UNIFORMED SERVICES UNIVERSITY OF THE HEALTH SCIENCES
4301 JONES BRIDGE ROAD
BETHESDA, MARYLAND 20814-4799

GRADUATE EDUCATION
(301) 295-3913
FAX (301) 295-6772

APPROVAL SHEET

Title of Dissertation: "Evaluation of Cervicography Screening for Cervical Cancer in a High Risk Population"

Name of Candidate: Diana Schneider
Doctor of Public Health Degree
16 February 2000

Dissertation and Abstract Approved:

22 March 00
David Cruess, Ph.D. — Date
Department of Preventive Medicine and Biometrics
Committee Chairperson

8 Feb 2000
Heidi Friedman, Ph.D. — Date
Department of Preventive Medicine and Biometrics
Committee Member

Date
CAPT Gerald Quinnan
Department of Preventive Medicine anda Biometrics
Committee Member

08 Feb 2000
CAPT William Haffner — Date
Department of OB/GYN
Committee Member

8 Feb 2000
Mark Schiffman, M.D. — Date
National Cancer Institute
Committee Member

The author hereby certifies that the use of any copyrighted material in the thesis manuscript entitled:

"Evaluation of Cervicography Screening for Cervical Cancer in a High-Risk Population

beyond brief excerpts is with the permission of the copyright owner, and will save and hold harmless the Uniformed Services University of the Health Sciences from any damage which may arise from such copyright violations.

Diana Schneider

Diana Schneider
Department of Preventive Medicine & Biometrics
Uniformed Services University of the Health Sciences

Abstract

Evaluation of Cervicography Screening for Cervical Cancer in a High-Risk Population

Diana L. Schneider, Dr. P.H., 2000

Dissertation directed by Heidi B. Friedman, Ph.D., David Cruess, Ph.D., and Gerald Quinnan, M.D., Department of Preventive Medicine and Biometrics, USUHS; William Haffner, M.D., Department of Obstetrics and Gynecology, USUHS; Mark Schiffman, M.D., M.P.H., National Cancer Institute

Statement of the problem: Cervicography™ was first described in 1981 as a visual screening system for early detection of cervical neoplasia and cancer. Early studies to assess the validity of cervicography showed the method to have an acceptable sensitivity but an unacceptably low specificity for mass screening. Following revision of the cervicography classification scheme, specificity improved, but at the expense of lowered sensitivity. Most previously published studies have had some methodologic inadequacies which may have affected the outcome.

Methods: Cervigrams were taken for 8460 women who enrolled into a population-based, natural history study of cervical neoplasia in Guanacaste Province, Costa Rica. Cervicography and three cytologic screening tests were the basis for referral for colposcopic examination and directed biopsy. Initial cervicography classification was compared with a referent diagnosis determined by histology and three cytologic tests, cytology, and presence of cancer-associated human papillomavirus types. Cervicography was submitted to additional review and arbitration to achieve an optimal classification. Interobserver agreement was assessed, and the performance of the optimal cervicography

result was compared with the referent diagnosis. Sensitivity, specificity, and predictive values were estimated, and results were stratified by characteristics of the woman and visual characteristics of the cervigram image. Digital colposcopic images were interpreted to evaluate the perceived appropriateness of the decision to biopsy and biopsy placement, and the impact of these on sensitivity and specificity.

Results: Moderate agreement on cervigram classification was observed (*kappa*=0.47 when cervigram results were classified into seven categories and 0.54 when cervigram results were classified into dichotomous categories of referred for colposcopy versus not referred). For the detection of high grade squamous intraepithelial lesion or cancer, optimized cervicography yielded a sensitivity of 55.2% and a specificity of 94.3%, which was only slightly improved over the initial estimates of 49.3% sensitivity and 95.0% specificity at enrollment. Higher sensitivity was associated with younger age, premenopausal status, the presence of metaplasia, the absence of cervicovaginal atrophy, and improved quality of the acetic acid effect.

Conclusions: Evaluator agreement with cervicography is moderate. The arbitrated cervigram classification improved the performance of cervicography only slightly over a single interpretation. Cytology performed better than cervicography for the detection of high grade squamous intraepithelial lesions, but the two methods performed similarly for the detection of invasive cervical cancer. Cervicography is not recommended for postmenopausal women and/or women ages 50 and older.

EVALUATION OF CERVICOGRAPHY SCREENING FOR CERVICAL CANCER IN A HIGH-RISK POPULATION

by

Diana Louise Schneider

Dissertation submitted to the Faculty of the Department of Preventive Medicine and Biometrics Graduate Program of the Uniformed Services University of the Health Sciences in partial fulfillment of the requirements for the degree of

