Chapter 13: Primary Screening of Cervical Cancer With Human Papillomavirus Tests

Eduardo L. Franco

Despite its history of success in cancer screening, Pap cytology has important limitations, particularly its high false-negative rate, which carries important public health implications. Since the mid-1990s, there has been substantial interest in the use of human papillomavirus (HPV) DNA testing in cervical cancer screening under the premise that the testing of cervical cells for the causative agent of cervical cancer could have acceptable screening performance, while being more reproducible in clinical practice than Pap cytology. There have been several studies assessing the utility of HPV testing compared with the Pap test as a screening tool. These studies varied widely in lesion-outcome definition and in methodology. No studies were based on cervical cancer incidence or mortality. No randomized controlled trials have yet been published; all of the studies were based on concomitant testing for HPV and cytology or additional tests. HPV testing has greater sensitivity (average, 27%) but somewhat lower specificity (average, 8%) than Pap cytology for detecting high-grade lesions. Screening of women aged 30 years or older tends to improve test specificity, but it also does so for cytology. The combination of cytology and HPV attained high-negative predictive values, which suggests that their joint use could allow screening intervals to be safely increased, thus lowering costs. Although evidence is yet to come from long-term studies and from randomized controlled trials with high-grade lesions and invasive cancer as outcomes, HPV testing is clearly one of the most promising new technologies and has the potential to improve cervical cancer-screening effectiveness in many settings. [J Natl Cancer Inst Monogr 2003;31:89–96]

In addition to its role in the studies of etiology and the natural history of cervical cancer, human papillomavirus (HPV) testing has been used for three main screening or management-related purposes defined as follows:

1) Primary screening—for the detection of cases of cervical cancer or of its precursor lesions among asymptomatic women without a referral diagnosis, i.e., as true population screening, either opportunistic or systematic. HPV testing in these studies has been used to complement the result from the screening Pap smear or as a screening test in isolation.

2) Secondary triage—for the detection of cases of cervical cancer or of its precursor lesions among women who were initially found to have an abnormal Pap smear that requires further evaluation. HPV testing in these circumstances is used to complement the result of a repeat Pap cytology that is conducted in a more controlled setting or as a substitute for the repeat smear as part of a management algorithm to triage women who should undergo immediate colposcopy and biopsy and the medical decisions that stem from the results of these procedures (1).

3) Follow-up of treated cases—for improved surveillance of recurrent cervical lesions after treatment to permit more aggressive management of cases that are likely to recur because of persistent HPV detection.

The focus of this overview is on HPV testing in primary cervical cancer screening. I summarize herein the published investigations of the efficacy of HPV testing that also presented the results of Pap cytology as a comparison standard. I also discuss methodologic and study design implications for future research on the screening for cervical cancer.

RESEARCH ON HPV TESTING IN PRIMARY SCREENING

Despite its long history of success as a cancer screening tool, Pap cytology has important limitations as a laboratory test. A recent meta-analysis (2) that included only studies unaffected by verification bias indicated that, although the Pap test has adequate specificity, its average sensitivity to detect cervical intraepithelial neoplasia (CIN) or invasive cervical cancer was much lower than what it was generally believed to be. The Pap test’s high false-negative rate is thus its most critical limitation, which carries important medical, financial, and legal implications; the latter is a particularly acute problem in the United States, where false-negative smears are among the most frequent reasons for medical malpractice litigation. Since the mid-90s, there has been substantial interest in the use of HPV DNA assays as a cervical cancer screening tool under the premise that standardized molecular testing of cervical exfoliated cells for the causative agent of cervical cancer could have acceptable diagnostic performance, while being more reproducible and more easily adapted for automated, high-volume testing in clinical practice than conventional Pap cytology.

Reid et al. (3) first proposed HPV testing as an adjunct to Pap cytology in the wake of U.S. federal legislation mandating a reappraisal of cervical cancer screening programs and tighter regulation aimed at improving the quality of cytology laboratories. Since then, there have been several studies assessing the relative use of HPV testing compared with the Pap test as a primary cervical cancer screening tool. These studies are shown in Table 1. Most investigations have used first- or second-generation Hybrid Capture (HC) systems (Digene, Gaithersburg, MD), the only HPV test currently approved by the U.S. Food and Drug Administration (FDA). A few studies have used different polymerase chain reaction (PCR) protocols to detect HPV (4,12,16). PCR has a lower threshold of detectability for HPV DNA than the HC assay, but HCII (the second-generation assay

Correspondence to: Eduardo L. Franco, DrPH, Division of Cancer Epidemiology, McGill University, 546 Pine Ave. West, Montreal, PQ, Canada H2W 1S6 (e-mail: eduardo.franco@mcgill.ca).

See “Notes” following “References.”

