

CADTH RAPID RESPONSE REPORT:  
SUMMARY WITH CRITICAL APPRAISAL

# HPV Self-Sampling for Primary Cervical Cancer Screening: A Review of Diagnostic Test Accuracy and Clinical Evidence – An Update

Service Line: Rapid Response Service  
Version: 1.0  
Publication Date: May 30, 2019  
Report Length: 42 Pages

**Authors:** Yi-Sheng Chao, Suzanne McCormack

**Cite As:** HPV Self-Sampling for Primary Cervical Cancer Screening: A Review of Diagnostic Test Accuracy and Clinical Evidence – An Update. Ottawa: CADTH; 2019 May. (CADTH rapid response report: summary with critical appraisal).

**ISSN:** 1922-8147 (online)

**Disclaimer:** The information in this document is intended to help Canadian health care decision-makers, health care professionals, health systems leaders, and policy-makers make well-informed decisions and thereby improve the quality of health care services. While patients and others may access this document, the document is made available for informational purposes only and no representations or warranties are made with respect to its fitness for any particular purpose. The information in this document should not be used as a substitute for professional medical advice or as a substitute for the application of clinical judgment in respect of the care of a particular patient or other professional judgment in any decision-making process. The Canadian Agency for Drugs and Technologies in Health (CADTH) does not endorse any information, drugs, therapies, treatments, products, processes, or services.

While care has been taken to ensure that the information prepared by CADTH in this document is accurate, complete, and up-to-date as at the applicable date the material was first published by CADTH, CADTH does not make any guarantees to that effect. CADTH does not guarantee and is not responsible for the quality, currency, propriety, accuracy, or reasonableness of any statements, information, or conclusions contained in any third-party materials used in preparing this document. The views and opinions of third parties published in this document do not necessarily state or reflect those of CADTH.

CADTH is not responsible for any errors, omissions, injury, loss, or damage arising from or relating to the use (or misuse) of any information, statements, or conclusions contained in or implied by the contents of this document or any of the source materials.

This document may contain links to third-party websites. CADTH does not have control over the content of such sites. Use of third-party sites is governed by the third-party website owners' own terms and conditions set out for such sites. CADTH does not make any guarantee with respect to any information contained on such third-party sites and CADTH is not responsible for any injury, loss, or damage suffered as a result of using such third-party sites. CADTH has no responsibility for the collection, use, and disclosure of personal information by third-party sites.

Subject to the aforementioned limitations, the views expressed herein are those of CADTH and do not necessarily represent the views of Canada's federal, provincial, or territorial governments or any third party supplier of information.

This document is prepared and intended for use in the context of the Canadian health care system. The use of this document outside of Canada is done so at the user's own risk.

This disclaimer and any questions or matters of any nature arising from or relating to the content or use (or misuse) of this document will be governed by and interpreted in accordance with the laws of the Province of Ontario and the laws of Canada applicable therein, and all proceedings shall be subject to the exclusive jurisdiction of the courts of the Province of Ontario, Canada.

The copyright and other intellectual property rights in this document are owned by CADTH and its licensors. These rights are protected by the Canadian *Copyright Act* and other national and international laws and agreements. Users are permitted to make copies of this document for non-commercial purposes only, provided it is not modified when reproduced and appropriate credit is given to CADTH and its licensors.

**About CADTH:** CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

**Funding:** CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

## Abbreviations

CADTH	Canadian Agency for Drugs and Technologies in Health
HPV	human papillomavirus
PCR	polymerase chain reaction
RCT	randomized controlled trial
SR	systematic review

## Context and Policy Issues

The introduction of cervical cancer screening and timely intervention is associated with the recent decrease in cervical cancer incidence.<sup>1</sup> There are several options to screen cervical cancer. Two of the methods commonly used in Canada are cytology and human papillomavirus (HPV) tests.<sup>2</sup> Cytology requires clinicians to obtain samples from the cervix for further examination.<sup>2</sup> HPV tests that detect the infection of HPV also requires samples from the cervix.<sup>2</sup> The HPV tests that detect certain types of carcinogenic HPV genotypes, especially genotypes 16 and 18, are called high-risk HPV tests.<sup>3</sup> The samples can be obtained via brushes or swabs or other devices not only by clinicians, but also by screening participants.<sup>3</sup> Clinician-sampled HPV tests are used in screening program in several countries, such as Italy<sup>4</sup> and Denmark.<sup>5</sup> Self-sampled HPV tests have been tested in the capital region in Denmark but have not replaced clinician-sampled tests.<sup>5</sup>

With feasibility to conduct at home and potentially better acceptability to participants, self-sampled HPV tests have been used to reach individuals that are unscreened or under-screened for cervical cancer.<sup>4</sup> In a previous CADTH report, there was some evidence to show similar diagnostic test accuracy between self- and clinician-sampled HPV tests.<sup>6</sup> For example, the diagnostic test accuracy of GP5+/6+ polymerase chain reaction (PCR) HPV tests using samples taken with brushes is similar for self- and clinician-collected samples.<sup>7</sup> In several primary studies, fair to high agreement between self- and clinician-sampled HPV tests has been found.<sup>6</sup>

Since the previous CADTH review, there have been primary studies comparing self- and clinician-sampled HPV tests published<sup>8,9</sup> and a systematic review has been updated.<sup>3</sup> This report updates the previous review on the difference in the diagnostic test accuracy of self-sampled HPV tests and the agreement between self- and clinician-sampled HPV tests.

## Research Questions

1. What is the diagnostic test accuracy of self-sampled HPV tests compared with clinician-sampled HPV tests or cytology for asymptomatic cervical cancer screening?
2. What is the clinical evidence regarding the agreement or concordance of self-sampled HPV tests and clinician-sampled HPV tests or cytology for asymptomatic cervical cancer screening?

## Key Findings

Similar to the meta-analysis reviewed in in the previous CADTH report, Arbyn et al. confirmed the difference in diagnostic test accuracy between self- and clinician-collected samples remained significant for signal amplification-based HPV tests. Since the publication of the CADTH report, there were new studies that have examined self- and clinician-sampled HPV tests.

Two systematic reviews, two randomized controlled trials (RCTs), and ten non-randomized studies were identified. In the updated meta-analysis by Arbyn et al., human papillomavirus (HPV) tests were categorized into polymerase chain reaction (PCR) and signal amplification-based tests. Self-sampled HPV tests based on PCR for the detection of cervical intraepithelial neoplasia (CIN) grade 2 or more severe were shown to not have statistically different sensitivity or specificity compared with clinician-sampled tests. However, self-sampled HPV tests based on signal amplification were not as accurate for the detection of CIN2+.

Moderate to excellent agreement between self- and clinician-sampled HPV tests was reported in primary studies. Various HPV tests were tested in different healthcare settings.

However, it was unclear whether the differences in the agreement were associated with the types of HPV tests. There was heterogeneity between the included studies and the impact on the diagnostic accuracy or agreement of self- and clinician-sampled HPV tests was unclear.

Further research on the diagnostic test accuracy and agreement between self- and clinician-sampled HPV tests in target populations could reduce the uncertainties in the application of self-sampled HPV tests.

## Methods

### Literature Search Methods

This report makes use of a literature search developed for a previous CADTH report of April 2018.<sup>6</sup>

For the current report, a limited literature search was conducted by an information specialist on key resources including Ovid Medline, Embase, the Cochrane Library, University of York Centre for Reviews and Dissemination (CRD) databases, Canadian and major international health technology agencies, as well as a focused Internet search. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were self-sampling and the human papilloma virus (HPV). No filters were applied to limit retrieval by study type. The search was also limited to English language documents published between January 1, 2017 and April 26, 2019.

### Selection Criteria and Methods

One reviewer screened citations and selected studies. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

**Table 1: Selection Criteria**

<b>Population</b>	Asymptomatic adults eligible for cervical cancer screening (≥ 21 years of age, or age at which screening starts in the jurisdiction)
<b>Intervention</b>	Q1-2: Self-sampled high-risk HPV tests for primary cervical cancer screening
<b>Comparator</b>	Q1-2: Clinician-sampled high-risk HPV tests for primary cervical cancer screening; cytology (conventional Pap smear or liquid based cytology) Q1 only: Colposcopy with histologic examination of tissue specimens, when indicated
<b>Outcomes</b>	Q1: Diagnostic test accuracy <ul style="list-style-type: none"> <li>• Number and proportion of patients positive and negative on each test using colposcopy as reference standard</li> <li>• Sensitivity, specificity, PPV, NPV, PLR, NLR, DOR to screen for high-grade cervical lesions (HSIL or CIN2+, AGC, AIS) and/or invasive cervical cancer (squamous cell carcinoma or adenocarcinoma)</li> </ul> Q2: Agreement between self-sampled HPV tests and clinician-sampled HPV tests or cytology (i.e., % agreement of positive test results, % agreement of negative test results)
<b>Study Designs</b>	Health technology assessments; systematic reviews; meta-analyses; randomized controlled trials; non-randomized studies

AGC = atypical glandular cell; AIS = adenocarcinoma in situ; CIN = cervical intraepithelial neoplasia; DOR = diagnostic odds ratio; HPV = human papillomavirus; HSIL = high-grade squamous intraepithelial lesion; NLR = negative likelihood ratio; NPV = negative predictive value; PLR = positive likelihood ratio; PPV = positive predictive value

### Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1, they were duplicate publications, or were published prior to 2017. If patients with known HPV status or cytology results were recruited, primary studies that did not specify the screening programs were not eligible for this report. Guidelines with unclear methodology were also excluded. Studies included in selected systematic reviews were also excluded.

### Critical Appraisal of Individual Studies

The included systematic reviews were critically appraised by one reviewer using the AMSTAR II checklist,<sup>10</sup> randomized and non-randomized studies were critically appraised using the Downs and Black checklist.<sup>11</sup> Diagnostic test accuracy studies were also assessed with the QUADAS-2 checklist.<sup>12</sup> Summary scores were not calculated for the included studies; rather, a review of the strengths and limitations of each included study were described narratively.

### Summary of Evidence

#### Quantity of Research Available

A total of 233 citations were identified in the literature search. Following screening of titles and abstracts, 210 citations were excluded and 23 potentially relevant reports from the electronic search were retrieved for full-text review. One potentially relevant publication was retrieved from the grey literature search for full text review. Of these potentially relevant articles, ten publications were excluded for various reasons, and 14 publications met the inclusion criteria and were included in this report. These comprised two systematic reviews,

two RCTs, and ten non-randomized studies. Appendix 1 presents the PRISMA<sup>13</sup> flowchart of the study selection.

Additional references of potential interest are provided in Appendix 6.

