefits? Cancer Epidemiol Biomarkers Prev 2015;24:773-82.
53. Public Health England. UK NSC evidence review process. Guidance. Appendix F: Requirements for UK NSC evidence summaries. Available from: https://www.gov.uk/government/publications/uk-nsc-evidence-review-process/appendix-f-requirements-for-uk-nsc-evidence-summaries. Last accessed 4th March 2017. 2016.
54. Shea BJ, Grimshaw JM, Wells GA, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007;7:10.
55. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.
56. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. Journal of Epidemiology and Community Health 1998;52:377-384.
57. Cook DA, Mei W, Smith LW, et al. Comparison of the Roche cobas 4800 and Digene Hybrid Capture 2 HPV tests for primary cervical cancer screening in the HPV FOCAL trial. BMC Cancer 2015;15 (1) (no pagination).
58. Gustinucci D, Rossi PG, Cesarini E, et al. Use of Cytology, E6/E7 mRNA, and p16INK4a-Ki-67 to Define the management of human papillomavirus (HPV)-positive women in cervical cancer screening. American Journal of Clinical Pathology 2016;145:35-45.
59. Veijalainen O, Kares S, Kujala P, et al. HPV test with cytology triage in organized screening for cervical cancer. Acta Obstetricia et Gynecologica Scandinavica 2016;3:3.
60. Zorzi M, Del Mistro A, Farruggio A, et al. Use of a high-risk human papillomavirus DNA test as the primary test in a cervical cancer screening programme: a population-based cohort study. BJOG: An International Journal of Obstetrics & Gynaecology 2013;120:1260-1268.