Journal of the National Cancer Institute Monographs, No. 31, © Oxford University Press 2003, all rights reserved.
that became available commercially in 1997) has substantially improved molecular sensitivity (as compared with its first-generation counterpart, the HC tube assay, or HCI as shown in Table 1) for detecting HPV DNA that approaches that of consensus-primer or type-specific PCR techniques. In fact, the HCII is likely to perform more reproducibly than the PCR techniques that are not based on proven consensus primer protocols. The HCII test is a nucleic acid hybridization assay with signal amplification for the qualitative detection in cervical specimens of HPV DNA of 13 high-risk, cancer-associated types: 16, 18, 31, 33, 35, 39, 45, 51, 52, and 56. The HCI and HCII assays cannot determine the specific HPV type present, because detection is performed with a combined probe mix. PCR protocols are based on target amplification with type-specific or consensus or general primers followed by hybridization with specific oligoprobes. Although a few biotechnology companies are currently developing PCR-based diagnostic systems for clinical use, none are yet available commercially.

Table 1 summarizes all of the published studies using HC or PCR in cervical cancer screening for the detection of high-grade squamous intraepithelial lesions (HSILs) (CIN 2 or 3). These investigations targeted European (4,5,12,15), African (7,10,14), Asian (13), Latin American (9), and North American (8,16) populations. Three investigations are represented by more than one publication each: France (5,15), South Africa (7,10), and Zimbabwe (11,14) studies, which presented results from different subsets of the data or information on expanded accrual.

Most studies assessed the screening performance concerning prevalent lesions on the basis of simple cross-sectional designs (most studies in Table 1), whereas others (5,12,15) assessed both the prevalent and short-term incident lesions based on cross-sectional investigations with short-term follow-up. Lesion definition varied across studies and included either CIN/squamous intraepithelial lesion (SIL) of all grades or CIN 2 or 3 (HSIL) or worse lesions, diagnosed by histology on specimens obtained by colposcopy-guided biopsy. In some studies, the colposcopic result was used if no biopsy was taken. The South African (7), Costa Rican (9), and Chinese (13) studies were based on direct community recruitment, but only the Costa Rican study sampled a representative fraction of the local population.

None of these studies was based on long-term follow-up for more relevant end points, such as the incidence of CIN 2 or 3 or cancer or mortality from invasive cervical cancer. None of the investigations were randomized controlled trials (RCT); all of the studies were based on concomitant testing for HPV and cytology alone or for additional tests. Such investigations are known as split-sample studies because the cervical specimen collected in single or multiple exfoliative procedures using a swab, a cytobrush, or other collection devices is split into multiple subsamples for testing. Studies varied in terms of timing of collection, collection method, or whether or not visual methods for cervical inspection were used as adjunct screening techniques.

### COMPARATIVE EFFICACY OF HPV TESTING AND PAP CYTOLOGY

Table 2 summarizes the findings from all of the studies shown in Table 1, in terms of screening performance indices of sensitivity and specificity. All estimates for HCII testing were

---

**Table 1.** Characteristics of studies comparing HPV testing with Pap cytology in primary screening for cervical cancer and its precursor lesions*  

<table>
<thead>
<tr>
<th>First author (reference No.)</th>
<th>Country, study site</th>
<th>Study size</th>
<th>Age, y</th>
<th>HPV test</th>
<th>Study features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cuzick (4)</td>
<td>U.K., London</td>
<td>2009</td>
<td>20–45</td>
<td>93% Type-specific PCR (16,18,31,33)</td>
<td>Women free of cytologic abnormalities at enrollment</td>
</tr>
<tr>
<td>Clavel (5)</td>
<td>France, Reims</td>
<td>1518</td>
<td>15–72</td>
<td>HCl HCII</td>
<td>Women free of cytologic abnormalities at enrollment, cross-sectional plus 15 mo of follow-up testing, HPV positivity alone not a criterion for immediate colposcopy referral</td>
</tr>
<tr>
<td>Cuzick (6)</td>
<td>U.K., London</td>
<td>2981</td>
<td>35+</td>
<td>HCl, HCII, MY09/11 PCR</td>
<td>Women free of cytologic abnormalities at enrollment</td>
</tr>
<tr>
<td>Kuhn (7)</td>
<td>South Africa, Cape Town</td>
<td>2944</td>
<td>35–65</td>
<td>HCl HCII</td>
<td>Unscreened population, community recruitment</td>
</tr>
<tr>
<td>Ratnam (8)</td>
<td>Canada, Newfoundland</td>
<td>2098</td>
<td>18–69</td>
<td>HCl HCII</td>
<td>Multiple screening practices, 10% random sample of Pap–HPV– women referred for colposcopy</td>
</tr>
<tr>
<td>Schiffman (9)</td>
<td>Costa Rica, Guanacaste</td>
<td>8554</td>
<td>18–90+</td>
<td>HCl, HCII</td>
<td>Population-based, HPV positivity not a criterion for colposcopy referral</td>
</tr>
<tr>
<td>Wright (10)</td>
<td>South Africa, Cape Town</td>
<td>1365</td>
<td>35–65</td>
<td>HCII</td>
<td>Subset of sample in Kuhn (7)</td>
</tr>
<tr>
<td>Womack (11)</td>
<td>Zimbabwe, Harare</td>
<td>481</td>
<td>25–55</td>
<td>HCII</td>
<td>Primary care clinics, no history of cancer at enrollment, 2 strata based on HIV serostatus, all women underwent colposcopy</td>
</tr>
<tr>
<td>Schneider (12)</td>
<td>Germany, East Thuringia</td>
<td>4761</td>
<td>18–70</td>
<td>GP5/6+ PCR</td>
<td>Multiple screening practices, cross-sectional plus 8 mo of follow-up testing</td>
</tr>
<tr>
<td>Belinson (13)</td>
<td>China, Shanxi</td>
<td>1997</td>
<td>35–45</td>
<td>HCII</td>
<td>Unscreened population, community recruitment, all women underwent colposcopy</td>
</tr>
<tr>
<td>Blumenthal (14)</td>
<td>Zimbabwe, Harare</td>
<td>2073</td>
<td>25–55</td>
<td>HCII</td>
<td>Completed recruitment (irrespective of HIV serostatus) of study in Womack (11) all women underwent colposcopy</td>
</tr>
<tr>
<td>Clavel (15)</td>
<td>France, Reims</td>
<td>7932</td>
<td>15–76</td>
<td>HCII</td>
<td>Same design as Clavel (5) with expanded patient accrual</td>
</tr>
<tr>
<td>Kulasingam (16)</td>
<td>United States, Washington State</td>
<td>4075</td>
<td>18–50</td>
<td>HCII, MY09/11 PCR</td>
<td>Family planning clinic recruitment, ThinPrep cytology, 41% random sample of Pap–HPV– women referred for colposcopy</td>
</tr>
</tbody>
</table>