## Summary of Study Characteristics

### *Study Design*

In the updated systematic review (SR) by Arbyn et al. in 2018, articles published up to April 2018 were searched and diagnostic studies that conducted HPV tests on self- and clinician-collected samples from the same individuals were included.<sup>3</sup> For a study to be included, the presence or absence of cervical intraepithelial neoplasia (CIN) grade 2 or more severe needed to be verified with colposcopy.<sup>3</sup> In the 2017 SR by Kelly et al., articles published between January 2004 and February 2017 were searched and studies that evaluated the performance of point-of-care HPV tests for the detection of CIN2+ or CIN3+ were eligible.<sup>14</sup> The review included seven studies (study design not specified) using careHPV for the detection of CIN2+ and four studies using careHPV for the detection of CIN3+.<sup>14</sup> The overlap in the included studies was shown in Appendix 5. There were 76 accuracy studies included in Arbyn et al.<sup>3</sup> and eight included in Kelly et al.<sup>14</sup> Three studies were included in both SRs.<sup>14</sup>

Polman et al. conducted one non-inferiority RCT to compare the sensitivity of self- and clinician-collected samples used for HPV tests to detect CIN2+ or CIN3+.<sup>15</sup> Ajenifuja et al. randomized participants into two groups: provider sampling before self-sampling and self-sampling before provider sampling to compare the agreement of HPV DNA tests with self- or provider-collected samples.<sup>16</sup>

Thay et al., Des Marais et al., Lam et al., Senomago et al., Wong et al., Zhang et al., Cremer et al., and Obiri-Yeboah et al. conducted cohort studies.<sup>5,8,9,17-21</sup> Zhang et al. follow-up the participants for 15 years.<sup>9</sup> The follow-up lengths in the remaining non-randomized studies ranged from 6 to 20 months, if reported.<sup>5,15,20</sup>

Toliman et al. and Phoolcharoen et al. reported cross-sectional findings in eligible populations.<sup>22,23</sup>

### *Country of Origin*

The first authors of the SRs were based in Belgium<sup>3</sup> and the UK.<sup>14</sup>

The first authors of the RCTs were based in the Netherlands<sup>15</sup> and Nigeria.<sup>16</sup> The first authors of the non-randomized studies were based in Cambodia (1 study),<sup>17</sup> Papua New Guinea and Australia (1 study),<sup>22</sup> the US (2 studies),<sup>18,19</sup> Denmark (1 study),<sup>5</sup> Thailand (1 study),<sup>23</sup> China (2 studies),<sup>8,9</sup> the US and Salvador (1 study),<sup>20</sup> and Ghana (1 study).<sup>21</sup>

### *Patient Population*

In the SR by Arbyn et al., individuals participating in diagnostic studies that used self- and clinician-collected samples in HPV tests for the detection of CIN2+ were eligible.<sup>3</sup> In the SR by Kelly et al., individuals recruited in cross-sectional or cohort studies that evaluated point-of-care HPV tests for the detection of CIN2+ or CIN3+ were included.<sup>14</sup>

In the RCT by Polman et al., individuals aged 29 to 61 years in a regular cervical cancer screening program were randomized.<sup>15</sup> In the RCT by Ajenifuja et al., individuals presenting for cervical cancer screening were recruited.<sup>16</sup>

In the non-randomized studies, populations of different characteristics were recruited. Thay et al. enrolled human immunodeficiency virus (HIV) positive or negative participants in a cohort study or from a hospital in Cambodia.<sup>17</sup> Toliman et al. recruited those attending cervical cancer screening at clinics.<sup>22</sup> Because self-sampled HPV tests had the potential to reach infrequently screened individuals, Des Marais et al. focused on low-income and infrequently screened individuals according to national guidelines in the US (no screening in the past four years).<sup>18</sup> Lam et al. recruited non-attenders (not screened for at least four or six years depending on the ages) in the Copenhagen Self-sampling Initiative (CSI) and used participants in the Horizon study (cytology and clinician-sampled HPV tests) as comparison.<sup>5</sup> Phoolcharoen et al. obtained samples from a colposcopy clinic.<sup>23</sup> Senkomago et al. and Wong et al. recruited female sex workers.<sup>8,19</sup> Zhang et al. followed up participants without a history of cervical cancer or hysterectomy aged 35 to 45 years beginning in 1999 for 15 years.<sup>9</sup> Cremer et al. analyzed participants aged 30 to 49 years in a cervical cancer screening program.<sup>20</sup> Obiri-Yeboah et al. recruited those attending HIV and outpatient clinics in a cervical cancer screening study.<sup>21</sup>

### *Interventions and Comparators*

In the SRs by Arbyn et al., the HPV tests were categorized into signal amplification and PCR-based tests and used for self- and clinician-collected samples.<sup>3</sup> Kelly et al. only included studies that evaluated the diagnostic test accuracy of two HPV tests: careHPV and OncoE6 using self- or clinician-collected samples.<sup>14</sup>

In the RCTs by Polman et al. and Ajenifuja et al., HPV tests on self- and clinician-collected samples were compared (GP5+/6+ PCR enzyme immunoassay and Hybridio GenoArray, respectively).<sup>15,16</sup>

In the non-randomized studies by Thay et al., Toliman et al., Des Marais et al., Poolcharoen et al., Senkomago et al., Wong et al., Zhang et al., Cremer et al., and Obiri-Yeboah et al. HPV tests on self- and clinician-collected samples were compared.<sup>8,9,17,19-23</sup>

Lam et al. compared self-sampled HPV tests and routine screening that included cytology or co-testing (HPV tests and cytology).<sup>5</sup>

In addition, Thay et al. also studied the effectiveness of visualization with acetic acid and digital colposcopy.<sup>17</sup> Zhang et al. also tested visual inspection with acetic acid.<sup>9</sup>

### *Outcomes*

The diagnostic test accuracy in this report was based on the detection of colposcopy-confirmed cases (CIN2+ or CIN3+). Sensitivity was the number of identified cases (positive on both HPV and colposcopy) divided by the total number of colposcopy confirmed cases.<sup>6</sup> Specificity was the number of non-cases (negative on both HPV and colposcopy) divided by the total number of colposcopy negative cases.<sup>6</sup> Positive predictive values were the number of identified cases divided by the total number of individuals with positive HPV test results.<sup>6</sup> Negative predictive values were the number of identified non-cases divided by the total number of participants with negative HPV test results.<sup>6</sup>

In the SR by Arbyn et al., the outcomes in the study selection criteria were diagnostic test accuracy and response rates to screening (i.e. proportions of the invited population who participated in screening).<sup>3</sup> Kelly et al. reported diagnostic test accuracy including sensitivity and specificity.<sup>14</sup>

In the RCT by Polman et al., the outcome was the detection of CIN2+ or CIN3+ and diagnostic test accuracy (sensitivity and specificity) was derived.<sup>15</sup> Ajenifuja et al. reported the degree of agreement between self- and provider-sampled HPV tests.<sup>16</sup>

In the non-randomized studies, Thay et al. estimated HPV prevalence and the proportions detected with self- and clinician-collected samples.<sup>17</sup> Toliman et al., Des Marais et al., Phoolcharoen et al., Senkomago et al., Wong et al., Cremer et al., and Obiri-Yeboah et al. reported the agreement of HPV detection with self- and clinician-collected samples.<sup>8,18-23</sup> Lam et al. reported the positive predictive values for the detection of CIN2+.<sup>5</sup> Zhang et al. reported the cumulative diagnostic test accuracy for the detection of CIN2+ at baseline, 6-, 11-, and 15-year follow-up.<sup>9</sup>

Additional details regarding the characteristics of included publications are provided in Appendix 2.

## Summary of Critical Appraisal

### *Systematic reviews*

In the SRs by Arbyn et al. and Kelly et al., the population, intervention, comparator, and outcome components were described.<sup>3,14</sup> The selection of study design was explained.<sup>3,14</sup> Comprehensive literature search strategies, details in the included studies, critical appraisal with published tools, and review authors' competing interests were described.<sup>3,14</sup> The risk of bias in the included studies were considered when interpreting the results and the heterogeneity across the included studies were explained.<sup>3,14</sup> There were no lists of excluded studies for either SRs.<sup>3,14</sup> Arbyn et al. and Kelly et al. conducted meta-analyses and adopted appropriate statistical methods (bivariate models in both SRs).<sup>3,14</sup> Arbyn et al. adjusted for inter-study heterogeneity by conducting sensitivity analysis.<sup>3</sup> However, Kelly et al. did not conduct sensitivity analysis due to an insufficient number of primary studies.<sup>14</sup> In both studies, the potential impact of the risk of bias in the include studies was assessed.<sup>3,14</sup> Arbyn et al. did not find evidence of publication bias.<sup>3</sup> Kelly et al. did not examine publication bias due to a small number of included studies.<sup>14</sup>

Arbyn et al. updated a meta-analysis that was previously published.<sup>3</sup> Kelly et al. did not publish the protocol *a priori*.<sup>14</sup>

Arbyn et al. selected studies and extracted data in duplicate<sup>3</sup> and Kelly et al. did not.<sup>14</sup>

### *RCTs*

In the RCTs by Polman et al. and Ajenifuja et al., the hypotheses, main outcomes, patient characteristics, interventions, distributions of principal confounders, main findings, and the random variability for the main outcomes were described.<sup>15,16</sup> The participants and outcomes assessors were not blinded.<sup>15,16</sup> The lengths of follow-up were similar between groups.<sup>15,16</sup> The statistical tests to assess the outcomes were appropriate.<sup>15,16</sup> The compliance (attending screening) and the outcome measures (HPV tests) were reliable.<sup>15,16</sup> Participants of different groups were recruited from the same populations within the same periods of time.<sup>15,16</sup> Participants were randomized to different groups.<sup>15,16</sup> Allocation to interventions was not concealed.<sup>15,16</sup> The lengths of follow-up were adequate.<sup>15,16</sup> Confounders were adjusted in the analysis.<sup>15,16</sup>

Polman et al. considered important adverse effects and described patients lost to follow-up.<sup>15</sup> It was not reported whether there was differential attrition between groups.<sup>15</sup>



Participants lost to follow-up were not considered in the analysis.<sup>15</sup> Patients lost to follow-up were not reported in the RCT by Ajenifuja et al.<sup>16</sup>

The population asked to participate in the RCT by Polman et al. was representative of the population of interest.<sup>15</sup> The representativeness of the sample in Ajenifuja et al. was not assessed.<sup>16</sup>

Polman et al. reported the 95% confidence intervals.<sup>15</sup> Ajenifuja et al. did not report the actual probability values (*P* values).<sup>16</sup>

The sample sizes were not estimated (power analysis) before the studies began.<sup>15,16</sup>

In addition, the RCT by Polman et al. was assessed with the QUADAS-2 tool.<sup>12</sup> Consecutive samples were enrolled and a case-control design was not adopted.<sup>15</sup> Inappropriate exclusion criteria that were associated with disease prevalence or other confounding factors were not used.<sup>15</sup> The HPV tests were interpreted without the knowledge of colposcopy results.<sup>15</sup> The diagnosis thresholds for the HPV tests were provided.<sup>15</sup> The reference standard, colposcopy, was likely to correctly classify the target condition.<sup>15</sup> There were appropriate intervals between index tests and the reference standard.<sup>15</sup> All patients were eligible for the same reference standard, colposcopy, if positive on the HPV screening test.<sup>15</sup> Participants with negative results in cytology or HPV tests were referred to routine screening.<sup>15</sup> The cases identified in routine screening could be used to calculate diagnostic test accuracy.<sup>15</sup> However, the reference standard results were not interpreted without the results of the index and comparator tests.<sup>15</sup> Participants lost to follow-up were not included in the analysis.<sup>15</sup>

### *Non-randomized studies*

In the nonrandomized studies, the hypotheses, main outcomes, patient characteristics, interventions, distributions of principal confounders, main findings, and the random variability for the main outcomes were described.<sup>5,8,9,17-23</sup> The participants and outcomes assessors were not blinded.<sup>5,8,9,17-23</sup> The lengths of follow-up were similar between groups in each study.<sup>5,8,9,17-23</sup> The statistical tests to assess the outcomes were appropriate.<sup>5,8,9,17-23</sup> The compliance (attending screening) and the outcome measures (HPV test results) were reliable.<sup>5,8,9,17-23</sup> Participants of different groups were recruited from the same populations within the same periods of time.<sup>5,8,9,17-23</sup> Participants were not randomized to different groups.<sup>5,8,9,17-23</sup> Allocation to interventions was not concealed.<sup>5,8,9,17-23</sup> The lengths of follow-up were adequate.<sup>5,8,9,17-23</sup>

The actual probability values (*P* values) were reported,<sup>5,8,9,17-22</sup> except for the non-randomized study by Phoolcharoen et al. that did not report probability values.<sup>23</sup>

Cremer et al., Des Marais et al., Obiri-Yeboah et al., Thay et al., Toliman et al., Wong et al., and Zhang et al. considered confounding in the analysis.<sup>8,9,17,18,20-22</sup>

Obiri-Yeboah et al., Phoolcharoen et al., Thay et al., and Toliman et al. did not report patients lost to follow-up.<sup>17,18,21,23</sup> Cremer et al., Des Marais et al., Lam et al., Senkomago et al., Wong et al., and Zhang et al. described patients lost to follow-up, but whether the distributions were significant was not determined.<sup>5,8,9,18-20</sup>

The population asked to participate in the non-randomized study by Lam et al. was representative of the population of interest.<sup>5</sup> However, the representativeness of the participants in other studies was not assessed.<sup>5,8,9,17-23</sup>

Wong et al. reported important adverse effects,<sup>8</sup> while others did not.<sup>5,9,17-23</sup>

In addition, the non-randomized studies by Lam et al. and Zhang et al. were assessed with the QUADAS-2 checklist.<sup>12</sup> Lam et al. recruited consecutive samples<sup>5</sup> and Zhang et al. enrolled random samples.<sup>9</sup> Both avoided case-control design.<sup>5,9</sup> Inappropriate exclusion criteria that were associated with disease prevalence and other factors were not used.<sup>5,9</sup> The index or comparator tests were not interpreted with the knowledge of the results of the reference standard (colposcopy).<sup>5,9</sup> The thresholds for the index and comparator tests were provided.<sup>5,9</sup> The reference standard, colposcopy, was likely to correctly classify the target condition.<sup>5,9</sup> There were appropriate intervals between the index or comparator tests and the reference standard.<sup>5,9</sup> All patients were eligible for the same reference standard if test positive.<sup>5,9</sup> Those with negative test results were followed-up in routine screening<sup>5</sup> or the subsequent rounds of screening.<sup>9</sup> However, in both studies, the reference standard results were not interpreted without the knowledge of the results of the index or comparator tests.<sup>5,9</sup> Participants lost to follow-up were not included in the analysis.<sup>5,9</sup>

Additional details regarding the strengths and limitations of included publications are provided in Appendix 3.

