Diagnostic accuracy of self collected vaginal specimens for human papillomavirus compared to clinician collected human papillomavirus specimens: a meta-analysis

G S Ogilvie, D M Patrick, M Schulzer, J W Sellors, M Petric, K Chambers, R White, J M FitzGerald

Background/objectives: Providing summary recommendations regarding self collection of vaginal specimens for human papillomavirus (HPV) testing is difficult owing to the wide range of published estimates for the diagnostic accuracy of this approach. To determine summary estimates from analyses of reported findings of the sensitivity, specificity and summary receiver operating characteristic curves (SROC) for self collected vaginal specimens for HPV testing compared to the reference standard, clinician collected HPV specimens.

Methods: Standard search criteria for a diagnostic systematic review were employed. Eligible studies were combined using a random effects model and summary ROC curves were derived for overall and for specific subgroups.

Results: Summary measures were determined from 12 studies. Six studies where patients used Dacron or cotton swabs or cytobrushes to obtain samples were pooled and had an overall sensitivity of 0.74 (95% CI 0.61 to 0.84) and specificity of 0.88 (95% CI 0.83 to 0.92), with diagnostic odds ratio of 22.3 and an area under the curve of 0.91. Self specimens using Dacron or cotton swabs or cytobrushes collected by women enrolled at referral clinics had an overall sensitivity of 0.81 (95% CI 0.65 to 0.91) and specificity of 0.90 (95% CI 0.80 to 0.95). Sensitivity and specificity of tampons ranged from 0.67–0.94 and 0.80–0.85 respectively.

Conclusions: Our findings indicate that the combined sensitivity for HPV-DNA is more than 70% when patients use Dacron swabs, cotton swabs, or cytobrushes to obtain their own vaginal specimens for HPV-DNA evaluation. Self collected HPV-DNA swabs may be an appropriate alternative for low resource settings or in patients reluctant to undergo pelvic examinations.

Although largely preventable, cervical cancer remains a common worldwide malignancy.1 The widespread screening for cervical cancer by Papanicolaou (Pap) smear has led to a substantial decrease in the prevalence of the disease, but this screening method has recognised limitations, including poor interobserver reproducibility,1 limited correlation with disease process,1,4 and poor uptake by women of lower socioeconomic status2 and women deemed at risk.5 In addition, Pap screening programmes require significant infrastructure and resources.6 All of these limitations have prompted the search for improved methods of screening for cervical cancer.

Human papillomavirus (HPV) is now well established as a common worldwide malignancy.1 The widespread screening for cervical cancer by Papanicolaou (Pap) smear has led to a substantial decrease in the prevalence of the disease, but this screening method has recognised limitations, including poor interobserver reproducibility,1 limited correlation with disease process,1,4 and poor uptake by women of lower socioeconomic status2 and women deemed at risk.5 In addition, Pap screening programmes require significant infrastructure and resources.6 All of these limitations have prompted the search for improved methods of screening for cervical cancer.

One possible advantage that HPV-DNA testing offers for cervical cancer screening programmes is the method of specimen collection. While Pap smear collection requires a pelvic examination, collection of a vaginal specimen for HPV-DNA testing can be performed by the patients themselves. In resource limited settings, patient collected specimens for HPV-DNA might be acceptable as the primary screening test for cervical cancer, thus decreasing the need for practitioners to conduct screening. However, in societies with adequate healthcare resources, self collected HPV-DNA testing could be combined with Pap testing to improve cervical screening. Self collection of HPV-DNA specimens may also be more acceptable in populations that have difficulty obtaining Pap smears, such as abuse survivors7 and women with cultural concerns.15 16 Also, annual self testing for HPV-DNA could be used to screen women from geographically isolated regions without access to regular medical care, again to determine women who need to be offered further clinical evaluation.

Numerous studies have evaluated the accuracy of testing for HPV-DNA on patient collected vaginal specimens compared to the clinician collected specimens for HPV-DNA.17–20 However, summary recommendations cannot be made from these studies, because their findings are heterogeneous and a variety of specimen collection devices have been used. Because much of the proposed value of self collected HPV-DNA testing is dependent on the consistency and quality of specimen collection,

Abbreviations: AUC, area under the curve; HPV, human papillomavirus; PCR, polymerase chain reaction; ROC, receiver operating characteristic; SROC, summary receiver operating characteristic
METHODS

Data sources

The search strategy followed established methods recommended for diagnostic systematic reviews. The Embase, Cochrane Database of Abstracts of Reviews of Effectiveness, Cochrane Controlled Database of Systematic Reviews and Cochrane Central Registry of Controlled Trials were searched. Medical subject (MeSH) headings of “human papillomavirus/HPV,” “cervical neoplasms,” “cervical intraepithelial neoplasia” were exploded and combined with “self$.” The studies were then limited to English language. In addition, an expert in the field was consulted to assist in identifying any studies not found through the electronic search.