Doctor of Public Health 2000

Table of Contents

List of Tables

Introduction and Review of the Literature

First Manuscript

Second Manuscript

List of Figures

Introduction and Review of the Literature

Overall Discussion

Appendix

Cervicography Screening for Cervical Cancer

Introduction and Review of the Literature

Diana L. Schneider

Introduction

The present research project was conducted to provide a rigorous and independent evaluation of Cervicography™ [National Testing Laboratories Worldwide (NTL), Fenton, MO] as a primary screening method for early identification and prevention of cervical cancer. Cervicography is a visual screening method which evaluates the macroscopic appearance of the cervix rather than the cytologic appearance of exfoliated cells [as with the more commonly used Papanicolaou (Pap) smear]. In cervicography, a trained health care provider or technician takes two high resolution photographic images (Cervigrams™) of the cervix after applying a 5% acetic acid wash. Cervigrams are interpreted by expert evaluators who classify the images using modified colposcopic criteria[1,2] as the basis for referral for colposcopy. Colposcopy is a microscopic examination of the cervix performed by a trained colposcopist using a specially designed instrument called a colposcope. (The colposcopy examination is described in greater detail below, under *Screening for Cervical Cancer*). Magnification is achieved in cervicography by projection of the 35mm slide (produced from film developed specifically for cervigrams) onto a screen. The magnification achieved depends on the width of the screen and the distance between the evaluator and the screen.[3] The evaluator may move closer to the screen (as close as one foot from the screen) to allow him/her to view the image with greater detail (due to the apparent increase in magnification caused by moving closer). At other times, the evaluator may view the image from 4-5 feet away. The viewing distance depends on the particular aspect of the evaluation that the evaluator is trying to achieve.[4] Appendix A provides the formulas used to calculate the

magnification and apparent magnification achieved from projecting the cervigram slide on the screen.

Cervicography may have a particularly important application in countries where specialized medical and laboratory expertise is concentrated in a few urban areas and/or where Papanicolaou reliability is problematic. Cervicography allows a wide variety of health professionals to take photographic images of the cervix in a standardized, controlled fashion. The film may then be sent elsewhere for developing and rapid, expert interpretation. A unique advantage of cervicography is that it provides an objective and permanent form of documentation of the visual appearance of the cervix.

This study is designed to assess the performance of cervicography screening, while overcoming some of the deficiencies of earlier studies of cervicography. The study evaluates the null hypothesis that cervicography does not perform adequately compared with the current standard of care, the Papanicolaou smear, for the detection of cervical cancer or serious cancer precursors (high grade squamous intraepithelial lesions), against an alternative hypothesis that cervicography performs equivalently or better than the Papanicolaou smear. Screening test performance is measured using standard contingency table methods (see *Evaluating a Screening Test* in this chapter) that compare the screening tests with a referent standard based primarily on histology. Performance estimates that are calculated include sensitivity, specificity, percent referred, positive predictive value, negative predictive value, and likelihood ratios. Additionally, interobserver agreement is evaluated using the *kappa* statistic. The difference between the classification provided by each of two cervigram evaluators, and between referral by

cervicography and conventional cytology, by the referent diagnosis, are directly compared using contingency table methods for paired data.

Previous studies have shown mixed results and were limited by their inclusion of selected populations; small sample size; and earlier, non-specific criteria for diagnostic classification.[5-16] This evaluation of cervicography was conducted as part of a large, population-based cohort study of the natural history of cervical neoplasia in the province of Guanacaste, Costa Rica, sponsored by the National Cancer Institute. The Guanacaste site was selected because of its consistently high age-adjusted rates of cervical cancer despite existing Papanicolaou screening services.[17]

Role of the Candidate

Some components of this study were underway before the candidate began her work on this study, while others were conducted independently by the candidate. At the time her participation began, all women had been selected and enrolled into the followup study (enrollment phase), cervigrams had been taken and interpreted by the initial cervigram evaluator, all cytology specimens had been collected and tested, specimens had been collected for human papillomavirus (HPV) testing, HPV testing had been performed using the first generation Hybrid Capture Tube Test, colposcopic examinations had been performed, digital colposcopic images were taken, and pathologic materials were collected, and a referent diagnosis was assigned. The candidate performed all statistical analyses, interpreted data, and wrote the first manuscript. She designed and managed the second phase of the study involving reviews of cervigrams and digital colposcopic

images. She selected women whose cervigrams and digital colposcopic images were included in reviews, prepared these materials for evaluation, designed data collection forms, interacted with expert evaluators, oversaw the data collection process, performed statistical analyses, and wrote the second manuscript. Additionally, the candidate independently researched and wrote the introduction and review of the literature. Finally, she interpreted all study data, developed conclusions, and synthesized these in the writing of the overall discussion.

Review of the Literature

Terminology

With the changing base of knowledge in the study of cervical carcinogenesis over the years, terminology used to describe advancing grades of cervical neoplasia has gone though several revisions. Table 1 outlines some of the terminology used to describe cervical abnormalities. The continuum of cervical abnormalities is often described as dysplasia, or cervical intraepithelial neoplasia (CIN). These terminologies are less commonly used in the U.S. since the implementation of the Bethesda system in 1988.[18] The terms low grade squamous intraepithelial lesions and high grade squamous intraepithelial lesions correspond to the Bethesda System of cytologic diagnosis. The Bethesda system was established to help provide consistency in the assignment of cytologic categories.[18,19] The terms CIN 1, CIN 2, and CIN 3 still correspond to the histologic classification system accepted by most pathologists. However, the Bethesda system terminology is sometimes used informally to describe histologic abnormalities.[20]

CIN terminology is also used to describe cytologic diagnoses, especially outside the United States, whereas the Bethesda system is most commonly used in the U.S. Changes in classification systems and in recommendations for colposcopic examination become apparent in reviewing the literature. Each study must be looked at within the context of the time the research was conducted, albeit with the appreciation that improvements in current knowledge may change the conclusions drawn during earlier time periods.