## Summary of Findings

### *Diagnostic test accuracy of self-sampled high-risk HPV tests*

#### **Systematic Reviews**

In the SR by Arbyn et al., the pooled sensitivity and specificity of high-risk HPV tests based on PCR (GP5+/5+ PCR-EIA, Abbott RT PCR hrHPV, Anyplex II HR, cobas 4800 HPV test, GP5+/6+-LMNX, Linear Array, HPV Risk assay, Xpert HPV) for the detection of CIN2+ using self-samples were not significantly different to those using clinician samples (96% and 79% respectively for both self and clinician samples).<sup>3</sup> There was a significant difference in the sensitivity and specificity of high-risk HPV tests based on signal amplification (Hybrid Capture and Cervista).<sup>3</sup> Self-sample HPV tests based on signal amplification were significantly less sensitive (77% versus 93%) and significantly less specific (84% versus 86%) than clinician-sampled HPV tests.<sup>3</sup>

Kelly et al. reported the sensitivity and specificity of self- and clinician-sampled HPV tests, but did not determine the statistical significance of the difference in the diagnostic test accuracy between these two sampling methods.<sup>14</sup> The pooled sensitivity and specificity of careHPV tests using clinician-collected samples were 88.1% and 83.7% respectively for the detection of CIN2+ and 90.3% and 85.3% respectively for the detection of CIN3+.<sup>14</sup> The pooled sensitivity and specificity of careHPV tests using self-collected samples were 73.6% and 88.0% respectively for the detection of CIN2+ and 75.2% and 90.6% respectively for the detection of CIN3+.<sup>14</sup> Kelly et al. concluded the diagnostic test accuracy was good using careHPV and considered the sensitivity using self-collected samples lower.<sup>14</sup>

#### **RCTs**

In the RCT by Polman et al., the sensitivity and specificity of self-sampled HPV tests were 92.9% and 93.9% respectively for the detection of CIN2+ and 95.1% and 93.4% for the detection of CIN3+ (GP5+/6+ PCR enzyme immunoassay).<sup>15</sup> The sensitivity and specificity of clinician-sampled HPV tests were 96.4% and 94.2% respectively for the detection of CIN2+ and 95.8% and 93.5% for the detection of CIN3+.<sup>15</sup> There was no significant difference in the sensitivity or specificity between self- and clinician-sampled HPV tests.<sup>15</sup>

### Non-randomized studies

Lam et al. reported that the positive predictive values were higher among the CSi attenders (self-sampled HPV tests) (36.5% respectively) than those in the Horizon study (cotesting with cytology and HPV tests, 25.6%).<sup>5</sup> The HPV tests included Hybrid Capture 2, CLART, and Aptima.<sup>5</sup> Lam et al. concluded that self-sampling was associated with higher detection rates than cytology and cotesting (HPV tests and cytology) in a group of screening non-attendees.<sup>5</sup>

Zhang et al. reported the baseline, 6-, 11- and 15-year cumulative sensitivity and specificity for the detection of CIN2+.<sup>9</sup> Zhang et al. concluded that single self-sampled HPV tests were less sensitive than clinician-sampled HPV tests.<sup>9</sup> However, both tests were equal in colposcopy referral rates and the detection rates of CIN2+ on cumulative cases.<sup>9</sup>

### *Agreement of self-sampled high-risk HPV tests and clinician-sampled high-risk HPV tests or cytology*

#### RCTs

Ajenifuja et al. reported statistically significant moderate correlation with a  $\kappa$  value as 0.47 (95% CI, 0.213 to 0.723) between self- and clinician-sampled HPV tests (Hybridio GenoArray).<sup>16</sup>

### Non-randomized studies

Thay et al. defined HPV infections as those detected by both self- and clinician-collected samples.<sup>17</sup> Self-sampled HPV tests identified 89% of HPV infections (50 out of 56) and clinician-sampled HPV test identified 80% (45 out of 56) (careHPV).<sup>17</sup> Thirty-nine HPV infections were identified by both self- and clinician-sampled HPV tests.<sup>17</sup> There was no significant difference in the detection rates between self- and clinician-collected samples.<sup>17</sup> Toliman et al. reported the agreement in high-risk HPV detection between self- and clinician-collected samples was substantial ( $\kappa > 0.6$  in 32 pair-wise comparisons of HPV tests: Xpert, Cobas, and Aptima).<sup>22</sup> Des Marais et al. reported moderate to good agreement between self- and clinician-sampled HPV tests (Aptima,  $\kappa = 0.56$  to  $0.66$ ).<sup>18</sup> Phoolcharoen et al. reported the moderate agreement of HPV tests between self- and clinician-collected samples was 74.5% with a  $\kappa$  value as 0.46 (Cobas).<sup>23</sup> Senkomago et al. reported moderate agreement between self- and clinician-sampled HPV tests that increased over time (Aptima,  $\kappa = 0.55$  and  $0.83$  at baseline and 24 months respectively).<sup>19</sup> Senkomago et al. considered operational proficiency by the participants to obtain samples might be related to the increased agreement over time.<sup>19</sup> Wong et al. identified substantial agreement between self- and clinician-sampled HPV tests (unspecified HPV test,  $\kappa = 0.69$ ).<sup>8</sup> Cremer et al. reported the agreement with a  $\kappa$  value as 0.70 (Hybrid Capture 2).<sup>20</sup> Obiri-Yeboah et al. identified excellent agreement (careHPV,  $\kappa = 0.88$ ).<sup>21</sup>

Phoolcharoen et al. reported the lowest agreement,  $\kappa = 0.46$ ,<sup>23</sup> and Obiri-Yeboah et al. reported the highest agreement,  $\kappa = 0.88$ .<sup>21</sup> There were several differences in health care settings and HPV tests. Phoolcharoen et al., recruited participants from a colposcopy clinic in Thailand.<sup>23</sup> Obiri-Yeboah et al. recruited individuals from a teaching hospital in Ghana.<sup>21</sup> The HPV tests used were Cobas and careHPV.<sup>21,23</sup> Without reference standards in both studies or other tests to understand the diagnostic test accuracy of self- and clinician-sampled HPV tests,<sup>21,23</sup> the exact reasons for the difference in the agreement were unclear.

Appendix 4 presents a table of the main study findings and authors' conclusions.

## Limitations

There were different degrees of heterogeneity in the primary studies included in the SRs.<sup>3,14</sup> The differences in health care settings, HPV tests, screening strategies, and participants' technical proficiency might influence the comparability of the study results of the primary studies and the generalizability to populations that were not considered in the SRs. The population characteristics in the primary studies also varied between studies.<sup>5,8,9,17,21,23</sup> It was unclear whether the differences in HPV prevalence attributed to the variations in the agreement between self- and clinician-sampled HPV tests. The diagnostic test accuracy was not uniform for all HPV tests and there were no sufficient sample sizes to determine the differences between devices. The types of HPV tests were not described in several studies and the impact on the results was not clear.<sup>8,20</sup>

## Conclusions and Implications for Decision or Policy Making

There two SRs,<sup>3,14</sup> two RCTs<sup>15,16</sup> and ten non-randomized studies included.<sup>5,8,9,17-23</sup>

### *Diagnostic test accuracy of self-sampled high-risk HPV tests*

The diagnostic test accuracy of self- and clinician-sampled HPV tests for the detection of CIN2+ were available in one low-quality SR,<sup>3</sup> one critically low-quality SR,<sup>14</sup> one fair-quality RCT,<sup>15</sup> and two fair-quality non-randomized studies.<sup>5,9</sup>

In the updated meta-analysis by Arbyn et al., self-sampled HPV tests based on PCR for the detection of CIN2+ did not have statistically different sensitivity or specificity compared with clinician-sampled tests.<sup>3</sup> However, self-sampled HPV tests based on signal amplification were statistically significantly less sensitive and specific for the detection of CIN2+.<sup>3</sup> Kelly et al. meta-analyzed the diagnostic test accuracy of careHPV (signal amplification-based)<sup>3</sup> and also concluded that careHPV tests using self-collected samples were less sensitive than the same test using clinician-collected samples for the detection of CIN2+ or CIN3+.<sup>14</sup> In the RCT by Polman et al., self- and clinician-sampled PCR- based HPV tests were similarly accurate for the detection of CIN2+ or CIN3+.<sup>15</sup>

### *Agreement of self- and clinician-sampled high-risk HPV tests*

The agreement of self- and clinician-sampled HPV tests was available in one fair-quality RCT,<sup>16</sup> seven fair-quality non-randomized studies,<sup>8,17-22</sup> and one poor-quality non-randomized study.<sup>23</sup>

In the RCT, moderate agreement between self- and clinician-sampled HPV tests was reported using HybriBio GenoArray.<sup>16</sup> In non-randomized studies, moderate to excellent agreement between self- and clinician-sampled HPV tests was found.<sup>8,18,20-22</sup> Similarly Thay et al. used careHPV and did not find significant differences in the HPV detection rates between self- and clinician-collected samples.<sup>17</sup>

In a difference from the earlier meta-analysis<sup>7</sup> reviewed in the previous CADTH report,<sup>6</sup> Arbyn et al. categorized the HPV tests into PCR- and signal amplification-based tests and did not focus on the types of storage medium and sampling devices, such as brushes or swabs.<sup>7</sup> The difference in diagnostic test accuracy between self- and clinician-collected samples remained significant for signal amplification-based HPV tests.<sup>6,7</sup> Since the publication of the CADTH report,<sup>6</sup> there were new studies that continued showing moderate to excellent agreement between self- and clinician-sampled HPV tests.<sup>8,17,18,21-23</sup>