*HC = Hybrid Capture; HPV = human papillomavirus; PCR = polymerase chain reaction.
Table 2. Estimates of screening performance indices from studies comparing HPV testing with Pap cytology* in primary screening for cervical cancer and its precursor lesions—estimates shown are for HSIL or cancers as disease outcome†

<table>
<thead>
<tr>
<th>First author (reference No.)</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HPV Pap</td>
<td>HPV Pap</td>
<td></td>
</tr>
<tr>
<td>Cuzick (4)</td>
<td>75 46</td>
<td>96 96</td>
<td>Recalculated for enrollment screening results</td>
</tr>
<tr>
<td>Clavel (5)</td>
<td>100 79</td>
<td>86 96</td>
<td></td>
</tr>
<tr>
<td>Cuzick (6)</td>
<td>95 79</td>
<td>94 99</td>
<td>HPV indices based on HCII (n = 1703)</td>
</tr>
<tr>
<td>Kuhn (7)</td>
<td>88 78</td>
<td>82 97</td>
<td>HPV indices based on PCR (n = 2988)</td>
</tr>
<tr>
<td>Ratnam (8)</td>
<td>68 27</td>
<td>91 96</td>
<td>LSIL in histology excluded, HPV indices based on HCII (n = 424)</td>
</tr>
<tr>
<td>Schiffman (9)</td>
<td>88 78</td>
<td>89 94</td>
<td>LSIL in histology excluded, HPV indices based on HCl (n = 2861)</td>
</tr>
<tr>
<td>Wright (10)</td>
<td>84 61</td>
<td>83 96</td>
<td>All ages, bias-adjusted specificity includes CIN 1</td>
</tr>
<tr>
<td>WOMACK (11)</td>
<td>66</td>
<td>81</td>
<td>30+ y, bias-adjusted, specificity includes CIN 1</td>
</tr>
<tr>
<td>Schneider (12)</td>
<td>89 20</td>
<td>94 99</td>
<td>Conventional cytology and HCII all ages</td>
</tr>
<tr>
<td>Belinson (13)</td>
<td>95 87</td>
<td>85 94</td>
<td>HCII, ages 18–30 y</td>
</tr>
<tr>
<td>Blumenthal (14)</td>
<td>80 44</td>
<td>61 91</td>
<td>HCII, ages 31–40 y</td>
</tr>
<tr>
<td>Clavel (15)</td>
<td>100 68</td>
<td>87 95</td>
<td>HCII, age &gt;40 y</td>
</tr>
<tr>
<td>Kulasingam (16)</td>
<td>91/88 57</td>
<td>73/79 90</td>
<td>Clinician collected HPV sample</td>
</tr>
<tr>
<td></td>
<td>74/70 46</td>
<td>71/78 89</td>
<td>Self-sampling for HPV</td>
</tr>
<tr>
<td></td>
<td>63/57 36</td>
<td>83/87 96</td>
<td>Bias-controlled‡</td>
</tr>
</tbody>
</table>

*LSIL threshold (majority of studies) or LSIL or persistent ASCUS [Kulasingam et al. (16)]. Unless otherwise indicated, results are based on conventional cytology.
†ASCUS = atypical squamous cells of undetermined significance; CIN = cervical intraepithelial neoplasia; HC = Hybrid Capture; HIV = human immunodeficiency virus; HPV = human papillomavirus; HSIL = high-grade squamous intraepithelial lesion; LSIL = low-grade squamous intraepithelial lesion; PCR = polymerase chain reaction.
‡Verification bias: bias-controlled denotes verification of disease status among all participants, and bias-adjusted denotes correction of estimates based on the verification of disease status in a random sample of test-negative women. See text for details.