The Natural History of Cervical Neoplasia

Much has been learned about the natural history of cervical neoplasia during the past 15 or 20 years. Studies during the 1960s and 1970s correctly associated cervical cancer with a sexually transmitted agent.[21] It is now known that certain types of human papillomaviruses (HPV) cause most cases of cervical neoplasia.[22] The current list of HPV DNA types known to be associated with cervical cancer includes at least types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68.[23] However, approximately 50% of cases worldwide are associated with HPV type 16, 15% with type 18, 10% with type 45, and 5% with type 31.[24]

HPV DNA is present in nearly all cases of cervical cancer. However, infection with an oncogenic HPV type does not always lead to a diagnosed cervical abnormality. HPV infections are transient and may frequently disappear without detection. Factors associated with persistence and regression of HPV infection are not fully understood, but some possible explanations include the effects of HPV type, viral load, and immune response.[20,23] Lesions classified as condylomatous atypia and CIN 1 are common, and

they typically become cytologically apparent within months to one to two years of infection. These lesions usually regress spontaneously over months to a few years through a process thought to be controlled by immune responses. Current epidemiologic data suggest that condylomatous atypia and CIN 1 are too common and benign to be considered a precancer.[20,23] Using data from two large cohort studies, Schiffman et al.[20] suggest that the true incidence of low grade or equivocal cytologic lesions cannot be precisely determined because 1) the number of women infected with HPV is unknown, 2) the subtlest cytologic diagnoses are subjective, and 3) cytologic screening is performed too infrequently to accurately monitor the rapid cytologic changes in acute HPV infection. These investigators argue that the distinctions between HPV infection and cytologically evident lesions is unclear and not standardized, and therefore they propose using a broad diagnosis of HPV infection as a single diagnosis to include HPV infection and low grade lesions.[20]

Since high grade lesions are the immediate precursors to nearly all cervical carcinomas, understanding factors associated with progression to high grade lesions is most critical in preventing invasive cervical cancer.[20] Some women may test positive for cancer-associated HPV types without developing a diagnosable cervical lesion. In a small minority of HPV positive women with or without low grade lesions, infection may progress to CIN 2 or CIN 3, and even fewer will progress to invasive cancer. Lesions classified as CIN 2/3 appear to qualify as the immediate precursors to cancer. Throughout the spectrum of advancing cervical neoplasia, the risk of progression to a higher grade of abnormality appears to be related to persistence of infection.[20,23] Other

cofactors associated with progression to cancer are still unknown. Current epidemiologic research is looking particularly at the role of immunologic and behavioral cofactors as possible contributors to progression from HPV infection to CIN 2/3.[20] It is also believed that some HPV infections may become latent, rather than completely disappearing. There is evidence that in some women, latent infections may reactivate and progress quickly after decades of quiescence.[20] The most important risk factor for progression from CIN 2/3 to invasive cancer is age. Women with invasive cancer are on average ten years older than women with CIN 2/3, with no other notable known differences in epidemiologic risk factors.[20] The natural history study of cervical neoplasia currently being conducted in Guanacaste[17] will contribute to our understanding of the natural history pathway leading to high grade lesions.[20]

Historical Perspective of Cervical Cancer Screening

Cytologic sampling of the uterine cervix for diagnosis of cancer is widely believed to have been introduced in 1928 by George N. Papanicolaou. In 1924, Papanicolaou observed that cancer cells derived from the uterine cervix may be observed in vaginal smears, and he presented his findings in 1928.[25] It is less well known that Papanicolaou was not the first researcher to employ cytologic sampling of the cervix. Romanian pathologist, Aurel Babes, used the technique at least two years earlier than Papanicolaou, and he published his results in 1928.[25-27] In 1939, Papanicolaou began collaborating with Dr. Herbert Traut to study vaginal smears from women with malignant tumors of the uterine cervix, and they observed the presence of abnormal cells. They presented their

work in 1941, and later wrote up their findings in a book in 1943. Their observations were soon replicated by other investigators.[25] It was soon discovered that analysis of cytologic samples could be used to detect precancerous changes still confined to the epithelium of the cervix, eventually leading to its application for cancer detection, early treatment, and prevention.[25] Mass screening for cervical cancer using the cytologic smear (which became known as the Papanicolaou smear) was implemented in British Columbia in 1949 and in the United States by the mid-1950s.[28] Cytologic testing using the Papanicolaou gradually advanced to a widely accepted screening method without the benefit of rigorous clinical trials.