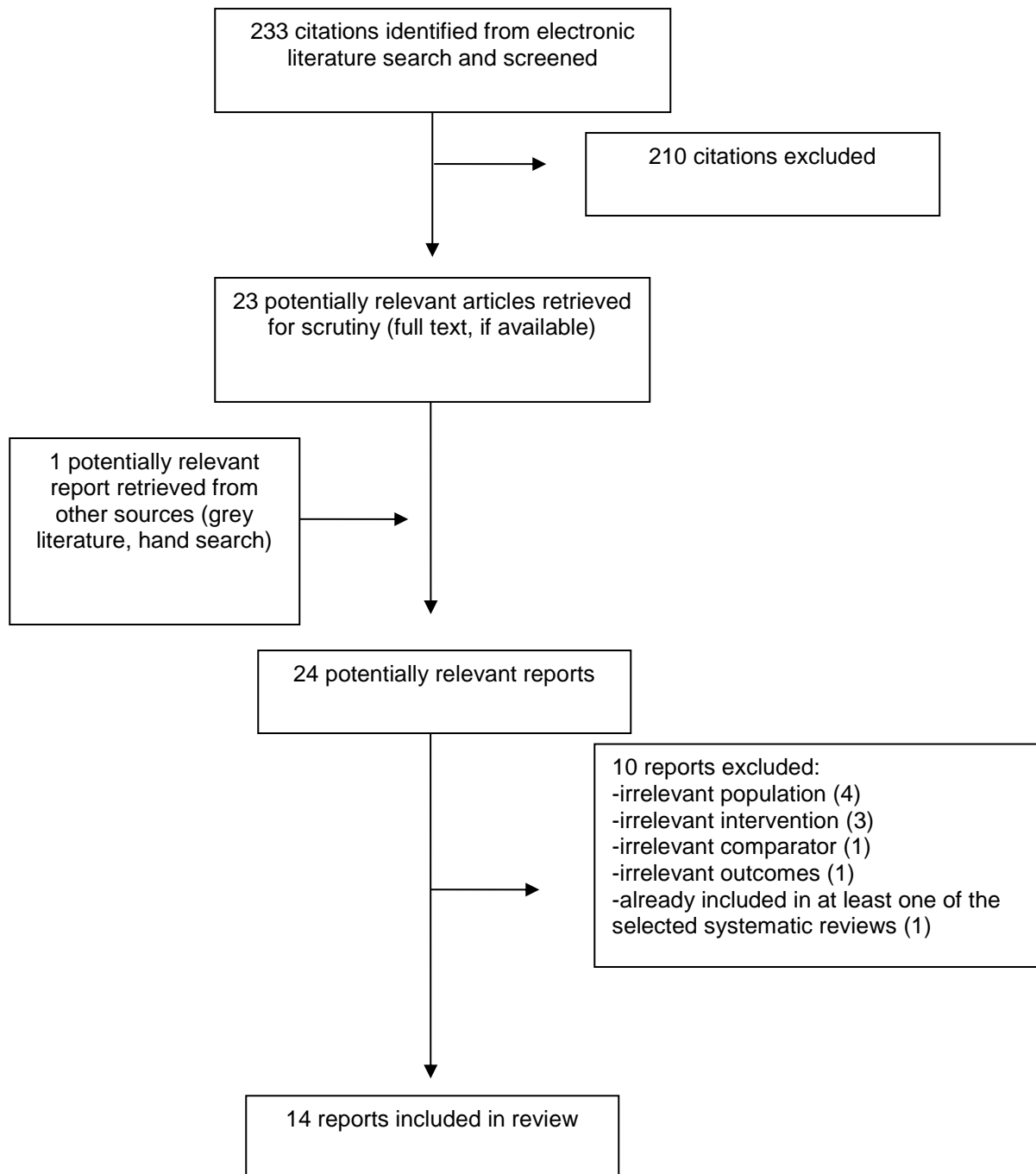
However, it was unclear whether the differences in the agreement were associated with the types of HPV tests being studied. There was heterogeneity between the included studies and the impact on the diagnostic accuracy or agreement of self- and clinician-sampled HPV tests was unclear.

Based on available evidence, self-sampled HPV tests could provide similar accuracy to clinician-sampled tests, particularly for PCR-based HPV tests. Moderate to excellent agreement between self- and clinician-sampled HPV tests was observed in primary studies conducted in various healthcare settings. Further research on the diagnostic test accuracy and agreement between self- and clinician-sampled HPV tests in target populations could reduce the uncertainties in the application of self-sampled HPV tests.

## References

- Canadian cancer statistics 2017 special topic: pancreatic cancer. Toronto (ON): Canadian Cancer Society; 2017: <http://www.cancer.ca/~media/cancer.ca/CW/publications/Canadian%20Cancer%20Statistics/Canadian-Cancer-Statistics-2017-EN.pdf>. Accessed 2019 May 29.
- HPV Testing for Primary Cervical Cancer Screening: A Health Technology Assessment. (CADTH optimal use report vol.7, no.1b). Ottawa (ON): Canadian Agency for Drugs and Technology in Health; 2019: <https://www.cadth.ca/sites/default/files/ou-tr/op0530-hpv-testing-for-pcc-report.pdf>. Accessed 2019 May 29.
- Arbyn M, Smith SB, Temin S, et al. Detecting cervical precancer and reaching underscreened women by using HPV testing on self samples: updated meta-analyses. *BMJ*. 2018;363:k4823.
- Melnikow J, Henderson JT, Burda BU, Senger CA, Durbin S, Weyrich MS. Screening for cervical cancer with high-risk human papillomavirus testing: updated evidence report and systematic review for the US Preventive Services Task Force. *JAMA*. 2018;320(7):687-705.
- Lam JUH, Elfstrom KM, Ejegod DM, et al. High-grade cervical intraepithelial neoplasia in human papillomavirus self-sampling of screening non-attenders. *Br J Cancer*. 2018;118(1):138-144.
- Chao Y-S, Clark M, Ford C. HPV self-sampling for primary cervical cancer screening: a review of diagnostic test accuracy and clinical evidence. (CADTH Rapid response report: summary with critical appraisal). Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2018: <https://www.cadth.ca/hpv-self-sampling-primary-cervical-cancer-screening-review-diagnostic-test-accuracy-and-clinical>. Accessed 2019 May 29.
- Arbyn M, Verdoodt F, Snijders PJF, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. *The Lancet Oncology*. 2014;15(2):172-183.
- Wong ELY, Cheung AWL, Huang F, Chor JSY. Can Human Papillomavirus DNA Self-sampling be an Acceptable and Reliable Option for Cervical Cancer Screening in Female Sex Workers? *Cancer Nurs*. 2018;41(1):45-52.
- Zhang L, Xu XQ, Hu SY, et al. Durability of clinical performance afforded by self-collected HPV testing: A 15-year cohort study in China. *Gynecol Oncol*. 2018;151(2):221-228.
- Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008.
- 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. *J Epidemiol Community Health*. 1998;52(6):377-384.
- 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(8):529-536.
- Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-e34.
- Kelly H, Mayaud P, Segondy M, Pant Pai N, Peeling RW. A systematic review and meta-analysis of studies evaluating the performance of point-of-care tests for human papillomavirus screening. *Sex Transm Infect*. 2017;93(S4):S36-S45.
- Polman NJ, Ebisch RMF, Heideman DAM, et al. Performance of human papillomavirus testing on self-collected versus clinician-collected samples for the detection of cervical intraepithelial neoplasia of grade 2 or worse: a randomised, paired screen-positive, non-inferiority trial. *Lancet Oncol*. 2019;20(2):229-238.
- Ajenifuja OK, Ikeri NZ, Adeteye OV, Banjo AA. Comparison between self sampling and provider collected samples for Human Papillomavirus (HPV) Deoxyribonucleic acid (DNA) testing in a Nigerian facility. *Pan Afr Med J*. 2018;30:110.
- Thay S, Goldstein A, Goldstein LS, Govind V, Lim K, Seang C. Prospective cohort study examining cervical cancer screening methods in HIV-positive and HIV-negative Cambodian Women: a comparison of human papilloma virus testing, visualization with acetic acid and digital colposcopy. *BMJ Open*. 2019;9(2):e026887.
- Des Marais AC, Zhao Y, Hobbs MM, et al. Home Self-Collection by Mail to Test for Human Papillomavirus and Sexually Transmitted Infections. *Obstet Gynecol*. 2018;132(6):1412-1420.
- Senkomago V, Ting J, Kwatampora J, et al. High-risk HPV-RNA screening of physician- and self-collected specimens for detection of cervical lesions among female sex workers in Nairobi, Kenya. *Int J Gynaecol Obstet*. 2018;143(2):217-224.
- Cremer M, Maza M, Alfaro K, et al. Scale-Up of an Human Papillomavirus Testing Implementation Program in El Salvador. *J Low Genit Tract Dis*. 2017;21(1):26-32.
- Obiri-Yeboah D, Adu-Sarkodie Y, Djigma F, et al. Self-collected vaginal sampling for the detection of genital human papillomavirus (HPV) using careHPV among Ghanaian women. *BMC Womens Health*. 2017;17(1):86.
- Toliman PJ, Kaldor JM, Badman SG, et al. Evaluation of self-collected vaginal specimens for the detection of high-risk human papillomavirus infection and the prediction of high-grade cervical intraepithelial lesions in a high-burden, low-resource setting. *Clin Microbiol Infect*. 2019;25(4):496-503.
- Phoolcharoen N, Kantathavorn N, Krisorakun W, et al. Agreement of self- and physician-collected samples for detection of high-risk human papillomavirus infections in women attending a colposcopy clinic in Thailand. *BMC Res Notes*. 2018;11(1):136.

## Appendix 1: Selection of Included Studies



## Appendix 2: Characteristics of Included Publications

**Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses**

First Author, Publication Year, Country	Study Designs and Numbers of Primary Studies Included	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
<b>Arbyn et al. 2018,<sup>3</sup> Belgium</b>	<p>Diagnostic test accuracy; RCTs for response outcomes</p> <p>56 accuracy studies and 25 participation trials</p> <p>Databases searched: Medline (PubMed), Embase, and CENTRAL</p> <p>Study selection criteria - - Diagnostic studies: “a vaginal sample was collected by a woman herself (self sample) followed by a cervical sample collected by a clinician (clinician sample); the same hrHPV assay was performed on both samples; and the presence or absence of CIN2+ was verified by colposcopy and biopsy in all enrolled women, or in women with one or more positive tests. Studies with cytological follow-up for women with negative colposcopy results at baseline assessment were accepted as well, but were indexed for sensitivity analyses” (p. 2)</p>	<p>Individuals participating in diagnostic test accuracy studies or RCTs</p>	<p>Self sampling arm (intervention arm) Invited to collect a self sample for hrHPV testing;</p> <p>versus</p> <p>Control arm Invited or reminded to undergo a screening test on a clinician sample</p> <p>HPV tests identified: signal amplification (including Hybrid Capture and Cervista), polymerase chain reaction (PCR, including GP5+/5+ PCR-EIA, Abbott RT PCR hrHPV, Anyplex II HR, cobas 4800 HPV test, GP5+/6+-LMNX, Linear Array, HPV Risk assay, Xpert HPV)</p>	<p>Diagnostic test accuracy</p> <p>Response rates to screening; test positivity rates, adherence to follow-up in women who were screened, and detection of CIN2+.</p> <p>Follow-up lengths not reported</p>
<b>Kelly et al. 2017,<sup>14</sup> UK</b>	<p>Study selection criteria: Cross-sectional or cohort studies evaluating HPV-POC tests against histological endpoint of CIN2+ or CIN3+</p>	<p>29,657 women in 7 studies using careHPV for the detection of CIN2+</p> <p>27,845 women in 4 studies using careHPV for the detection of</p>	<p>Self- versus Physician-collected samples HPV tests</p>	<p>Diagnostic test accuracy including sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV; including 95% CIs).</p>



**Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses**

First Author, Publication Year, Country	Study Designs and Numbers of Primary Studies Included	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
	<p>Databases searched: Medline, Embase, Global Health and CINAHL</p> <p>English language only, human subjects only, 1 January 2004 to 25 February 2017</p>	<p>CIN3+</p> <p>Population inclusion criteria: Any sexually active populations consistent with the WHO screening guidelines in any geographical location, female only, include HIV-positive patients</p>	careHPV or OncoE6	Follow-up lengths not reported

CENTRAL = Cochrane Central Register of Controlled Trials; CIN = cervical intraepithelial neoplasia; EIA = enzyme immunoassay; HPV = human papilloma virus; hrHPV = high-risk human papilloma virus; PCR = polymerase chain reaction; RCT = randomized controlled trial; WHO = World Health Organization

**Table 3: Characteristics of Included Primary Clinical Studies**

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
<b>Randomized controlled trials</b>				
<b>Polman et al. 2019,<sup>15</sup> the Netherlands</b>	<p>RCT, non-inferiority, part of a regular screening program, IMPROVE (full terms not specified) study</p> <p>Dutch Trial register (NTR5078)</p>	<p>13,925 women analyzed</p> <p>7,643 women were included in the self-sampling group and 6,282 in the clinician-based sampling group</p> <p>Aged 29 to 61 years</p>	<p>Self-sampling group: women requested to collect their own cervicovaginal sample using an Evalyn Brush (Rovers Medical Devices BV, Oss, Netherlands)</p> <p>versus</p> <p>Clinician-based sampling group: samples collected by a general practitioner with a Cervex-Brush (Rovers Medical Devices BV)</p> <p>HPV tests: validated GP5+/6+ PCR enzyme immunoassay (Labo Biomedical Products BV, Rijswijk, Netherlands).</p>	<p><b>Primary endpoints</b> Detection of cervical intraepithelial neoplasia (CIN) of grade 2 or worse (CIN2+) and grade 3 or worse (CIN3+)</p> <p>Diagnostic test accuracy (sensitivity and specificity)</p> <p>Non-inferiority of HPV testing on self-collected versus clinician-collected samples: evaluated against a margin of 90% for the relative sensitivity and 98% for the relative specificity</p> <p>Median follow-up duration for HPV-positive women: 20 months</p>