Based on the manufacturer’s recommended threshold of positivity of 1 pg/mL (equivalent to 5000 viral copies), whereas those from HCl were based on a detectability threshold that was approximately 10-fold higher. Some studies presented information for different age groups or also included the performance estimates for liquid-based cytology using ThinPrep (Cytex, Boxborough, MA), the thin-layer cytology smear system approved by the FDA in 1996 as a replacement for the conventional Pap test.

Taking into account all of the paired comparisons between Pap cytology and HPV testing (with PCR or HCII) shown in Table 2, the latter test has, on average (unweighted by study size), 27% higher sensitivity than cytology in absolute terms but somewhat lower specificity, i.e., on average, 8.4% lower for detecting high-grade lesions. The equivalent comparisons based on results using HCII yielded differences of 18.6% for sensitivity and 12.6% for specificity. Irrespective of combination or study restriction, the average sensitivity of Pap cytology was 60% and its specificity was 95%. The equivalent estimates for HPV testing were 85% and 84%, respectively. The screening of women aged 30 years or older or 35 years or older tended to improve the performance of HPV testing because viral infections in this age group are less likely to be of a transient nature than in the younger women. The average sensitivity and specificity estimates for study combinations, including these women, were 89% and 90%, respectively. However, a comparable net increase was also seen with conventional cytology, most likely for the same reasons, i.e., low-grade squamous intraepithelial lesion (LSIL) that decreases in prevalence relative to HSIL among the older women. The above interpretation of differences between tests is not affected by restricting the comparison only to one entry per study or by whether or not conventional or liquid-based cytology serves as the reference.

**Verification Bias**

It is important to note that the sensitivity and specificity estimates of many studies shown in Table 2 are relative, not absolute, because of verification bias. The latter occurs whenever the probability of disease verification via the gold standard is dependent on the screening test result. In general, such studies used a design in which only women with one or more positive screening tests were referred for colposcopy and biopsy, which prevented the unbiased estimation of absolute sensitivity and specificity (their estimates should be considered to be relative). These studies relied on the fact that, with two or more tests, there were always combinations of either Pap-negative or HPV-negative women with verified disease status available for analy-
sis. However, the biasing effects of the unequal verification of disease status can be strong and may lead to estimates of screening efficacy that cannot be generalized for cost considerations and other public health uses (17). Such verification bias was either averted in the Zimbabwe (11) and in the Chinese (13) studies (all of the women underwent colposcopy by design) or adjusted for in the Canadian (8), in the German (12), and in the U.S. (16) studies (the screening results from a random fraction of women with negative screening tests were extrapolated to those without colposcopic verification). A discussion of verification bias and other statistical pitfalls of HPV screening studies has been presented elsewhere (17).

ISSUES IN DEFINING THE DIAGNOSTIC GOLD STANDARD OF CERVICAL NEOPLASIA

It is noteworthy that the issue of verification bias is not as simple as the above discussion indicates. The previously mentioned studies that either avoided or corrected for the putative bias assumed that a colposcopy-guided biopsy accurately reveals the existence of cervical lesional tissue, which was then used to ascertain the distribution of diseased and nondiseased women to allow the computation of adjusted estimates of screening efficacy. While the approach is correct for its intended purpose, i.e., to obtain an improved estimate of the distribution of disease conditional on test results, it should be recognized that a simple colposcopy or even a colposcopy-guided biopsy cannot provide a guarantee that a lesion will be detected. In many test-negative women, the colposcopist cannot visualize lesional tissue and may decide that the colposcopic impression of no disease alone serves as the definitive diagnosis. However, this would be a mistaken decision because a lesion could be hidden in the endocervical canal and not be visible. Although this pitfall could be minimized by adopting a colposcopy protocol in which blind biopsy specimens are collected, it is still possible that a fraction of the existing lesions will remain undetected, because of either their location or their size. Preliminary results from the Atypical Squamous Cells of Undetermined Significance/Low-Grade Squamous Intraepithelial Lesions Triage Study (ALTS) and the Guanacaste cohort study indicate that a colposcopy-guided biopsy misses about one third of the CIN3 lesions that are present or develop within 2 years (Schiffman M: personal communication). Therefore, in any cross-sectional survey of screening efficacy, the ethically acceptable gold standard for cervical lesions (colposcopy-guided biopsies) is an imperfect one because of the inadequate sampling of the entire cervical tissue that is at risk of squamous cell carcinoma. One could speculate that only a more aggressive diagnostic approach, such as a detailed histologic examination of serial sections from cone biopsies or from specimens collected via the loop electrosurgical excision procedure (LEEP) or via the large loop excision of the transformation zone, would approach the definition of an acceptable gold standard of disease. Adopting such an approach, even in a sample of test-negative women, would be not only unethical but also impractical.