The Impact of Cytologic Screening

Cervical cancer is the third most common cancer among women worldwide, the second in developing countries, and the first in Central American nations.[29] This is a change from 1985 when cervical cancer was the second most common cancer among women worldwide. Much of this decline is due to apparently reduced incidence rates in China over recent years.[29,30] This may be influenced by the strict sexual mores of the Cultural Revolution. Cervical cancer incidence and mortality have markedly declined during recent decades in countries implementing widespread cytologic screening using the Papanicolaou smear.[31-33] Gustafsson and his colleagues studied trends in cervical cancer incidence before[34] and after[33] implementation of screening programs in 17 developed and developing nations. These investigators found that the age-specific incidence curves before widespread implementation of screening were similar among the

countries studied.[34] Using these data as a baseline, they compared the pre-screening incidence rates with rates from a later time period (i.e., after implementation of screening), allowing sufficient lag time since the initial implementation of screening programs to show an effect. Overall, age-standardized incidence rates of invasive cervical cancer decreased by at least 25 percent in 11 of the 17 populations studied. Only England and Scotland did not show a consistent decreasing trend in incidence rates between the two time periods. In all populations studied except Singapore and India, the reductions in age-standardized cervical cancer incidence rates from baseline compared with the followup period were larger for women ages 35–64 years than they were for populations including women of all ages. When including only the truncated populations of women ages of 35–64 (which allowed direct standardization to the world age distribution), all 17 populations studied experienced the decreasing trend in incidence rates, including England and Scotland. The authors concluded that these reductions were largely explained by screening, because the incidence remained stable in other populations not exposed to widespread screening.

Some studies have been conducted to evaluate cervical cancer screening programs in specific countries or regions. Two larger, recent studies relevant to Central America bear mention here. These studies also demonstrate the association of screening with reduced risk of cervical cancer, even given the potential for inaccuracy of a single Papanicolaou test. Herrero and colleagues[35] conducted a case control study of cervical cancer screening in four Latin American cities. They interviewed 759 cases and 1430 controls regarding screening patterns and demographic, reproductive, sexual,

contraceptive, medical, and dietary histories. An age-adjusted cervical cancer odds ratio of 2.5 (C.I. 2.1–3.3) was reported for women who never had a Papanicolaou smear compared with women with at least one lifetime Papanicolaou smear. This estimate was affected very little by adjustment for other risk factors. Women with only one lifetime Papanicolaou smear had a two-fold greater risk of developing cervical cancer compared with women with ten or more Papanicolaou smears (OR 2.2, C.I. 1.5–3.1), whereas women with two to nine Papanicolaou smears had no excess risk of cervical cancer. This suggests that in this population the first Papanicolaou smear offered some protection over never having had a Papanicolaou smear, and the second Papanicolaou smear offered even more protection. However, subsequent Papanicolaou smears did not result in reduced odds of developing cervical cancer.

Hernández-Avila and others[36] evaluated a cervical cancer screening program in Mexico using a case-control design. In their study of 397 cases of invasive cancer and 1005 population controls, they found an overall relative risk of 1.3 (C.I. 1.0–1.7) among women with no history of Papanicolaou smear as compared to women with a history of a Papanicolaou smear. Compared to women with a history of Papanicolaou smear but no history of gynecologic symptoms, women with no screening history had a 2.6-fold risk of cervical cancer (C.I. 1.9–3.6). Frequency of Papanicolaou smear screening was also associated with developing cervical cancer. Compared with women who had a Papanicolaou smear every 1–2 years, women screened every 3–4 years had a 1.3-fold risk (C.I. 0.7–2.3), women screened every 5–9 years had a 1.7-fold risk (C.I. 1.0–3.1), and women screened every 10 or more years had a 3.6-fold risk (C.I. 1.9–6.8). These results

showed a statistically significant increasing risk with length of screening interval ($p < 0.001$).

Although cytologic screening was never submitted to randomized clinical testing, it is so widely accepted as the appropriate standard of care that ethical considerations preclude the prospective followup of individual women to assess the impact of screening on cervical cancer incidence rates. One group of investigators employed a mathematical modeling technique to attempt to predict the association of screening on the incidence of cervical cancer based on the probabilities of various data inputs.[37] Their study focused on evaluating potential screening programs in developing countries. Assuming a paucity of resources, they modeled possible screening modalities involving a once lifetime screening, a ten-year screening interval, and a five-year screening interval. They based their model on women 15 years of age and performed a life-table analysis using the death rates for the female population of Brazil. Probability inputs included false-positive and false-negative outcomes for cytology and HPV DNA screening tests by grade of abnormality, screening participation rates, progression rates from HPV infection and/or CIN (grades 1, 2, and 3) to cancer, and regression rates. Some of the input data were from earlier sources for which more current data are now available. However, the authors performed a sensitivity analysis to account for variability in their data. The results showed that if once per lifetime screening were established with 80% coverage, a 23% reduction in the incidence of invasive cervical cancer would result from cytologic screening. If a ten-year screening interval were established with 50% coverage, a 30% reduction in the incidence of invasive cervical cancer would result from cytologic

screening. Variation in the rates of progression (among grades of CIN or of CIN to cancer) of CIN = 20% had little effect on the baseline results. If there were 50% more misclassified Papanicolaou smears (false positive or false negative), a smaller reduction in cervical cancer incidence would occur compared to the baseline predictions.