Selection of studies

Studies that met the widely accepted methodological criteria for diagnostic studies were included:

- consecutively/randomly recruited women
- reference (criterion) standard applied uniformly (that is, clinician collected specimen)
- Hybrid Capture-II (HC-II) or polymerase chain reaction (PCR) analysis of the sample
- Blinded analysis of the sample(s).

Two of the authors (GO, DP) identified and reviewed studies to be included, and agreement scores of inclusion/exclusion were calculated using Cohen’s kappa. Disagreements were resolved by consensus. Data that were abstracted from each article included number of patients tested, clinical setting (outreach, primary care, referral setting), recruitment (consecutive or random), sample type (patient and clinician), diagnostic method (PCR/HC-II), high risk, or low risk HPV evaluated, and data for a $2 \times 2$ table (true positive, false positive, false negative, true negative). If studies involved several self sampling methods (swab, tampon, cervicovaginal lavage), the first method described was used in the analysis. If we were unable to construct a $2 \times 2$ table from data available in the paper, the author was contacted in order to obtain paired data from each patient enrolled in the study.

Analysis

Heterogeneity of odds ratios was determined by the Q test. The presence or absence of a threshold effect was determined by Spearman’s rho. The kappa value between clinician and patient collected specimens was determined for each study. Summary estimates of sensitivity and specificity were pooled and weighted using Meta-test software (Joseph Lau, MD, New England Medical Center, Boston, MA, USA). DerSimonian and Laird random effects model was employed for all estimates. Summary estimates were generated for studies that used similar swab types, while ranges for sensitivity and specificity were provided for other subgroups such as diagnostic method or recruitment site. Reference standard in each case was the clinician obtained specimen.

Heterogeneity in diagnostic studies arises from a variety of different sources, including study design and patient populations. Given the interdependent nature of sensitivity and specificity different diagnostic thresholds will provide varying sensitivities and specificities. In order to address this, receiver operating characteristic (ROC) curves were used to provide a graphical representation of the diagnostic characteristics of tests at varying thresholds. Using the method described by Moses, a summary ROC (SROC) was plotted for studies that used similar swab types (Dacron or cotton swab or cytobrush). Cox adjustment was employed to avoid undefined transformations, and outlier studies were identified by visual inspection of logit regression plots. Q* estimate was also provided. The Q* estimate is the optimal estimate of the performance of the test, corresponds to the meta-analytically estimated values of the sensitivity and specificity of the test at the point where the pooled ROC curve crosses the negative diagonal. It is the point where sensitivity and specificity are equal, and is an indicator of the proximity of the ROC curve to the upper left hand corner. Area under the curve (AUC), which is a measure of the ability of a test to assign the correct value to a random pair of infected and non-infected individuals, was calculated for each SROC from Meta-test software.

RESULTS

Of the 821 studies identified in the search, 106 studies that included either clinician or self collected specimen for HPV-DNA were reviewed. Abstracts were reviewed for the entry criteria. Agreement on inclusion/exclusion of a study for the meta-analysis was $k = 0.98$ (95% CI, 0.96 to 1.00), and inclusion of the disputed study was made after careful

Table 1 Summary of study data

<table>
<thead>
<tr>
<th>Study author</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sellors</td>
<td>109</td>
<td>16</td>
<td>7</td>
<td>68</td>
<td>200</td>
</tr>
<tr>
<td>Hillermann</td>
<td>85</td>
<td>9</td>
<td>32</td>
<td>121</td>
<td>247</td>
</tr>
<tr>
<td>Harper</td>
<td>16</td>
<td>8</td>
<td>11</td>
<td>58</td>
<td>93</td>
</tr>
<tr>
<td>Nobbenhuis</td>
<td>30</td>
<td>12</td>
<td>0</td>
<td>12</td>
<td>54</td>
</tr>
<tr>
<td>Morrison</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Wright</td>
<td>170</td>
<td>132</td>
<td>128</td>
<td>985</td>
<td>1415</td>
</tr>
<tr>
<td>Rombolo</td>
<td>72</td>
<td>40</td>
<td>34</td>
<td>173</td>
<td>319</td>
</tr>
<tr>
<td>Chang</td>
<td>144</td>
<td>11</td>
<td>0</td>
<td>1039</td>
<td>1194</td>
</tr>
<tr>
<td>Harper</td>
<td>28</td>
<td>4</td>
<td>63</td>
<td>103</td>
<td>103</td>
</tr>
<tr>
<td>Lorenzato</td>
<td>45</td>
<td>27</td>
<td>10</td>
<td>171</td>
<td>253</td>
</tr>
<tr>
<td>Coullie</td>
<td>31</td>
<td>4</td>
<td>2</td>
<td>11</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>138</td>
<td>9</td>
<td>24</td>
<td>99</td>
<td>270</td>
</tr>
</tbody>
</table>