Even if tissue sampling could be done optimally with respect to the lesion site and the time of development, one also needs to consider the misclassification of lesion outcome status that exists, even with histopathologic ascertainment. Findings from a study (18) that involved multiple expert pathologists indicated that the reproducibility in grading histopathology specimens is not high, even with large specimens such as LEEP-obtained tissue samples. Therefore, a study that is simply based on lesion ascertainment provided by a single expert pathologist will be more prone to lesion misclassification than one employing a panel of readers that reaches a consensus diagnosis in every case.

Furthermore, as the design of screening efficacy studies evolves from the traditional single-opportunity sampling, cross-sectional layout to long-term, repeated sampling investigations over many years, disease case definition becomes a more dynamic process, requiring the juxtapositioning of screening and diagnostic test results obtained from multiple samples collected over time. This process involves combining the results from different diagnostic approaches, which may be differentially triggered by the severity of the lesion grade presumed by the test (HPV or cytology), e.g., colposcopy with simple biopsy for equivocal or low-grade lesions and LEEP for high-grade lesions. Ongoing natural history investigations of HPV and cervical neoplasia, such as the National Institutes of Health’s Guanacaste study or the Ludwig–McGill cohort, are examples of studies that have to grapple with this added complexity by having to differentiate between prevalent and incident lesions, progression and regression, and relating them to screening test performance. Calculation of sensitivity and specificity in these studies involves the combination of diagnostic information over multiple samples, which greatly minimizes the chance that any lesions are missed by the pitfalls described above for a cross-sectional study relying on colposcopy-guided biopsies alone. On the other hand, the repeated sampling layout of these investigations obviates the need for invasive diagnostic procedures among women testing consistently negative for both HPV and Pap tests over many visits. The longitudinal nature of the investigation ends up providing the test and diagnostic data that approach the true distribution of disease dynamics conditional on study duration. Therefore, correction for verification bias is not a critical issue in these longitudinal studies, with intensive follow-up of test-negative women and repeated histologic sampling of test-positive cases. However, these studies do have to contend with the issue of distinguishing between the prevalent and incident lesions to properly assign the distribution of disease for the purposes of gauging screening test efficacy.

HPV TESTING AS AN ADJUNCT TO PAP CYTOLOGY SCREENING

An important finding of some of the studies summarized in Tables 1 and 2 was the realization that the combination of cytology and HPV testing attained very high sensitivity and negative predictive values (approaching 100%) (8,9,13). A testing combination with such a high-negative predictive value could potentially allow increasing screening intervals safely, e.g., from 1–3 to 3–5 years, depending on the population and risk profile. The drawback of this approach is the loss in specificity with respect to either test in isolation because of the excessive number of patients who would need to be referred for colposcopy, many of which will turn out to be false-positive results. Resource-rich countries can absorb the extra costs related to the secondary triage of cases that will be referred via a dual-testing screening approach because this strategy may be cost-saving on long-term assessment, via the reduced patient flow for primary screening clinics. Economic models based on valid estimates of screening efficacy across different settings are urgently needed to assess
the potential benefit of combined screening in relation to its costs.

With HPV testing assumed as an adjunct to Pap cytology, women with either a positive cytology or a positive HPV test are referred for colposcopy. This scenario should not be confused with reflexive HPV testing that is considered in triage of atypical squamous cells of undetermined significance (ASCUS) smears. In the latter, the HPV testing information indicates the most likely clinical risk category to which a woman would be assigned for management purposes, whereas in the former (true primary screening), the HPV test supplements the information obtained by cytology. Using this adjunctive testing approach, the Canadian study (8) found a relative sensitivity of 100% (95% confidence interval [CI] = 89 to 100) for high-grade lesions at a cytologic threshold of LSIL, with a negative predictive value of 100% (95% CI = 91 to 100), with the HPV assay being the primary contributor to the high-joint sensitivity of the testing combination as shown in Table 2. Similar results were also seen in the Costa Rican study (9) and in the U.S. study (16). A large follow-up study of the Portland, Oregon, cohort also indicated that having dual negative Pap and HPV tests at baseline was equated with negligible subsequent risk of CIN3 or cancer during the next 4 years, mostly due to the contribution of the HPV assay (19).