In summary, the studies described here, as well as others, lead us to the general conclusion that cytologic screening has contributed greatly to a reduction in cervical cancer incidence worldwide. However, as will be discussed, the Papanicolaou smear is prone to error, and as knowledge of the disease and technology has improved, alternative techniques have been introduced in attempts to achieve further gains in cervical cancer prevention.

Screening for Cervical Cancer

Cervical cancer screening traditionally is performed as a two tier system (Figure 1). In the first tier, women have a screening test performed (traditionally the Papanicolaou smear). Women with an abnormal screening test result are referred for colposcopic examination (second tier). The colposcopist performs a visual examination of the cervix using a high powered microscope combined with application of acetic acid and iodine stain to highlight lesions, if present. Colposcopy can achieve 6-40x magnification, though 10x enlargement is considered suitable for routine use.[2] Colposcopic evaluation involves the active participation of the colposcopist who may manipulate the cervix or mucus to view as much of the cervix as possible. He or she is also able to observe the uptake and fading of stain, which provides information about

cervical abnormalities. The use of green filters in colposcopy allows for enhanced visualization of vessels.[2] If a lesion is visualized, the colposcopist takes a biopsy of the most severe part of the lesion(s). More than one biopsy may be taken at this time. Histologic confirmation of the lesion is usually based on this method of colposcopy-directed biopsy, but histologic analysis may also be performed on tissue removed during excisional treatment of more severe lesions or during surgical procedures.

The use of acetic acid in colposcopic examination dates back to Hinselmann, a gynecologist who invented the first colposcope in 1925. Hinselmann used acetic acid to clean mucus from the cervix.[2] Acetic acid has the effect of enhancing colposcopic features by causing cells to swell. The acetowhitening effect may be due to light being reflected from the swollen epithelial cells,[38] although the reasons for this effect are not known with certainty. One team of investigators[38] examined acetowhite and nonacetowhite lesions for the presence of cytokeratins, since the presence of specific keratins varies with the pattern of differentiation in stratified squamous epithelium. They found that acetowhite lesions had a significant quantity of cytokeratin 10, whereas nonacetowhite lesions had minimal cytokeratin 10. Their findings suggest that the presence of cytokeratin 10 may be associated with the acetowhitening effect in cells.

After applying acetic acid, low grade lesions appear shiny or snow white and semitransparent, whereas high grade lesions appear with a dull, oyster white color.[1,39] Acetic acid causes epithelial lesions to become more distinct by accentuating their color and making various structures more distinguishable from each other. The dark red ectopic columnar epithelium becomes paler with shades of white and pink after

application of acetic acid. The grape-like structures of the columnar epithelium become more pronounced due to swelling and enlargement of the villi. Epithelial swelling caused by acetic acid gives a white appearance to abnormal epithelium and accentuates its surface contour. Patterns of mosaicism and punctation (abnormal blood vessel patterns often characteristic of high grade lesions) become more distinct. The effect of acetic acid on pathologic epithelium is not as rapid as on ectopic columnar epithelium. Therefore, observation by the colposcopist of the acetic acid uptake plays an important role in the colposcopic examination.[2]

The colposcopic examination is performed by a trained colposcopist, who is typically a gynecologist, family practice physician, or nurse practitioner. Colposcopy examination usually lasts 15-30 minutes and is not feasible as a primary screening method because of the long duration of the examination, the high cost of the colposcope, the high cost of the colposcopic examination, and the lack of sufficient numbers of trained colposcopists. Were these limitations not a concern, some studies have found that colposcopy would perform better than conventional cytology in a primary screening setting.[40-42]

Controversy still surrounds the issue of screening endpoints. The Bethesda system of cytologic classification was developed in 1988 in an attempt to clarify diagnostic terminology in cytopathology and to alleviate some of the confusion in the classification of atypical smears.[18,19] The Bethesda system categorizes results as unsatisfactory, normal, reactive changes, atypical squamous cells of undetermined significance (ASCUS), low grade squamous intraepithelial lesions (LSIL), high grade

squamous intraepithelial lesions (HSIL), microinvasive cancer, and invasive cancer. Cytology results of normal or reactive changes are not evaluated further. The LSIL category subsumes categories of koilocytotic atypia and CIN 1 (using the CIN classification scheme). The HSIL category includes the categories of CIN 2 and CIN 3. Until recently, most clinicians referred patients with any cytologic abnormality (ASCUS or more severe) for colposcopic examination. However, since the late 1980s, knowledge about the association of HPV and cervical cancer, and of the natural history of cervical neoplasia, has led some researchers and clinicians to rethink this approach. LSIL are now believed to be too common and likely to regress to be considered a "precancer".[20] Therefore, many clinicians now recommend colposcopic examination only for women with cytology results of HSIL or cancer, and continued followup for women with LSIL, unless it is anticipated that the patient may not return for regular followup.