**Table 3: Characteristics of Included Primary Clinical Studies**

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
<b>Ajenifuja et al. 2018,<sup>16</sup> Nigeria</b>	RCT, two-arm, multiple visits, single centre	194 women presenting for cervical cancer screening underwent both self- and provider sampling for HPV DNA testing using HybriBio GenoArray.	self- versus provider sampling  Group A: provider sampling before self sampling  Group B: self sampling before undergoing provider sampling  HPV tests: HPV DNA testing using HybriBio GenoArray	Degree of agreement between self and provider sampling for HPV DNA tests  Follow-up time: not reported
<b>Non-randomized studies</b>				
<b>Thay et al. 2019,<sup>17</sup> Cambodia</b>	Cohort, prospective	250 Cambodian women between 30 and 49 years of age  129 HIV-positive and 121 HIV-negative  Recruited from the National Center for HIV/AIDS Dermatology and sexually transmitted disease cohort, the Sihanouk Hospital Center of Hope's Rural Outreach Teams and the Pochentong Medical Center.	(1) self-sampled human papilloma virus (HPV) testing (careHPV system)  versus  (2) clinician-collected HPV testing  versus  (3) visualization with acetic acid  versus  (4) digital colposcopy with the Enhanced Visual Assessment System  HPV tests: careHPV system for 14 genotypes of hrHPV (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68)	HPV prevalence and cervical intraepithelial neoplasia (CIN) status  Agreement of self- and clinician-sampled HPV tests  HPV infection detected with self- or clinician-collected samples  Follow-up duration: not reported
<b>Toliman et al. 2019,<sup>22</sup> Papua New Guinea and Australia</b>	Cross-sectional	1,005 women attending for cervical cancer screening at 2 clinics	Self-collected vaginal specimens	Agreement in HPV detection

**Table 3: Characteristics of Included Primary Clinical Studies**

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
		Aged 30 to 59 years	versus  Clinician-collected cervical specimens  HPV tests: Cepheid Xpert HPV, Roche Cobas 4800 HPV and Hologic Aptima HPV assays	Sensitivity and specificity of self-sampled HPV tests for the detection of HPV infection identified with clinician-collected samples  High-grade squamous intraepithelial lesions (HSIL) detection for cytology  Follow-up not mentioned
<b>Des Marais et al. 2018,<sup>18</sup> US</b>	Observational, 2 <sup>nd</sup> phase of My Body, My Test observational study	193 women overdue for cervical cancer screening by national guidelines  Low-income, infrequently screened women  Inclusion criteria: 30–64 years of age; no history of Pap testing in the past 4 years (overdue for screening by national guidelines at the start of the study); household income below 250% of the poverty level; not pregnant; not had a hysterectomy; and uninsured, underinsured, or had Medicaid insurance.	1) a cervicovaginal sample self-collected by brush at home and returned by mail (self-home sample), 2) a cervicovaginal sample self-collected by brush in a clinic and handed to a nurse (self-clinic sample),  versus  3) a cervical sample collected by brush by a clinician during a pelvic examination (clinician sample)  HPV tests: Aptima HPV assay (E6/E7 mRNA of 14 high-risk HPV genotypes (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68)	High risk HPV, <i>Chlamydia trachomatis</i> , <i>Neisseria gonorrhoeae</i> , <i>Trichomonas vaginalis</i> , and <i>Mycoplasma genitalium</i> infection  Agreement between diagnostic test accuracy  Follow-up time not reported
<b>Lam et al. 2018,<sup>5</sup> Denmark</b>	Cohort, including part of the Copenhagen Self-sampling Initiative (CSi) implementation Study and Horizon study	4865 non-attenders who participated in self-sampling and 4291 women attending routine screening	Self-sampling in non-attenders  versus  Routine screening in	Cervical intraepithelial neoplasia grade 2 or worse (≥CIN2) detection rate  Positive predictive

**Table 3: Characteristics of Included Primary Clinical Studies**

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
		<p>Analyzed with 3347 samples collected in the Horizon study (cytology and HPV tests with physician-collected samples)</p> <p>Non-attendees: women who had not been screened for at least 4 (if aged 27–49 years) or 6 years (if aged 50–65 years).</p>	<p>women attending routine screening (cytology or contesting with HPV and cytology)</p> <p>CSi: opt-in pilot project, self sampling, HPV+ followed up by HPV and cytology contesting, some screened by general practitioner-collected cytology</p> <p>CSi non-responders: possibly symptom-based diagnosis</p> <p>Horizon study: routine screening, physician-collected cytology, additionally tested for HPV</p> <p>HPV tests: HC2, cobas, CLART, and APTIMA</p>	<p>values for the detection of CIN2+</p> <p>HPV positivity defined by CLART (Genomica, Madrid, Spain) and Onclarity (BD, Sparks, MD, USA) assays</p> <p>Follow-up: 18 months in the CSi study</p>
<p><b>Phoolcharoen et al. 2018,<sup>23</sup> Thailand</b></p>	<p>Cross-sectional, in a colposcopy clinic</p>	<p>247 pairs of samples</p> <p>Inclusion criteria: attending the colposcopy clinic, aged 30–70 years, no history of cervical cancer, no hysterectomy, and currently not pregnant.</p>	<p>Self-sampling with a dry brush</p> <p>versus</p> <p>Physician-collected endocervical samples from the same individuals</p> <p>HPV tests: Cobas4800 HPV test (Roche Molecular Diagnostics, Pleasanton, California, USA)</p>	<p>Concordance between vaginal self- and endocervical physician-collected high-risk HPV testing</p> <p>Follow-up not reported</p>
<p><b>Senkomago et al. 2018,<sup>19</sup> US</b></p>	<p>Cohort, prospective</p>	<p>350 female sex workers</p> <p>Aged 18 to 50 years</p>	<p>Self-collected cervico-vaginal specimens for hrHPV RNA testing</p>	<p>HPV prevalence</p> <p>Agreement between physician- and self-</p>

**Table 3: Characteristics of Included Primary Clinical Studies**

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
			<p>versus</p> <p>Physician collected cervical specimens for hrHPV-RNA testing and conventional cytology</p> <p>hrHPV-RNA testing every 3 months</p> <p>Conventional cytology every 6 months</p> <p>HPV tests: Aptima (qualitatively detecting E6/E7 mRNA of 14 hrHPV types: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68)</p>	<p>collected hrHPV-RNA results</p> <p>Diagnostic test accuracy of HPV tests for the detection of high-grade squamous intraepithelial cervical lesions or worse</p> <p>Follow-up between December 2, 2009, and February 15, 2013</p>
<b>Wong et al. 2018,<sup>8</sup> China</b>	Cohort	<p>68 female sex workers</p> <p>Inclusion criteria: aged 18 years or older, not currently pregnant, no known abnormal Papanicolaou test results and no symptoms of cervical cancer, genital cancer, cervical surgery, or immune treatment of the cervix during the 6 months before recruitment into the study</p>	<p>Self-sampling for HPV testing</p> <p>versus</p> <p>Clinician-obtained sample for HPV testing</p> <p>HPV DNA testing in the university laboratory</p>	<p>Agreement in HPV detection rates between clinician and HPV DNA self-sampling</p> <p>Agreement definition:                      k &lt; 0: poor                      k = 0 to 0.20: slight                      k = 0.21 to 0.40: fair                      k = 0.41 to 0.60: moderate                      k = 0.61 to 0.80: substantial                      k = 0.81 to 1.00: perfect</p> <p>Follow-up lengths not reported</p>
<b>Zhang et al. 2018,<sup>9</sup> China</b>	Cohort, prospective, the Shanxi Province Cervical Cancer Screening Study I (SPOCCS I)	<p>1,997 women</p> <p>Inclusion criteria: aged 35 to 45 with no history of cervical cancer or hysterectomy in 1999</p>	<p>HPV testing on self-collected and physician-collected samples, cytology and visual inspection with acetic acid (VIA)</p> <p>HPV tests: HC2 assays</p>	<p>Cumulative diagnostic test accuracy for the detection of CIN2+ at 6-year, 11-year and 15-year follow-up</p> <p>Follow-up in 1999 (baseline), 2005, 2010</p>

**Table 3: Characteristics of Included Primary Clinical Studies**

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
			(Hybrid Capture II, Qiagen Inc.)	and 2014
<b>Cremer et al. 2017,<sup>20</sup> US and El Salvador</b>	Cohort, phase 2 of Cervical Cancer Prevention in El Salvador, 3 phases in total	N = 8050 in phase 2, aged 30 to 49 years	self- and provider-collected specimens  HPV tests not reported	Agreement of diagnostic test accuracy  Follow-up: 6 months after screening
<b>Obiri-Yeboah et al. 2017,<sup>21</sup> Ghana</b>	Cohort, comparative frequency-matched study (1:5), part of a larger HPV and cervical cancer study conducted in the Cape Coast Teaching Hospital (CCTH)	194 women attending HIV and outpatient clinics in the Cape Coast Teaching Hospital, Ghana  Mean age 44.1 years (SD ± 11.3)  191 paired results	Self-collection of vaginal samples using the careHPV brush  versus  Clinician-collected cervical sample  HPV DNA (14 high-risk types) tests: careHPV assay (Qiagen) and HPV genotyping (Anyplex II, Seegene)	HPV detection concordance  Follow-up available if cytology required

CCTH = Cape Coast Teaching Hospital; CIN = cervical intraepithelial neoplasia; CSI = Copenhagen Self-sampling Initiative; DNA = deoxyribonucleic acid; HC2 = Hybrid Capture 2; HPV = human papillomavirus; hrHPV = high-risk human papillomavirus; HSIL = high-grade squamous intraepithelial lesions; mRNA = messenger ribonucleic acid; PCR = polymerase chain reaction; RCT = randomized controlled trial; RNA = ribonucleic acid; SD = standard deviation; SPOCCS = Shanxi Province Cervical Cancer Screening Study

## Appendix 3: Critical Appraisal of Included Publications

**Table 4: Strengths and Limitations of Systematic Reviews and Meta-Analyses using AMSTAR 2 checklist<sup>10</sup>**

Strengths	Limitations
<b>Arbyn et al., 2018<sup>3</sup></b>	
<ul style="list-style-type: none"> <li>- PICO components included in the research questions and inclusion criteria</li> <li>- Review protocol published <i>a priori</i> (update to a previous meta-analysis)</li> <li>- Selection of study designs explained</li> <li>- Comprehensive literature searches</li> <li>- Study selection in duplicate</li> <li>- Data extraction in duplicate</li> <li>- Included studies described</li> <li>- Risk of bias of the included studies appraised with published tools</li> <li>- Appropriate statistical methods used for synthesis</li> <li>- Risk of bias of the primary studies considered in meta-analysis</li> <li>- Risk of bias of the primary studies considered when discussing the results</li> <li>- Heterogeneity discussed</li> <li>- Publication bias investigated</li> <li>- Review authors' conflict of interest reported</li> </ul>	<ul style="list-style-type: none"> <li>- Excluded studies not provided</li> <li>- Sources of funding for the included studies not provided</li> </ul>
<b>Kelly et al., 2017<sup>14</sup></b>	
<ul style="list-style-type: none"> <li>- PICO components included in the research questions and inclusion criteria</li> <li>- Selection of study designs explained</li> <li>- Comprehensive literature searches</li> <li>- Included studies described</li> <li>- Risk of bias of the included studies appraised with published tools</li> <li>- Appropriate statistical methods used for synthesis</li> <li>- Risk of bias of the primary studies considered in meta-analysis</li> <li>- Risk of bias of the primary studies considered when discussing the results</li> <li>- Heterogeneity discussed</li> <li>- Publication bias investigated</li> <li>- Review authors' conflict of interest reported</li> </ul>	<ul style="list-style-type: none"> <li>- Excluded studies not provided</li> <li>- Sources of funding for the included studies not provided</li> <li>- Review protocol not published <i>a priori</i></li> <li>- Study selection not in duplicate</li> <li>- Data extraction not in duplicate</li> </ul>

PICO = population, intervention, comparator, and outcome

**Table 5: Strengths and Limitations of Clinical Studies using the Downs and Black checklist<sup>11</sup>**

Strengths	Limitations
<b>Randomized controlled trials</b>	
<b>Polman et al., 2019<sup>15</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> </ul>