TP: true positive; TN: true negative; FP: false positive; FN: false negative.
<table>
<thead>
<tr>
<th>No</th>
<th>Author</th>
<th>Pap smear</th>
<th>Recruitment site</th>
<th>No</th>
<th>Patient collection device</th>
<th>Criterion standard</th>
<th>Diagnostic method</th>
<th>HPV types</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>-LR</th>
<th>-LR</th>
<th>OR</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sellors</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>200</td>
<td>Dacron swab</td>
<td>Cone-shaped brush</td>
<td>HC-II</td>
<td>HR</td>
<td>0.87 (0.8 to 0.92)</td>
<td>0.91 (0.81 to 0.96)</td>
<td>9.67</td>
<td>0.14</td>
<td>60.6</td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>Hillemanns</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>247</td>
<td>Cytobrush</td>
<td>Cytobrush</td>
<td>HC-II</td>
<td>Both</td>
<td>0.90 (0.82 to 0.95)</td>
<td>0.79 (0.72 to 0.85)</td>
<td>4.29</td>
<td>0.13</td>
<td>33.65</td>
<td>0.66</td>
</tr>
<tr>
<td>3</td>
<td>Harper</td>
<td>ABN</td>
<td>Gynaecology</td>
<td>93</td>
<td>Tampon</td>
<td>Dacron swab</td>
<td>HC-II</td>
<td>Both</td>
<td>0.67 (0.45 to 0.83)</td>
<td>0.84 (0.73 to 0.91)</td>
<td>4.19</td>
<td>0.39</td>
<td>9.87</td>
<td>0.49</td>
</tr>
<tr>
<td>4</td>
<td>Nobbenhuis</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>54</td>
<td>CVL</td>
<td>Cervex brush</td>
<td>PCR</td>
<td>HR</td>
<td>0.71 (0.55 to 0.84)</td>
<td>1.00 (0.74 to 1.00)</td>
<td>1.00</td>
<td>0.05</td>
<td>145</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>Morrison</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>16</td>
<td>CVL</td>
<td>Not stated</td>
<td>PCR</td>
<td>Both</td>
<td>1.00 (0.77 to 1.00)</td>
<td>0.88 (0.86 to 0.9)</td>
<td>4.67</td>
<td>0.45</td>
<td>9.87</td>
<td>0.45</td>
</tr>
<tr>
<td>6</td>
<td>Wright</td>
<td>No Pap</td>
<td>Outreach</td>
<td>1415</td>
<td>Dacron swab</td>
<td>Cone-shaped brush</td>
<td>HC-II</td>
<td>HR</td>
<td>0.56 (0.50 to 0.62)</td>
<td>0.84 (0.78 to 0.88)</td>
<td>4</td>
<td>0.43</td>
<td>9.00</td>
<td>0.48</td>
</tr>
<tr>
<td>7</td>
<td>Rompalo</td>
<td>Sx</td>
<td>Primary care</td>
<td>319</td>
<td>Dacron swab</td>
<td>Dacron swab</td>
<td>PCR</td>
<td>Both</td>
<td>0.64 (0.55 to 0.73)</td>
<td>1.00 (1.00 to 1.00)</td>
<td>93</td>
<td>0.07</td>
<td>26123.09</td>
<td>0.96</td>
</tr>
<tr>
<td>8</td>
<td>Chang</td>
<td>No Pap</td>
<td>Outreach</td>
<td>1194</td>
<td>Cotton</td>
<td>Not stated</td>
<td>HC-II</td>
<td>HR</td>
<td>0.93 (0.87 to 0.96)</td>
<td>0.94 (1.00 to 1.00)</td>
<td>13</td>
<td>0.23</td>
<td>47.31</td>
<td>0.74</td>
</tr>
<tr>
<td>9</td>
<td>Harper</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>103</td>
<td>Dacron swab</td>
<td>Dacron swab</td>
<td>PCR</td>
<td>Both</td>
<td>0.78 (0.60 to 0.89)</td>
<td>0.94 (0.85 to 0.98)</td>
<td>10.5</td>
<td>0.39</td>
<td>27.02</td>
<td>0.62</td>
</tr>
<tr>
<td>10</td>
<td>Lorenzato</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>253</td>
<td>Cotton swab</td>
<td>Cytobrush</td>
<td>PCR</td>
<td>Both</td>
<td>0.63 (0.50 to 0.73)</td>
<td>0.85 (0.90 to 0.97)</td>
<td>5.93</td>
<td>0.13</td>
<td>32.20</td>
<td>0.7</td>
</tr>
<tr>
<td>11</td>
<td>Fairley</td>
<td>ABN</td>
<td>Colposcopy</td>
<td>48</td>
<td>Tampon</td>
<td>Ayre Spatula</td>
<td>PCR</td>
<td>Both</td>
<td>0.89 (0.72 to 0.96)</td>
<td>0.80 (0.54 to 0.97)</td>
<td>4.7</td>
<td>0.075</td>
<td>59.21</td>
<td>0.75</td>
</tr>
<tr>
<td>12</td>
<td>Coutlee</td>
<td>HIV+/HIV-</td>
<td>Women</td>
<td>270</td>
<td>Tampon</td>
<td>CVL</td>
<td>PCR</td>
<td>Both</td>
<td>0.94 (0.88 to 0.97)</td>
<td>1.00 (0.72 to 0.87)</td>
<td>4.7</td>
<td>0.075</td>
<td>59.21</td>
<td>0.75</td>
</tr>
</tbody>
</table>