Another area of difficulty is how to manage women with ASCUS cytology results and/or how to clarify that diagnosis.[43,44] The ASCUS diagnosis is reported to have poor reproducibility and expert pathologists often interpret it differently.[44] One recommendation has been made to subcategorize ASCUS results into those which are more likely to represent underlying HSIL and those which favor LSIL.[44] Women with cytologic diagnoses of equivocal/HSIL might be managed as HSIL because of the potential risk, whereas women with equivocal/LSIL results may require supplemental HPV testing to help guide a management plan.[44] A large, multicenter trial is currently underway to help resolve some of the questions surrounding these difficult cases.[45]

Evaluating a Screening Test

Several definitions are important to understanding how screening tests are evaluated. In assessing the performance of a screening test, "true disease" status is established by some referent test (i.e., the gold standard), which is presumably better at diagnosing disease than the screening test. The gold standard represents the state of the art for diagnosis. Screening tests are therefore evaluated against the existing gold standard. reference test. Table 2 illustrates the relationship between the screening test and the referent diagnosis, and provides formulas for calculating key estimates.

- Sensitivity refers to the proportion of women with "true disease" who are detected by the screening test.
- The false negative rate is 1 minus sensitivity, or, the proportion of women with true disease who are misclassified as disease-free by the screening test.
- Specificity refers to the proportion of women without true disease who are correctly identified as disease-free by the screening test.
- The false positive rate is 1 minus specificity, or, the proportion of women without " true disease" who are misclassified as positive by the screening test.
- Positive predictive value is the proportion of subjects who screen positive who have true disease.

- Negative predictive value is the proportion of women who test negative by the screening test who truly do not have the disease.

It would be ideal to have a screening test that is both highly sensitive and highly specific. Perfect sensitivity and specificity would each be 100%. In practice, screening tests will always fall short of this perfect mark. Additionally, there is usually a tradeoff between sensitivity and specificity. An increase in sensitivity is necessarily accompanied by a reduction in specificity by the same screening method, and vice-versa. However, there is no uniform standard to define acceptable sensitivity and specificity. Standards will vary by the disease in question and the evolving technology of screening tests. Acceptability is determined based on judgement about the natural history of the disease and consequences of different types of error (e.g., false negativity versus false positivity). Knowledge of the natural history of the disease is important to this judgement because if a diseased individual is missed by the screening test, conditions with a long preclinical phase may provide opportunities for detection by screening at a later time, but at a time at which treatment is still effective in reducing morbidity and mortality.

The obvious consequence of misclassifying individuals with true disease as negative is missed disease among individuals who are not detected by the screening test. Such error could lead to the individual being detected later at a more advanced stage of disease, and possibly to more serious sequelae that require more frequent treatment, more invasive treatments, greater cost of treatments, and greater likelihood of death. One consequence of misclassifying individuals without true disease as positive is that

additional testing is likely to be performed on truly disease-free individuals who are incorrectly identified as positive or of high risk of being positive by the screening test. Unnecessary testing on nondiseased individuals may overburden available resources, may result in financial and emotional costs associated with performing these tests, and may contribute to patients' fear that she or he may have a serious and possibly life-threatening illness.[46,47] For many clinical tests, some individuals fall into a group which is clearly classified as positive, and others fall into a group which is clearly classified as negative. Others, however, have results which are more ambiguous. In order to identify a greater proportion of individuals with true disease, some patients with in-between screening results who truly do not have disease will be misclassified as positive by the screening test. Decisions regarding the optimal cutpoint to define positive screening test results are made with the objective of achieving as high a sensitivity as possible while at the same time achieving an acceptable specificity.[46,47]

For cervical cancer screening, high sensitivity is essential, but high specificity is also desirable because insufficient resources are available to provide high volume colposcopic services. The limited availability of trained colposcopists and the high cost of colposcopic examination contribute to the need to limit colposcopic referral to women who are truly at high risk of developing cervical cancer. The median reimbursement cost of colposcopy and biopsy in the U.S. was estimated to be $428 in 1994.[48] The long process of carcinogenesis lends itself to screening at regular intervals, so that even if some sensitivity is sacrificed in order to increase specificity, screening at one to three year intervals (depending on the risk category of the woman) will allow for the detection of

some of the cases that were initially missed by the screening test. Cytology test results of ASCUS and LSIL are problematic as a cutpoint for cancer prevention because of low specificity. Newer screening tests for cervical cancer, including cervicography, are generally measured against the conventional Papanicolaou smear, the current standard of care. A successful alternate test should strive for equal or greater sensitivity than the Papanicolaou smear achieves, without a significant loss of relative specificity.