**Table 5: Strengths and Limitations of Clinical Studies using the Downs and Black checklist<sup>11</sup>**

Strengths	Limitations
<ul style="list-style-type: none"> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Important adverse effects explicitly described</li> <li>- Patients lost to follow-up described</li> <li>- Individuals asked to participate representative of the population from which they were recruited</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Participants randomized to different groups</li> <li>- Confounding adjusted in the analysis</li> </ul>	<ul style="list-style-type: none"> <li>- Participants lost to follow-up not taken into account</li> <li>- Power analysis for sample sizes not conducted</li> </ul>
<b>Ajenifuja et al., 2018<sup>16</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- No patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Participants randomized to different groups</li> <li>- Confounding adjusted in the analysis</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Actual probability values (<i>P</i> values) not reported</li> </ul>
<b>Non-randomized studies</b>	
<b>Thay et al., 2019<sup>17</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- No patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> </ul>



**Table 5: Strengths and Limitations of Clinical Studies using the Downs and Black checklist<sup>11</sup>**

Strengths	Limitations
<ul style="list-style-type: none"> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	
<b>Toliman et al., 2019<sup>22</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- No patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> </ul>
<b>Des Marais et al., 2018<sup>18</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Patients lost to follow-up described</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> <li>- Participants lost to follow-up not taken into account</li> </ul>

**Table 5: Strengths and Limitations of Clinical Studies using the Downs and Black checklist<sup>11</sup>**

Strengths	Limitations
<b>Lam et al., 2018<sup>5</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Patients lost to follow-up reported</li> <li>- Participants asked to participate representative of the population which they were recruited from</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> <li>- Confounding not adjusted in the analysis</li> <li>- Participants lost to follow-up not taken into account</li> </ul>
<b>Phoolcharoen et al., 2018<sup>23</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- No patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> <li>- Confounding not adjusted in the analysis</li> </ul>
<b>Senkomago et al., 2018<sup>19</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> <li>- Confounding not adjusted in the analysis</li> </ul>

**Table 5: Strengths and Limitations of Clinical Studies using the Downs and Black checklist<sup>11</sup>**

Strengths	Limitations
<ul style="list-style-type: none"> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Actual probability values (<i>P</i> values) reported</li> <li>- Participants lost to follow-up taken into account in the analysis</li> </ul>	
<b>Wong et al., 2018<sup>8</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> <li>- Important adverse effects explicitly described</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Participants not randomized to different groups</li> <li>- Participants lost to follow-up not taken into account in the analysis</li> </ul>
<b>Zhang et al., 2018<sup>9</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> <li>- Participants lost to follow-up not taken into account in the analysis</li> </ul>

**Table 5: Strengths and Limitations of Clinical Studies using the Downs and Black checklist<sup>11</sup>**

Strengths	Limitations
<b>Cremer et al., 2017<sup>20</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- Patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> <li>- Participants lost to follow-up not taken into account in the analysis</li> </ul>
<b>Obiri-Yeboah et al., 2017<sup>21</sup></b>	
<ul style="list-style-type: none"> <li>- Hypothesis described</li> <li>- Outcomes to be measured described</li> <li>- Patient characteristics described</li> <li>- Intervention described</li> <li>- Distributions of principal confounders described</li> <li>- Main findings described</li> <li>- Estimates of random variability provided</li> <li>- No patients lost to follow-up reported</li> <li>- No significant changes to health care that the participants received, compared to that the majority received</li> <li>- Similar lengths of follow-up for participants in different groups</li> <li>- Appropriate statistical tests to assess the outcomes</li> <li>- Measures of compliance reliable</li> <li>- Accurate outcome measures</li> <li>- Participants recruited from the same period of time</li> <li>- Participants recruited from the same population</li> <li>- Confounding adjusted in the analysis</li> <li>- Actual probability values (<i>P</i> values) reported</li> </ul>	<ul style="list-style-type: none"> <li>- Participants not blinded</li> <li>- Outcome assessors not blinded</li> <li>- Intervention allocation not concealed</li> <li>- Power analysis for sample sizes not conducted</li> <li>- Important adverse effects not explicitly described</li> <li>- Participants not randomized to different groups</li> </ul>

**Table 6: Strengths and Limitations of Diagnostic Test Accuracy Studies using the QUADAS-2 checklist<sup>11</sup>**

Strengths	Limitations
Randomized controlled trials	
<b>Polman et al., 2019<sup>15</sup></b>	
<ul style="list-style-type: none"> <li>- Consecutive or random samples enrolled</li> <li>- Case-control design avoided</li> <li>- Inappropriate exclusion criteria avoided</li> <li>- Selection of patients not likely to introduce bias</li> <li>- Index test results interpreted without the knowledge of the results of the reference standard (colposcopy)</li> <li>- Index test threshold provided</li> <li>- Comparator test results interpreted without the knowledge of the results of the reference standard (colposcopy)</li> <li>- Comparator test threshold provided</li> <li>- Reference standard likely to correctly classify the target condition</li> <li>- Appropriate intervals between index tests and reference standards</li> <li>- All patients eligible for the reference standard if test positive</li> <li>- The same reference standard for all patients</li> </ul>	<ul style="list-style-type: none"> <li>- Reference standard results not interpreted without the knowledge of the results of index or comparator tests</li> <li>- Participants lost to follow-up not included in the analysis</li> </ul>
Non-randomized controlled trials	
<b>Lam et al., 2019<sup>5</sup></b>	
<ul style="list-style-type: none"> <li>- Consecutive or random samples enrolled</li> <li>- Case-control design avoided</li> <li>- Inappropriate exclusion criteria avoided</li> <li>- Selection of patients not likely to introduce bias</li> <li>- Index test results interpreted without the knowledge of the results of the reference standard (colposcopy)</li> <li>- Index test threshold provided</li> <li>- Comparator test results interpreted without the knowledge of the results of the reference standard (colposcopy)</li> <li>- Comparator test threshold provided</li> <li>- Reference standard likely to correctly classify the target condition</li> <li>- Appropriate intervals between index tests and reference standards</li> <li>- All patients eligible for the reference standard if test positive</li> <li>- The same reference standard for all patients</li> </ul>	<ul style="list-style-type: none"> <li>- Reference standard results not interpreted without the knowledge of the results of index or comparator tests</li> <li>- Participants lost to follow-up not included in the analysis</li> </ul>
<b>Zhang et al., 2019<sup>9</sup></b>	
<ul style="list-style-type: none"> <li>- Consecutive or random samples enrolled</li> <li>- Case-control design avoided</li> <li>- Inappropriate exclusion criteria avoided</li> <li>- Selection of patients not likely to introduce bias</li> <li>- Index test results interpreted without the knowledge of the results of the reference standard (colposcopy)</li> <li>- Index test threshold provided</li> <li>- Comparator test results interpreted without the knowledge of the results of the reference standard (colposcopy)</li> </ul>	<ul style="list-style-type: none"> <li>- Reference standard results not interpreted without the knowledge of the results of index or comparator tests</li> <li>- Participants lost to follow-up not included in the analysis</li> </ul>

**Table 6: Strengths and Limitations of Diagnostic Test Accuracy Studies using the QUADAS-2 checklist<sup>11</sup>**

Strengths	Limitations
<ul style="list-style-type: none"> <li>- Comparator test threshold provided</li> <li>- Reference standard likely to correctly classify the target condition</li> <li>- Appropriate intervals between index tests and reference standards</li> <li>- All patients eligible for the reference standard if test positive</li> <li>- The same reference standard for all patients</li> <li>- Long-term (15 years) follow-up of the participants</li> </ul>	

## Appendix 4: Main Study Findings and Authors' Conclusions

**Table 7: Summary of Findings Included Systematic Reviews and Meta-Analyses**

Main Study Findings	Authors' Conclusion
<b>Arbyn et al., 2018<sup>3</sup></b>	
<p>Self- versus clinician-collected samples</p> <p>hrHPV assays based on polymerase chain reaction for the detection of CIN2+</p> <ul style="list-style-type: none"> <li>- Pooled sensitivity and specificity using self samples: 96% and 79%</li> <li>- Pooled sensitivity and specificity using clinician samples: 96% and 79%</li> <li>- No significant difference in sensitivity between self samples and clinician samples to detect CIN2+ or CIN3+ (pooled ratio 0.99, 95% CI, 0.97 to 1.02)</li> <li>- Self samples with significantly lower specificity to exclude CIN2+ (2%) than clinician samples</li> </ul> <p>hrHPV assays based on signal amplification for the detection of CIN2+</p> <ul style="list-style-type: none"> <li>- Pooled sensitivity and specificity using self samples: 77% (95% CI 69% to 82%) and 84% (95% CI, 77% to 88%)</li> <li>- Pooled sensitivity and specificity using clinician samples: 93% (95% CI, 89% to 96%) and 86% (95% CI, 81% to 90%)</li> <li>- Self samples significantly less sensitive (pooled ratio 0.85, 95% CI, 0.80 to 0.89) than clinician samples</li> <li>- Self samples with significantly lower specificity to exclude CIN2+ (4%) than clinician samples</li> </ul> <p>Response rates to screening invitation</p> <ul style="list-style-type: none"> <li>- Mailing self-sampling kits to the woman's home address with higher response rates than the other two options: 1) to have a sample taken by a clinician than invitation or 2) reminder letters (pooled relative participation in intention-to-treat-analysis of 2.33, 95% CI, 1.86 to 2.91)</li> <li>- Opt-in strategies (had to request a self-sampling kit) generally not more effective than invitation letters (relative participation of 1.22, 95% CI, 0.93 to 1.61)</li> <li>- Direct offer of self-sampling devices to women in communities that were under-screened with high participation rates (&gt;75%)</li> <li>- Substantial inter-study heterogeneity (<math>I^2 &gt; 95%</math>).</li> </ul>	<ul style="list-style-type: none"> <li>- <i>"When used with hrHPV assays based on polymerase chain reaction, testing on self samples was similarly accurate as on clinician samples"</i> (p. 1)</li> <li>- <i>"Offering self sampling kits generally is more effective in reaching underscreened women than sending invitations"</i> (p. 1)</li> </ul>
<b>Kelly et al., 2017<sup>14</sup></b>	
<p>Self- versus clinician-collected samples</p> <ul style="list-style-type: none"> <li>- Pooled prevalence for CIN2+ and CIN3+: 2.3% and 1.1% respectively</li> </ul> <p>careHPV tests using clinician-collected cervical specimen</p> <ul style="list-style-type: none"> <li>- Sensitivity for CIN2+: 88.1% (95% CI, 81.4% to 92.7%)</li> <li>- Specificity for CIN2+: 83.7% (95% CI, 74.9% to 89.8%)</li> <li>- Sensitivity for CIN3+: 90.3% (95% CI, 83.4% to 94.5%)</li> <li>- Specificity for CIN3+: 85.3% (95% CI, 73.1% to 92.5%)</li> </ul> <p>careHPV tests using self-collected vaginal swabs</p>	<ul style="list-style-type: none"> <li>- <i>"CareHPV has good sensitivity and specificity for the detection of CIN2+ and CIN3+, but sensitivity was lower using self-collected vaginal samples"</i> (p. S36)</li> <li>- <i>"The specificity is lower in high HPV prevalence populations such as women living with HIV"</i> (p. S36)</li> </ul>

**Table 7: Summary of Findings Included Systematic Reviews and Meta-Analyses**

Main Study Findings	Authors' Conclusion
<ul style="list-style-type: none"> <li>- Sensitivity for CIN2+: 73.6% (95% CI, 64.9% to 80.8%)</li> <li>- Specificity for CIN2+: 88.0% (95% CI, 79.1% to 93.5%)</li> <li>- Sensitivity for CIN3+: 75.2% (95% CI, 66.8% to 82.0%)</li> <li>- Specificity for CIN3+: 90.6% (95% CI, 83.4% to 94.9%)</li> </ul> <p>OncoE6 tests (n = 2)</p> <ul style="list-style-type: none"> <li>- Sensitivity for CIN2+: 31.3% to 42.4%</li> <li>- Specificity for CIN2+: 99.1% to 99.4%</li> <li>- Sensitivity for CIN3+: 53.5%</li> <li>- Specificity for CIN3+: 98.9%</li> </ul>	