ABN, women with normal and abnormal Pap smears included; Sx, women with symptoms of, contacts of, or at risk for sexually transmitted infections included; CVL, cervicovaginal lavage; HC-II, Digene Hybrid Capture II technology; PCR, polymerase chain reaction; +LR, positive likelihood ratio; -LR, negative likelihood ratio; OR, odds ratio; HR, high risk.
revision of the abstract. Sixteen studies were deemed eligible for the meta-analysis.17-29 However, one study was a case-control study and therefore was excluded at the time of data extraction.22 Although six studies did not provide the raw data needed for calculations, on request these data were provided by three authors. The remaining two authors refused to provide raw data, leaving 12 studies to be combined (table 1).

Q test for heterogeneity was conducted overall and for subgroups. All were significant (p<0.01) with the exception of the studies enrolling women with abnormal Pap smears at referral centres. Overall, Spearman’s rho was –0.1, indicating no threshold effect.

As part of constructing the SROC, SROC logit regression plots were generated. Visual inspection of D on S regression identified one study24 as an extreme outlier, with extreme values of both D[0.2] and S[-5.1] compared to the other studies. In addition, inspection of this study demonstrated a specificity of 1.0 (95% CI 1.0 to 1.0), with no false positive specimens, compared to specificities of 0.79 to 0.94 in the other studies. This study also had a sample size of 1194 and therefore exerted a strong influence on the SROC and summary estimates. As such, estimates are presented with and without this outlier study.

Sensitivity of self collected vaginal specimens for HPV-DNA ranged from 0.56–1.00 and specificity ranged from 0.79-1.00 (table 2). The kappa values between patient and clinician obtained samples in individual studies ranged from 0.45–1.00. Six studies,17 18 20 21 25 27 including 2537 subjects where patients used Dacron or cotton swabs or cytobrushes to obtain samples, were pooled and had an overall sensitivity of 0.74 (95% CI 0.61 to 0.84), specificity of 0.88 (95% CI 0.83 to 0.92), and a diagnostic odds ratio of 22.3 (95% CI 11.7 to 42.6). When the outlier study was included,24 there were 3731 subjects included and the summary sensitivity for Dacron swabs, cotton swabs, or cytobrushes was 0.78 (95% CI 0.65 to 0.95) and a diagnostic odds ratio was 35.5, with a 95% CI 15.3 to 82.3.

Four studies where self specimens were obtained with Dacron swabs, cotton swab, or cytobrush from women recruited at referral clinics18 20 21 25 27 included 803 subjects and had an overall sensitivity of 0.81 (95% CI 0.65 to 0.91), specificity of 0.80 (95% CI 0.80 to 0.95), and a diagnostic odds ratio of 37.6 (95% CI 24.2 to 58.4). Three studies using tampons17 18 25 27 with 411 subjects had a range for sensitivity between 0.67 and 0.94 and a specificity ranging from 0.8 to 0.85. Seven studies used PCR as the diagnostic method,17 18 20 21 24 25 26 and specimen collection types included cervicovaginal lavage, Dacron swabs, cotton swabs, and tampons. Sensitivity for PCR ranged from 0.63–1.00 and specificity ranged from 0.80–1.00. Five studies used Hybrid Capture-II as the diagnostic method,17 18 21 24 26 and had sensitivities ranging from 0.56–0.93 and specificities ranging from 0.79–1.00. Three studies were conducted in an outreach/primary care setting,20 21 24 and Dacron or cotton swabs were used to obtain the sample. Sensitivity ranged from 0.56–0.93 and specificity ranged from 0.84–1.00 (table 3).