In addition to the previously mentioned caveat of an increased number of women who will be referred because of combined testing in cervical cancer screening, another methodologic issue requires our attention. A nominal increase in sensitivity always occurs by chance whenever an adjunct test (e.g., HPV) is used in parallel with a conventional test (e.g., Pap cytology), even if the new tests were totally random with respect to the disease being evaluated. This increase in sensitivity can be misleading, even if it is deemed to be statistically significant. Combined testing prevents a loss in specificity but may offer no real sensitivity gain in certain conditions (20). An empirically valuable adjunct test, such as the HPV assay, should complement the Pap cytology so that the net combined sensitivity and specificity will be truly superior to those of the cytology alone (21).

**Other Studies**

A few large RCTs of HPV testing in primary cervical cancer screening are currently ongoing. Of note are the United Kingdom’s HPV in Addition to Routine Testing (HART) investigation (22), the United Kingdom’s A Randomized Trial in Screening to Improve Cytology (ARTISTIC) (Kitchener H: personal communication), and the Canadian Cervical Cancer Screening Study (CCaST), led by the author. In addition, a large demonstration project of HPV testing has been ongoing in The Netherlands for several years. The definitive results from these investigations will likely be available within 2–5 years.

The HART study began in 2001 and has recruited more than 11,000 women aged 30–60 years in five centers. All of the women with mild dysplasias or worse lesions on cytology are referred for colposcopy, whereas those with equivocal or borderline abnormalities or with a positive HCII result (but cytologically normal) are randomly assigned to receive an immediate colposcopy or a 6-month follow-up by HCII and cytology. Although strictly speaking, the HART study is best described as a secondary triage investigation, it assesses the value of HPV testing as a procedure immediately after an initial primary screen by cytology, and thus its results will contribute to the knowledge base concerning the use of HPV testing in cervical cancer screening in the future. Preliminary results from the HART study indicate a sensitivity of 94% and a specificity of 93% for HPV and equivalent figures of 76% and 96%, respectively, for a borderline or worse cytology result.

The ARTISTIC trial will enroll 28,000 women aged 20–64 years who are attending routine cervical screening in the Manchester region. All of the women receive a Pap test (ThinPrep) but will be randomly assigned at a ratio of 3:1 to the HPV result being revealed (study arm) or concealed (control arm). All of the women will have Pap and HPV tests at 36 months after entry and will be followed up by record linkage for at least 6 years. Management will be according to standard U.K. guidelines, but knowledge of the HPV test results in the study arm will guide decisions via an algorithm that is contingent on the grade of the cytologic abnormality and the timing of the follow-up. Psychological outcomes and detailed health economic evaluation are important aspects of this trial.

The CCaST study will enroll 12,000 women between 30 and 69 years of age from multiple centers in Quebec and in Newfoundland. Women will be enrolled to have both an HPV test and a Pap smear but will be randomly assigned as to the order in which these tests will be done. All women with an ASCUS Pap smear or a positive HPV test will undergo colposcopy and biopsy, as will a random sample with negative tests. Women with normal colposcopy and biopsy specimens will undergo a repeat colposcopy 6 months later. Women with both negative screening tests will be rescreened 1 year later. Screening performance will be estimated with respect to which test was done first (the index test). All instances of HSILs detected by the second test but not by the index test will be treated as false-negative cases. Having lesion information on a random sample of women who were negative by either test will allow correcting for verification bias.

**Study Design and Ethical Issues**

Future studies should rely on the RCT design to be able to produce the weight of evidence that is needed for policy recommendations. Although adequate to produce proof of principle in test comparisons, split-sample testing designs, such as those of studies shown in Tables 1 and 2, do not allow the assessment of long-term outcomes and do not represent real screening conditions. A case in point is the contribution of the ALTIS trial (1), an RCT of triage options that brought a solid knowledge base for policy decisions concerning management of ASCUS smears. Split-sample testing designs are also prone to sampling interference, because exfoliated cells that are destined to HPV testing have to be obtained at the same time as the specimen that is layered onto a glass smear for Pap testing. Although these problems can be minimized by randomly changing the order of the specimen collection, the final comparison will have been made under circumstances that are not the same as those in real screening conditions when only one test is used. This problem is probably not serious, but the question will always exist in the absence of an RCT in which each testing methodology is treated separately.