CI = confidence interval; CIN = cervical intraepithelial neoplasia; HIV = human immunodeficiency virus; HPV = human papillomavirus; hrHPV = high-risk human papillomavirus

**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
Randomized controlled trials	
<b>Polman et al., 2019<sup>15</sup></b>	
<p>Self- versus clinician-collected samples using a PCR-based HPV test</p> <p>HPV prevalence</p> <ul style="list-style-type: none"> <li>- Self-collected samples: 569 (7.4%)</li> <li>- Clinician-collected samples: 451 (7.2%)</li> <li>- Relative risk = 1.04 (95% CI, 0.92 to 1.17)</li> </ul> <p>CIN2+ sensitivity and specificity of HPV testing (unadjusted)</p> <ul style="list-style-type: none"> <li>- Sensitivity using self-samples: 92.9% (95% CI, 87.3% to 98.4%)</li> <li>- Specificity using self-samples: 93.9% (95% CI, 93.4% to 94.5%)</li> <li>- Sensitivity using clinician-samples: 96.4% (95% CI, 92.9% to 99.9%)</li> <li>- Specificity using clinician-samples: 94.2% (95% CI, 93.6% to 94.8%)</li> <li>- Not statistically differ between self-sampling and clinician-based sampling</li> <li>- Relative accuracy of sensitivity = 0.96 (95% CI, 0.90 to 1.03)</li> <li>- Relative accuracy of specificity = 1.00 (95% CI, 0.99 to 1.01)</li> </ul> <p>CIN3+ sensitivity and specificity of HPV testing (unadjusted)</p> <ul style="list-style-type: none"> <li>- Sensitivity using self-samples: 95.1% (95% CI, 88.5% to 100%)</li> <li>- Specificity using self-samples: 93.4% (95% CI, 92.9% to 94.0%)</li> <li>- Sensitivity using clinician-samples: 95.8% (95% CI, 91.2% to 100%)</li> <li>- Specificity using clinician-samples: 93.5% (95% CI, 92.9% to 94.1%)</li> <li>- Relative accuracy of sensitivity = 0.99 (95% CI, 0.91 to 1.08)</li> <li>- Relative accuracy of specificity = 1.00 (95% CI, 0.99 to 1.01)</li> </ul>	<p><i>“HPV testing done with a clinically validated PCR-based assay had similar accuracy on self-collected and clinician-collected samples in terms of the detection of CIN2+ or CIN3+ lesions” (p. 229)</i></p>



**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
<b>Ajenifuja et al., 2018<sup>16</sup></b>	
<p>Self- versus provider-collected samples using HybriBio GenoArray</p> <p>HPV prevalence</p> <ul style="list-style-type: none"> <li>- Self sampling: 12 (6.2%)</li> <li>- Provider collected samples: 19 (9.8%)</li> </ul> <p>The most common HPV type</p> <ul style="list-style-type: none"> <li>- Both: HPV 58 (2.6%)</li> </ul> <p>Prevalence of multiple HPV genotypes</p> <ul style="list-style-type: none"> <li>- Self-collected samples: 5 cases (2.6%)</li> <li>- Provider-collected samples: 1 (0.5%)</li> </ul> <p>High risk-HPV detection rate</p> <ul style="list-style-type: none"> <li>- Self sampled: 7.2%</li> <li>- Provider sampled: 6.8%</li> </ul> <p>Agreement between self- and provider-collected samples</p> <ul style="list-style-type: none"> <li>- Moderate correlation</li> <li>- <math>\kappa = 0.47</math> (95% CI, 21.3% to 72.3%, <math>P &lt; 0.05</math>)</li> </ul>	<p><i>“moderate correlation between both sampling techniques” (p. 1)</i></p>
<b>Non-randomized studies</b>	
<b>Thay et al., 2019<sup>17</sup></b>	
<p>Self-sampled samples versus clinician-sampled samples (careHPV) versus visualization with acetic acid versus digital colposcopy</p> <p>hrHPV+ (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, 68) prevalence</p> <ul style="list-style-type: none"> <li>- 56 (22.4%) overall</li> <li>- 37 (28.6%) in HIV+ women (<math>P = 0.0154</math>, compared to HIV-)</li> <li>- 19 (15.7%) in HIV- women</li> <li>- 50 (89%) in self-sampling HPV specimens (<math>P = 0.174</math>, compared to physician-collected)</li> <li>- 45 (80%) in physician-collected specimens</li> </ul> <p>Comfort in obtaining self-samples</p> <ul style="list-style-type: none"> <li>- 95.2%</li> </ul> <p>VIA+</p> <ul style="list-style-type: none"> <li>- 37/250</li> <li>- 30 underwent confirmatory biopsies for cervical intraepithelial neoplasia (CIN) (26 CIN1, 4 CIN2+)</li> </ul> <p>Confirmed dysplasia</p> <ul style="list-style-type: none"> <li>- 20 (15.5%) in the HIV+ group</li> <li>- 10 (8.26%) in HIV- women (<math>P = 0.0291</math>)</li> <li>- Accurately differentiated between CIN1 and CIN2+ lesions by contemporaneous physician impressions of the DC images</li> </ul>	<p><i>- “potential modifications of the current cervical screening strategy that is currently being employed in Cambodia” (p. 1)</i></p> <p><i>- “The first step in this new strategy would be self-swabbing for hrHPV. Subsequently, hrHPV+ patients would have DC and immediate treatment based on colposcopic findings: cryotherapy for suspected CIN1 and loop electrosurgical excision procedure (LEEP) for suspected CIN2+” (p. 1)</i></p>
<b>Toliman et al., 2019<sup>22</sup></b>	
<p>Self-collected versus clinician-collected samples using Xpert, Cobas, and Aptima</p>	<p><i>“Xpert HPV, using self-collected vaginal specimens, has sufficient accuracy for use in point-of-care ‘test-and-treat’</i></p>

**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
<p>Agreement in hrHPV detection between self- and clinician-samples specimens across all three assays</p> <ul style="list-style-type: none"> <li>- substantial (<math>k &gt; 0.6</math> in 32 pairwise comparisons of HPV tests)</li> </ul> <p>Sensitivity, specificity, and positive and negative predictive values for the detection of HPV type 16 according to the constructed reference standard (HPV infection detected with clinician-collected samples) using self-collected specimens</p> <ul style="list-style-type: none"> <li>- Xpert HPV: 92.1%, 93.1%, 63.6% and 98.9%</li> <li>- Cobas 4800 HPV: 90.4%, 94.3%, 67.8% and 98.7%</li> <li>- Aptima HPV: 63.2%, 97.2%, 75.0% and 95.3%</li> <li>- Similar results observed for all hrHPV types (combined) and for HPV types 18/45, on all three assays</li> </ul> <p>Detection of any hrHPV using self-collected specimens on all assays for HSIL positivity</p> <ul style="list-style-type: none"> <li>- High sensitivity (86% to 92%)</li> <li>- High specificity (87% to 94%)</li> <li>- High negative predictive value (&gt;98%)</li> </ul>	<p><i>cervical screening strategies in high-burden, low-resource settings</i>" (p. 496)</p>
<b>Des Marais et al., 2018<sup>18</sup></b>	
<p>Self-home versus self-clinic versus clinician samples using Aptima</p> <p>Prevalence of high-risk HPV</p> <ul style="list-style-type: none"> <li>- Self-home samples: 12.4%</li> <li>- Clinical samples: 11.4% [not significantly different from self-home samples (<math>P = 0.79</math>)]</li> <li>- Self clinic samples: 15.5% (not significantly different from self-home samples, <math>P = 0.21</math>)</li> </ul> <p>Positivity for high-risk HPV</p> <ul style="list-style-type: none"> <li>- Increased with increasing grades of cervical abnormality in all sample types (<math>P &lt; 0.001</math>)</li> </ul> <p>Positivity for high-risk HPV in all identified cases of high-grade squamous intraepithelial lesions and of cervical intraepithelial neoplasia 2 or worse</p> <ul style="list-style-type: none"> <li>- Self-home samples: detected all cases</li> <li>- Comparable across sample types for <i>T vaginalis</i> (range 10.2% to 10.8%), <i>M genitalium</i> (3.3% to 5.5%), <i>C trachomatis</i> (1.1% to 2.1%), and <i>N gonorrhoeae</i> (0% to 0.5%)</li> </ul> <p>Agreement measured by Kappa values between sample types</p> <ul style="list-style-type: none"> <li>- high-risk HPV: 0.56 to 0.66 (moderate to good)</li> <li>- <i>T vaginalis</i>: 0.86 to 0.91</li> <li>- <i>M genitalium</i>: 0.65 to 0.83</li> </ul> <p>Instruction understanding</p> <ul style="list-style-type: none"> <li>- No difficulty understanding self-collection instructions: 93.6%</li> <li>- Willing to use self collection in the future: 96.3%</li> </ul>	<p><i>"Mail-based, at-home self-collection for high-risk HPV and sexually transmitted infection detection was valid and well accepted among infrequently screened women in our study"</i> (p. 1412)</p>

**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
<b>Lam et al., 2018<sup>5</sup></b>	
<p>Self sampling (HC2, Cobas, CLART, and Aptima) versus routine screening (cytology)</p> <p>CIN2+ detection</p> <ul style="list-style-type: none"> <li>- Self-sampling: higher than routine cytology-based screening (OR = 1.83, 95% CI, 1.21 to 2.77)</li> <li>- Self-sampling: similar to routinely screening with cytology and HPV testing (OR = 1.03, 95% CI, 0.75 to 1.40)</li> </ul> <p>Positive predictive value for CIN2+</p> <ul style="list-style-type: none"> <li>- Screening non-attenders: higher than routinely HPV- and cytology-screened screened women (36.5% vs 25.6%, respectively)</li> <li>- Among the adequate biopsies                             <ul style="list-style-type: none"> <li>- CSI-attenders (self-sampling) and GP-attenders (clinician-collected cytology) with higher PPV (36.5% and 32.7%, respectively) than non-responders (20.0%) and women included in the Horizon study (contesting with cytology and HPV test 25.6%)</li> </ul> </li> </ul>	<p><i>“Self-sampling offered to non-attenders showed higher detection rates for ≥CIN2 than routine cytology-based screening, and similar detection rates as HPV and cytology co-testing” (p. 138)</i></p>
<b>Phoolcharoen et al., 2018<sup>23</sup></b>	
<p>Self- versus clinician-collected samples using Cobas</p> <p>hrHPV prevalence</p> <ul style="list-style-type: none"> <li>- Self- collected samples: 41.3%</li> <li>- Physician-collected samples: 36.0%</li> </ul> <p>Agreement between the methods</p> <ul style="list-style-type: none"> <li>- 74.5% with <math>\kappa = 0.46</math> (<math>P &lt; 0.001</math>)</li> </ul>	<p><i>“Our study revealed moderate agreement between self- and physician-collected methods for hrHPV testing” (p. 1)</i></p>
<b>Senkomago et al., 2018<sup>19</sup></b>	
<p>Self- versus clinician-collected samples using Aptima</p> <p>hrHPV-RNA prevalence from baseline to 24 months</p> <ul style="list-style-type: none"> <li>- Self-collected samples: decreased slightly from 28.5% (98/344) to 24.3% (53/218)</li> <li>- Clinician-collected samples: decreased slightly from 29.9% (103/344) to 24.3% (53/218)</li> </ul> <p>Agreement between the sampling methods</p> <ul style="list-style-type: none"> <li>- Increase over time</li> <li>- Baseline <math>k = 0.55</math> (95% CI, 0.45 to 0.65)</li> <li>- 24 months <math>k = 0.83</math> (95% CI, 0.74 to 0.91)</li> </ul> <p>Among 21 patients with HSIL+ over 24 months</p> <ul style="list-style-type: none"> <li>- Clinician-collected samples: 18 (86%) hrHPV-RNA-positive results at baseline</li> <li>- Self-collected samples: 17 (81%) hrHPV-RNA-positive results at baseline</li> <li>- hrHPV-RNA-positive results or cytology anomalies: 20 (95%) at baseline</li> </ul>	<p><i>- “Overall agreement between physician- and self-collected hrHPV-RNA results was moderate and appeared to increase over time” (p. 217)</i></p> <p><i>- “Baseline physician- and self-collected hrHPV-RNA tests were similarly strong indicators of cumulative HSIL+ over 24 months” (p. 217)</i></p>