Summary ROC curves are shown in figure 1. Areas under the curve for studies using Dacron swab, cotton swab, or cytobrush was 0.91 and the Q* estimate was 0.85 (95% CI

---

### Table 3 Summmary of findings

<table>
<thead>
<tr>
<th>No of studies</th>
<th>No</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Odds ratio (95% CI)</th>
<th>AUC</th>
<th>Q* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dacron/cotton swab or cytobrush</td>
<td>6</td>
<td>2537</td>
<td>0.74 (0.61 to 0.84)</td>
<td>0.88 (0.83 to 0.92)</td>
<td>22.3 (11.7 to 42.6)</td>
<td>0.91</td>
</tr>
<tr>
<td>Dacron/cotton swab or cytobrush</td>
<td>7</td>
<td>3731</td>
<td>0.78 (0.65 to 0.88)</td>
<td>0.90 (0.85 to 0.94)</td>
<td>35.5 (15.3 to 82.3)</td>
<td>NC</td>
</tr>
<tr>
<td>Referral clinics</td>
<td>4</td>
<td>803</td>
<td>0.81 (0.65 to 0.91)</td>
<td>0.90 (0.78 to 0.95)</td>
<td>37.6 (24.2 to 58.4)</td>
<td>0.93</td>
</tr>
<tr>
<td>Tampons</td>
<td>3</td>
<td>411</td>
<td>0.67–0.94*</td>
<td>0.8–0.95*</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>PCR</td>
<td>7</td>
<td>1063</td>
<td>0.63–1.00*</td>
<td>0.8–1.00</td>
<td>NC</td>
<td>NC</td>
</tr>
<tr>
<td>Hybrid Capture-II</td>
<td>5</td>
<td>3149</td>
<td>0.56–0.93*</td>
<td>0.79–1.00*</td>
<td>9.0–26123.1*</td>
<td>NC</td>
</tr>
<tr>
<td>Primary care setting</td>
<td>3</td>
<td>2928</td>
<td>0.56–0.93*</td>
<td>0.84–1.00*</td>
<td>9.0–26123.1*</td>
<td>NC</td>
</tr>
</tbody>
</table>

NC, not calculated; AUC, area under curve; Q*, Q* estimate.
*Range for studies.
†Summary estimate generated from combining studies.
‡Diagnostic odds ratio generated from combining studies.
§Outlier study included.

---

**Figure 1** Summary receiver operating characteristic curves. (A) Dacron swab, cotton swab, or cytobrush. (B) Referal setting.
0.79 to 0.91). In women recruited at referral centres with abnormal Pap smears, AUC was 0.93 and the Q* estimate was of 0.87 (95% CI 0.84 to 0.90).

**DISCUSSION**

This systematic review offers summary estimates of the diagnostic accuracy of self collected vaginal specimens using Dacron swabs, cotton swabs, or cytobrushes for HPV-DNA test compared with clinician collected samples. The overall sensitivity for self collected specimens for HPV-DNA when Dacron, cotton swabs, or cytobrushes are used was 0.74 and specificity is 0.88 compared to clinician obtained specimens using these same devices. Summary estimates increased to 0.81 and 0.90, respectively, for sensitivity and specificity when self samples are conducted in referral settings. This is likely because women with active cervical disease, as reflected by abnormal Pap tests requiring referral, are likely to have a higher burden of viral shedding, thus enabling easier detection of the HPV-DNA with the self collected specimens in this population. Tampons offered sensitivity between 0.67–0.94, but given that fewer than four studies were available, we were unable to combine them to generate a summary findings. Both PCR and HC-II offered similar ranges in terms of their sensitivities and specificities. Future studies that examine both the acceptability and the diagnostic accuracy of tampons for specimen collection would enable researchers to generate summary estimates of diagnostic accuracy for tampons, and determine if tampons offer an advantage over swabs and cytobrushes for self testing for HPV-DNA.

One outlier study was both included and excluded from the overall estimate. This study had extreme values of both D and S, a specificity of 1.00, and no false positive samples in a study with over 1100 patients enrolled. Given its large sample size, this study was strongly influential on the regression and SROC. Addition of this outlier study increased the sensitivity and specificity of 0.87 (95% CI 0.84 to 0.90).

**REFERENCES**