Another statistic employed in the evaluation of a screening test is the likelihood ratio.[49,50] The likelihood ratio is the ratio of the probability of having a given test result among diseased persons to the probability of having the same test result among nondiseased persons. Likelihood ratios may be calculated for each classification category of the screening test. Therefore, they avoid loss of information resulting from the grouping of results into dichotomous categories, i.e., as is required to calculate sensitivity, specificity, and predictive value.[50] Likelihood ratios are useful in the clinical management of patients by adding to clinician's predictive ability about their patients' condition. Likelihood ratios calculated from screening test results may be combined with information on patients' signs, symptoms, and risk profile, subsequently allowing clinicians to estimate their patients' probability of having the disease in question.[49-51]

Limitations of the Papanicolaou smear

Although the reductions in cervical cancer incidence point to the success of cytologic screening, it is not a perfect tool. False negative rates of the Papanicolaou smear compared to a gold standard (usually histologic) diagnosis have been reported between 0.02% to 99%.[28,52-62] Fahey and colleagues[62] conducted a meta-analysis of Papanicolaou smear accuracy using a summary receiver operating characteristic (ROC) curve analysis. In the studies reviewed, sensitivity ranged from 11 to 99 percent and specificity ranged from 14 to 97 percent. An unweighted ROC curve showed that a specificity of 90-95 percent corresponds to a sensitivity of 20 to 35 percent, indicating that the Papanicolaou smear is unable to achieve concurrently high sensitivity and specificity. Study characteristics were not found to be important predictors of Papanicolaou smear accuracy.

Optimal screening using the Papanicolaou smear requires strict clinical and laboratory protocols in obtaining the sample of cervical cells by a clinician, processing the sample, screening the smear by a cytotechnologist, and interpreting the smear by a cytopathologist.[25,28,63] Cervical cells must be collected from the squamocolumnar junction, the site of changing cellular composition (metaplasia) called the transformation zone,[2,64] and from the endocervical epithelium. The transformation zone is the site where most neoplasia occurs, and therefore is critical to cervical sampling.[2,64] Endocervical cells may be high up in the endocervical canal (especially in older women), and must be sampled using an appropriate device. Even with proper collection instruments, endocervical cells

will be missed in 2–4% of samples.[28] Additionally, smears must be fixed rapidly in the clinic and stained properly in the laboratory. Variability in environmental conditions, especially in tropical climates, may result in air-drying artifacts.[25] The interpretation of cytology slides requires careful screening by a trained cytotechnologist and subsequent evaluation by a cytopathologist. Appropriate screening requires the tedious task of viewing every cell and marking any abnormality. The average Papanicolaou smear contains between 50,000 and 300,000 cells.[25] In an effort to reduce screening errors, laboratory accreditation guidelines state that a cytotechnologist should not review more than 100 slides per 24 hour period.[65] Other factors that may affect Papanicolaou smear interpretation include obscuring by mucus, blood, inflammatory exudate, and/or necrotic debris covering the lesion.[66]

In addition to the questionable accuracy of cytologic screening, the Papanicolaou smear diagnosis has also been reported to suffer from poor reproducibility. Several studies have indicated that agreement on overall and category-specific diagnoses, and on specimen adequacy, is poor to moderate in some studies[19,61,67-72] and moderate to good in others.[69,72-76] This varying degree of reproducibility has been shown for both inter-[61,67-70,72,73,76] and intra-observer[67-69,71,73-76] agreement, among cytotechnologists[61,67-69,72,74,75] and cytopathologists[67,70-74,76] and for laboratories using the dysplasia[61,67-70,74-76] and Bethesda[71-73] classification systems. The ASCUS category used in the Bethesda System, which reflects a wide spectrum of cytologic changes, has shown especially poor reproducibility.[19,44,72,77]

In summary, many studies have demonstrated some of the limitations of Papanicolaou smear screening. The poor accuracy and reproducibility of the test leave

room for improvement by the implementation of quality assurance programs for laboratories or by the introduction of other cytologic techniques and/or alternate screening methods. Several attempts at improving cervical cancer screening have been implemented. Guidelines and quality assurance programs are in use and are intended to reduce laboratory error.[65,66,78-81] New cytologic techniques (e.g., TriPath, Elon, NC; ThinPrep, Cytyc Corporation, Boxborough, MA) are being used in some laboratories and are still undergoing evaluation on their validity.[82-91] In addition to the newer cytologic techniques, visual methods, including cervicography, speculoscopy, and aided visualization (each described in greater detail below) have been investigated for over a decade, and are in use (often as an adjunct to cytologic screening) in developed and developing countries, though to a lesser extent than cytology. HPV DNA testing is probably not suitable for general screening unless the cost is reduced, but a role will likely be developed for its use in the screening or triage of certain subgroups of women.[92,93]

Cervical cancer is an important problem in both developing and developed countries. Screening modalities that are implemented depend on the financial cost of each method, availability of adequately trained clinical and laboratory personnel, access to screening services, and local infrastructure. Because the resources in each region are different, it is likely that different screening methods, used individually or in combination with other(s), will provide the basis of screening services.