**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
<b>Wong et al., 2018<sup>8</sup></b>	
<p>Preference to self sampling using unspecified tests</p> <ul style="list-style-type: none"> <li>- 65.6% preferred HPV DNA self-sampling in the future</li> <li>- 86.7% in those without previous experience of Papanicolaou tests (<i>P</i> = 0.055)</li> </ul> <p>Overall crude agreement in HPV detection rates</p> <ul style="list-style-type: none"> <li>- 85.3% (58/68)</li> <li>- <i>k</i> = 0.69 (95% CI, 0.51 to 0.87), substantial</li> </ul> <p>Sensitivity and specificity for the detection of ASCUS+</p> <ul style="list-style-type: none"> <li>- Self-collected samples: 66.7% and 66.1%</li> </ul> <p>Positive and negative predicted values</p> <ul style="list-style-type: none"> <li>- Self-collected samples: 24.0% and 92.5%</li> </ul> <p>Prevalence of HPV</p> <ul style="list-style-type: none"> <li>- Slightly higher in self-collected samples (39.7%, 27/68) than in clinician-collected samples (36.8%, 25/68)</li> </ul> <p>Attitudes toward self-sampling</p> <ul style="list-style-type: none"> <li>- Positive, but less confident in their skills of self-sampling compared with clinicians (70.6% versus 91.2%)</li> </ul>	<p><i>“The findings showed that self-sampling could be incorporated into current cervical cancer screening approaches”</i> (p. 46)</p>
<b>Zhang et al., 2018<sup>9</sup></b>	
<p>HPV testing on self-collected and physician-collected samples (HC2), cytology and visual inspection with acetic acid (VIA) compared with each other</p> <p>Sensitivity for cumulative CIN2+ using self-collected samples</p> <ul style="list-style-type: none"> <li>- Baseline: 83.1% (95% CI, 73.7% to 89.7%)</li> <li>- 6 years: 83.3% (95% CI, 74.9% to 89.3%)</li> <li>- 11 years: 70.3% (95% CI, 62.5% to 77.2%)</li> <li>- 15 years: 63.3% (95% CI, 55.7% to 70.2%)</li> </ul> <p>Specificity for cumulative CIN2+ using self-collected samples</p> <ul style="list-style-type: none"> <li>- Baseline: 85.9% (95% CI, 84.3% to 87.4%)</li> <li>- 6 years: 87.2% (95% CI, 85.5% to 88.7%)</li> <li>- 11 years: 87.9% (95% CI, 86.1% to 89.5%)</li> <li>- 15 years: 87.0% (95% CI, 85.0% to 88.8%)</li> </ul> <p>Sensitivity for cumulative CIN2+ using clinician-collected samples</p> <ul style="list-style-type: none"> <li>- Baseline: 97.6% (95% CI, 91.6% to 99.3%)</li> <li>- 6 years: 96.1% (95% CI, 90.4% to 98.5%)</li> <li>- 11 years: 82.1% (95% CI, 75.0% to 87.5%)</li> <li>- 15 years: 73.5% (95% CI, 66.3% to 79.6%)</li> </ul> <p>Specificity for cumulative CIN2+ using clinician-collected samples</p> <ul style="list-style-type: none"> <li>- Baseline: 84.7% (95% CI, 83.0% to 86.3%)</li> <li>- 6 years: 86.2% (95% CI, 84.4% to 87.7%)</li> <li>- 11 years: 86.7% (95% CI, 84.8% to 88.4%)</li> <li>- 15 years: 86.1% (95% CI, 84.0% to 87.9%)</li> </ul>	<p><i>“Self-collected HPV testing demonstrates lower sensitivity than physician-collected HPV testing but performs comparably to cytology prospectively and provides satisfactory assurance against CIN2+”</i> (p. 222)</p>

**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
<p>Prospective PPV of cumulative CIN2+</p> <ul style="list-style-type: none"> <li>- Self-collected: 83.3% (95% CI, 74.9% to 89.3%), 70.3% (95% CI, 62.5% to 77.2%) and 63.3% (95% CI, 55.7% to 70.2%) at 6 years, 11 years and 15 years</li> </ul> <p>Relative cumulative sensitivity of clinician-collected versus self-collected HPV testing</p> <ul style="list-style-type: none"> <li>- Stable over 15 years at about 1.16. Cumulative sensitivity of self-collected HPV testing</li> <li>- Comparable to cytology</li> <li>- Significantly higher than VIA</li> </ul> <p>CIN2+ during 6-year follow-up and 15 years after baseline among women positive HPV tests at baseline</p> <ul style="list-style-type: none"> <li>- Self-collected: 26.2% (95% CI, 21.5% to 30.9%)</li> <li>- Physician-collected: similar</li> </ul> <p>Protection against CIN2+ of negative self-collected HPV results</p> <ul style="list-style-type: none"> <li>- Greater than VIA</li> <li>- CIN2+ cumulative incident rates: 1.1% at the 6-year follow-up</li> </ul>	
<b>Cremer et al., 2017<sup>20</sup></b>	
<p>Response rates</p> <ul style="list-style-type: none"> <li>- CM (colposcopy management) cohort: 216 (44.2%) completed (203 treated, 11 diagnosed negative, 2 pregnant)</li> <li>- ST (screen and treat) cohort: 411 (88.4%) completed (407 treated, 2 diagnosed negative, 1 pregnant)</li> </ul> <p>Overall agreement between HPV test results from self-collected and provider-collected specimens (unspecified HPV test)</p> <ul style="list-style-type: none"> <li>- 93.7%</li> <li>- <math>\kappa</math> value = 0.70 (95% CI, 0.68 to 0.73)</li> </ul>	<ul style="list-style-type: none"> <li>- <i>“Human papillomavirus testing with ST management resulted in an approximately twice completion rate compared with CM management”</i> (p. 26)</li> <li>- <i>“Agreement between self- and provider-based sampling was good and might be used to extend screening to women in areas that are more difficult to reach”</i> (p. 26)</li> </ul>
<b>Obiri-Yeboah et al., 2017<sup>21</sup></b>	
<p>Overall HPV detection concordance using careHPV and Anyplex</p> <ul style="list-style-type: none"> <li>- 94.2% (95% CI, 89.9 to 97.1)</li> <li>- Kappa value = 0.88 (<math>P &lt; 0.0001</math>), showing excellent agreement</li> <li>- Agreement similar between HIV positive (93.8%) and negative (94.7%) women</li> </ul> <p>Sensitivity and specificity of self-collected samples for the detection of HPV infection identified by clinician-collected samples</p> <ul style="list-style-type: none"> <li>- 92.6% (95% CI, 85.3 to 97.0) and 95.9% (95% CI, 89.8 to 98.8)</li> <li>- Highest sensitivity: HIV positive women (95.7%, 95% CI, 88.0 to 99.1)</li> <li>- Highest specificity: HIV negative women (98.6%, 95% CI, 92.4 to 100)</li> </ul> <p>User experience</p> <ul style="list-style-type: none"> <li>- 76.3% women found SC very easy/easy to obtain</li> <li>- 57.7% preferred SC to CC</li> </ul>	<p><i>“The feasibility, acceptability and performance of SC using careHPV support the use of this alternative form of HPV screening among Ghanaian women”</i> (p. 1)</p>

**Table 8: Summary of Findings of Included Primary Clinical Studies**

Main Study Findings	Authors' Conclusion
- 61.9% felt SC would increase their likelihood to access cervical cancer screening	

ASCUS = atypical squamous cells of undetermined significance; CC = clinician-collected; CI = confidence interval; CIN = cervical intraepithelial neoplasia; CM = colposcopy management; CSi = Copenhagen Self-sampling Initiative; DC = digital colposcopy; GP = general practitioner; HC2 = Hybrid Capture 2; HIV = human immunodeficiency virus; HPV = human papillomavirus; hrHPV = high-risk human papillomavirus; HSIL = high-grade squamous intraepithelial lesion; LEEP = loop electrosurgical excision procedure; OR = odds ratio; PCR = polymerase chain reaction; PPV = positive predictive value; RNA = ribonucleic acid; SC = self-collection; ST = screen-and-treat; VIA = visual inspection with acetic acid

## Appendix 5: Overlap between Included Systematic Reviews

**Table 9: Primary Study Overlap between Included Systematic Reviews**

Primary Study Citation	Systematic Review Citation (n = 81)	
	Arbyn et al., 2018 <sup>3</sup> Accuracy studies (n = 76)	Kelly et al., 2017 <sup>14</sup> (n = 8)
Morrison 1992	X	
Hillemanns 1999	X	
Sellors 2000	X	
Wright 2000	X	
Belinson 2001	X	
Lorenzato 2002	X	
Nobbenhuis 2002	X	
Garcia 2003	X	
Salmerón 2003	X	
Brink 2006	X	
Daponte 2006	X	
Girianelli 2006	X	
Holanda 2006	X	
Seo 2006	X	
Szarewski 2007	X	
Qiao YL 2008	X	X
Bhatla 2009	X	
Balasubramanian 2010	X	
Gustavsson 2011	X	
Taylor 2011	X	
Twu 2011	X	
Belinson 2012	X	
Dijkstra 2012	X	
Longatto-Filho 2012	X	
van Baars 2012	X	
Zhao FH 2012	X	
Darlin 2013a	X	
Darlin 2013b	X	
Geraets 2013	X	

**Table 9: Primary Study Overlap between Included Systematic Reviews**

Primary Study Citation	Systematic Review Citation (n = 81)	
	Arbyn et al., 2018 <sup>3</sup> Accuracy studies (n = 76)	Kelly et al., 2017 <sup>14</sup> (n = 8)
Guan 2013	X	
Jentschke 2013a	X	
Jentschke 2013b	X	
Nieves 2013	X	
Bais 2007	X	
Gök 2010	X	
Giorgi Rossi 2011	X	
Lazcano-Ponce 2011	X	
Piana 2011	X	
Szarewski 2011	X	
Virtanen 2011	X	
Wikström 2011	X	
Gök 2012	X	
Sancho-Garnier 2013	X	
Broberg 2014	X	
Cadman 2015	X	
Haguenoer 2014	X	
Arrossi 2015	X	
Giorgi Rossi 2015	X	
Tranberg 2018	X	
Zhao FH 2013	X	X
Chernesky 2014	X	
Hesselink 2014	X	
Jeronimo 2014	X	X
Wang 2014	X	
Zhang S 2014	X	
Boggan 2015	X	
Porras 2014	X	
Chen Q 2016	X	
Chen K 2016	X	



**Table 9: Primary Study Overlap between Included Systematic Reviews**

Primary Study Citation	Systematic Review Citation (n = 81)	
	Arbyn et al., 2018 <sup>3</sup> Accuracy studies (n = 76)	Kelly et al., 2017 <sup>14</sup> (n = 8)
Jentschke 2016	X	
Qin Y 2016	X	
Stanczuk 2016	X	
Aiko 2017	X	
Asciutto 2017	X	
Catarino 2017	X	
Leeman 2017	X	
Asciutto 2018	X	
Leinonen 2018	X	
Enerly 2016	X	
Moses 2015	X	
Racey 2016	X	
Sultana 2016	X	
Zehbe 2016	X	
Kitchener 2018	X	
Modibbo 2017	X	
Kellen 2018	X	
Segondy 2016		X
Tuerxun 2016		X
Bansil 2015		X
Gage 2012		X
Chibweshwa 2016		X

## Appendix 6: Additional References of Potential Interest

### Primary studies using participants with known HPV or cytology status

Aiko KY, Yoko M, Saito OM, et al. Accuracy of self-collected human papillomavirus samples from Japanese women with abnormal cervical cytology. *J Obstet Gynaecol Res.* 2017;43(4):710-717.

El-Zein M, Bouten S, Louvanto K, et al. Validation of a new HPV self-sampling device for cervical cancer screening: The Cervical and Self-Sample In Screening (CASSIS) study. *Gynecol Oncol.* 2018;149(3):491-497.

Tranberg M, Jensen JS, Bech BH, Blaakaer J, Svanholm H, Andersen B. Good concordance of HPV detection between cervico-vaginal self-samples and general practitioner-collected samples using the Cobas 4800 HPV DNA test. *BMC Infect Dis.* 2018;18(1):348.