Another critical element of HPV screening studies is the choice of outcome. For logistical reasons, primary screening studies have relied on a cross-sectional assessment of prevalent lesions to estimate the diagnostic efficacy parameters for the various individual tests or combinations thereof. Focusing on an early end point, such as histologically defined HSILs or (more
stringently) CIN3, is an adequate strategy to obtain the proof of principle that a given test may fare better than Pap cytology. However, this proof is indirect because it is based on the assumption that, by having enhanced sensitivity and at least comparable specificity with respect to the latter, HPV testing will end up detecting more lesions that are likely to progress to cancer than could be found in a conventional cytology screening program. If qualitative prognostic differences exist between the subsets of HSILs that are detected by HPV and Pap cytology in terms of their ability to progress to invasive cervical cancer, it is conceivable that the gain in screening yield may not translate into reduced cancer incidence if a program were to adopt HPV testing in lieu of Pap cytology. Consequently, and apart from cost considerations, the most conservative proof that HPV testing may truly represent an improvement as testing technology over existing Pap cytology programs can only be obtained by demonstration of a reduction in invasive cervical cancer. This would of course require much larger trials because of the rarity of invasive lesions relative to their preinvasive stages. In addition, the need for more downstream end points, such as microinvasive and early invasive cervical cancer, creates an obvious ethical dilemma, because RCTs, with active monitoring of lesion development, cannot be conducted without safeguards to identify and to treat all of the prevalent and incident HSILs. The solution to this impasse may eventually come from effectiveness trials or demonstration projects using HSILs as the primary outcome and with subsequent passive follow-up via record linkage with tumor registries and mortality databases. Meanwhile, however, while the studies described above proceed, we are likely to see a gradual adoption of HPV testing anyway because of the base of knowledge already available, even if it is not of the highest standard. An important ethical quandary is represented by the fact that Pap screening is considered to be the standard of care in most middle- and high-income countries. Study designs based on split-sample testing avoid the problem of omitting Pap cytology by having all of the women undergo all competing test strategies, including cytology and HPV. Consequently, subjects and ethical review boards are assured that the standard of care was provided to each and every woman enrolled in the study. The same does not apply to RCTs, the design proposed above providing better quality of evidence for deciding for or against a particular screening test. In most of the industrialized countries, it would be ethically inconceivable to propose a screening-efficacy RCT in which Pap cytology was not offered to all of the participants. The ethical argument could perhaps be made for studies in low-resource countries where Pap cytology is not available either as opportunistic or as organized screening. In this case, an investigator could safely assume that an RCT omitting Pap cytology in one or more arms would still be providing benefit (and causing no harm) because of the availability of competing screening strategies that are likely to be efficacious in detecting lesions.

If an efficacy trial were to be carried out in a setting where women have easy access to Pap cytology screening as the standard of care, all of its comparison arms will have to include the latter, and comparisons will be made by combining competing testing strategies as adjunct tests. Another solution would be to make Pap cytology available to all of the study arms but in unmasked form only in the one in which cytology is the sole screening modality. This strategy may work well in short-term studies in which minor delays in managing lesions will be of no consequence. However, in long-term studies, one would have to intervene on the basis of the masked result, which would prevent a direct comparison among tests because the differences in lesion-detection rates among arms would disappear. Designs that maintain Pap cytology in all of the comparison arms need to compensate for the detection bias that occurs when HPV or other competing tests are added to cytology (17).

**Needed Research Directions**

A relevant goal for HPV screening studies is the evaluation of efficacy and cost-effectiveness of multiple HPV testing modalities and sampling strategies. Most of the published studies have relied on the HC assay, which is the only test approved by the FDA. Other competing HPV testing systems based on PCR or other techniques have been used only in research and may or may not soon become available commercially. Little is known about their performance in screening conditions. Also desirable are studies that evaluate strategies aimed at simplifying the screening process, e.g., self-sampling (10,23). These strategies will need to be assessed in both resource-rich and resource-poor settings, because the results may vary substantially as a consequence of differences in the prevalence of sexually transmitted diseases and cultural differences in accepting active participation in the screening process as a patient.

As with any cancer screening strategy, efficacy and effectiveness are directly related to the chosen management and follow-up algorithms to triage test-positive women. Although they vary in different settings, a national (24) and an international (25) consensus exists with respect to such options for cases detected by Pap cytology, but no agreed approaches exist for managing women with positive HPV test results in the absence of cytologic abnormalities. In resource-poor settings, the colposcopy–biopsy paradigm to assess the efficacy of screening approaches may not exist altogether. In such settings, it would be desirable to conduct demonstration projects examining the cost-effectiveness of management approaches that are independent of cytology–colposcopy ascertainment to monitor the occurrence of cervical lesion outcomes. For instance, one could gauge the comparative efficacy of HPV and other tests on the basis of the reduction in CIN3 or cancer incidence after using cryotherapy to treat lesions that were triaged by the competing methods and judged to be of high invasive potential on standardized visual inspection. Such a screen-and-treat approach is likely to have maximal benefit in settings where compliance is poor and no facilities or expertise exists for performing colposcopies and histology.

Will HPV testing serve as a true cervical cancer screening tool or will we use it as an adjunctive approach to make lesion risk predictions in programs where the central screening strategy continues to be cervical cytology? In addition to the ongoing research on HPV testing in screening, the results from long-term cohort studies of the natural history of HPV infection and cervical neoplasia have helped our understanding of the long-term predictive value of HPV tests (26–28). At present, the evidence, as summarized in this overview, indicates that HPV testing offers superior sensitivity for detecting prevalent HSILs than Pap cytology but lags behind in specificity as compared with the latter. Although restriction to older women seems to improve the specificity of the test, it also does so for cytology. Combined testing offers superior sensitivity and negative predictive values than cytology alone. These arguments have led the company
(Digene) that has the only commercially available HPV test (HC assay) to file an application to the FDA to enable it to distribute its assay as an adjunct to cytology in cervical cancer screening. Other competing proprietary technologies such as liquid-based cytology (ThinPrep) have been granted FDA approval and have taken a substantial market share in the United States and in other Western industrialized countries.

**PUBLIC HEALTH IMPLICATIONS**

Taken together, the results from the epidemiologic and screening studies and the directions already decided by health policymakers in many high- and middle-income countries indicate that we are currently moving from a Pap cytology-based paradigm of cervical cancer screening to one in which both cytology and HPV testing are used in combination for an augmented program sensitivity and negative predictive value, with consequent lengthening of screening intervals. In resource-rich countries, the cytology component of this new paradigm may also gradually move to one in which the conventional Pap test is replaced by a liquid-based cytology technique. In a few years, once there is accrued experience with the effectiveness of this combined testing approach, we may eventually witness a shift to a screening strategy based on HPV testing alone. In low-income countries, two mutually exclusive directions are envisaged, depending on the perceived importance of a local Pap test-based screening activity. In settings with no screening activity, it may be more cost-effective to implement a de novo program based on HPV testing, provided that the demonstration projects can provide the evidence that such a system may improve the detection of HSILs and cervical cancers without incurring an overburdening of the health care system and competition with other health priorities. On the other hand, the settings in which there is substantial dependence on cytology screening may want to reassess the effectiveness of their cervical cancer control programs in light of the availability of HPV testing. This will assist local health authorities in deciding whether the existing cytology-based program can be improved via the adoption of established World Health Organization (WHO) quality-control guidelines with acceptable costs or whether it would be preferable to shift to an HPV-based strategy, either alone or in an adjunctive form with cytology.

Central to this discussion is the fear by public health stakeholders, particularly outside the United States, that cervical cancer screening—which in the last half century was based on a technique in the public domain, the Pap test—may eventually become the province of a few commercial monopolies. This is the present situation with HPV testing, although this may change in the future as this screening approach becomes more widely accepted and other biotechnology companies sense the opportunity for making market inroads. Local government or the WHO regulation may become imperative after the existing HPV test producer loses its patent protection. Although the status quo of a single test maker (Digene) is clearly undesirable in many countries because of the commercial dependence, it guarantees that all HPV testing is of consistently high quality. In an unregulated market environment with multiple HPV test providers, the efficacy of HPV-based screening may eventually falter, bringing back the conditions that prevail today in settings with ineffective Pap cytology screening. The plausibility of this scenario will gradually increase as HPV testing makes the transition from a research tool to an established and widely adopted medical procedure.

Contributing to the uncertainty in forecasting the role of HPV testing in cervical cancer screening is the fact that the technology has evolved faster than it has been assessed clinically and epidemiologically. The previously mentioned move from HCl to HCII was based not only on the grounds of better immunoassay technology and chromogen chemistry but also on the grounds of a simpler and more reproducible microassay platform for multiple volume testing. Ongoing research on HPV assay designs is likely to produce even simpler testing formats, with the ultimate goal of perfecting a rapid technique that could be completed in fewer reaction steps and while the patient waits.

Public health authorities in middle- and low-income countries have monitored closely the ongoing debate on the role of new screening technologies. Between the fear of increased health care costs consequent to the adoption of such new screening tests and the promising results coming from the research front on HPV vaccines, it is all too tempting to take a wait-and-see attitude concerning cervical cancer prevention. It is not unreasonable to consider that this posture could become widespread in developing countries and lead to decreased funding of cytology quality-control programs, training, and needs assessment, in the false hope that HPV vaccines will soon be available to reduce the burden of cervical cancer, thus obviating the need for substantial expenditures implementing and maintaining screening programs. This scenario would no doubt prove to be disastrous.

It is necessary for us, as clinical scientists assessing the role of new technologies for cervical cancer, to place our work in a broader, more useful public health context where competing priorities come in many varieties and degrees of urgency. It is paramount in this respect that societal variables be incorporated in efficacy and cost-effectiveness studies so that the right lessons will be learned in each instance. Producing and sifting through the evidence on the role of HPV testing and other emerging technologies for cervical cancer screening require complex mobilization of resources by multidisciplinary investigations. The most definitive answers to come from these studies will take at least a decade to be obtained. Therefore, we need to alert stakeholders that cytology-based programs are not to be relaxed or we are at risk of wiping out the gains from the past 50 years in cervical cancer prevention.

**REFERENCES**


NOTES

Supported by grants from the National Institutes of Health, Department of Health and Human Services; and by the Canadian Institutes of Health Research. E. L. Franco is the recipient of a Distinguished Scientist award from the latter agency and of a National Scholar award from the Fonds de la recherche en santé du Québec.

E. L. Franco has conducted research that uses the commercial human papillomavirus assay manufactured by Digene, Inc.

I thank the monograph editors and the reviewers for their many excellent suggestions to my earlier drafts of the manuscript.