Test Referral Bias

The manner of handling of women with normal screening tests in studies designed to evaluate screening tests may lead to the presentation of incomplete information, or the drawing of inaccurate conclusions by the reader. Since many studies have reported a lack of sensitivity (high false negative rates) of the Papanicolaou smear,[28,58-61,94,95] women with a single negative Papanicolaou smear (or other screening test) should not necessarily be assumed normal. Such an assumption may more appropriately be made if the investigators are able to verify that women with negative screening test results are truly disease-free (e.g., by referring a subsample of women with all negative screening tests for colposcopy and having them confirmed as normal). In a study in which some women have a negative screening test, and no validation procedures are carried out, several of the usual measures of screening test performance (sensitivity, specificity, and negative predictive value) are likely to be biased.[50,96] In this situation, women with true disease may be missed by the screening protocol and subsequently excluded from colposcopy referral. Women with true disease but negative screening tests may therefore not have the opportunity for a biopsy to be taken. Figure 1 is a flow chart that outlines the screening, referral, and diagnosis protocol used in screening for cervical abnormalities.

Misclassifying screening results as negative would subsequently bias the estimates of sensitivity, specificity, the false positive rate, and negative predictive value.[50,97-100] Test referral bias occurs when women with true "disease" who are missed by the screening protocol are excluded from the denominator of the equation for estimating sensitivity and

are incorrectly included in the denominator of the equation for estimating the false positive rate (and in both the numerator and the denominator of the equation for estimating specificity). The misclassified cases also inappropriately fall into the numerator of the equation for estimating negative predictive value. Tables 3 and 4 illustrate algebraically how test referral bias affects the estimates of sensitivity, specificity, false positive rates, and negative predictive value. True disease missed by the screening protocol would bias the estimate of sensitivity toward a higher estimate (due to the falsely low denominator), assuming that the disease would have been picked up by colposcopic biopsy if the referral had occurred. Though less obvious, the estimate of specificity is also slightly overestimated, because the missed referrals are included in both the numerator and the denominator of the equation for specificity. Overestimating specificity corresponds to an underestimate of the false positive rate (i.e., 1–specificity), due to the falsely high denominator in the calculation of the false positive rate. True disease missed by the screening protocol would similarly bias the estimate of negative predictive value toward a falsely high estimate, because these cases would incorrectly be included in the numerator of the equation for negative predictive value. The estimate of positive predictive value would not be affected by test referral bias of missed true cases alone because both the numerator and the denominator of this equation are obtained from the sample undergoing the referral procedure (e.g., colposcopy).

The effects of test referral bias were tested using simulations with hypothetical data (Table 5). As indicated by Tables 3 and 4, similar effects of test referral bias will always be true when x is greater than or equal to zero. Testing a hypothetical cohort of

1100 women, the effects of not referring 1, 50, and 100 women with true disease for a confirmatory test resulted in shifts in the performance estimates in the directions indicated above. With increasing numbers of women misclassified due to test referral bias, a corresponding increase in sensitivity, specificity, and negative predictive value resulted. Similar patterns were seen when performing simulations of the effects of test referral bias on a hypothetical cohort of 10,000 women.

Misclassification due to test referral bias is different than the misclassification that can occur due to error in other steps of the diagnostic process. Error may also result in the colposcopic interpretation. Should a colposcopist not see a lesion on the cervix of a woman who has true disease, no biopsy will be taken, and the woman will be misclassified as nondiseased. Similarly, misclassification may occur if the colposcopist takes a biopsy from the wrong site, or from a part of a lesion that is not the most severe. Misclassification in the estimates of the performance of a screening test may also result from error in the referent diagnosis.

Cervicography

Over the years, concerns about errors in the conventional Papanicolaou smear have motivated some researchers to evaluate alternate and/or adjunctive screening methods. Earlier studies evaluated colposcopy as a possible alternative to the Papanicolaou smear for primary screening for cervical cancer with favorable results.[101] However, the time involved, the high cost of colposcopy, and the lack of sufficient numbers of highly trained colposcopists make such examination infeasible for mass

primary screening. Colposcopic examination is reserved for verification of lesions initially detected by the screening test. Cervicography was developed as a potential substitute for colposcopy.[3,102] Several authors have evaluated cervicography as an alternate primary and/or adjunctive screening method for the detection of cervical lesions. However, most of these studies are limited by one or more methodological consideration. In most previous studies of cervicography, women with normal screening test results were not referred for colposcopy, and no verification of the negative screening tests was performed. In many of these studies, these women were either assumed normal or excluded from subsequent analyses.

Cervicography was first described in 1981 by Adolf Stafl.[3] Stafl's initial cervigram classification scheme categorized results into the following four groups:

> *Negative*: the entire squamocolumnar junction is fully visible, no abnormal lesion is present
>
> *Suspicious*: abnormal lesion is present, characterized by white epithelium, punctation, mosaicism, and/or atypical vessels
>
> *Unsatisfactory*: the squamocolumnar junction is not fully visible, and
>
> *Technically defective*: the picture is out of focus, underexposed, overexposed, or the entire cervix is not visible).

This classification scheme was used in many subsequent studies[6,13,16,33,103-106] with mixed success, and has since been improved. The previously labeled "suspicious" lesions are now subclassified as "positive" or "atypical". Subgroupings of the positive and atypical categories indicate the location of the lesion relative to the cervical transformation zone and the evaluator's impression of the severity of the lesion. Women with positive lesions

are referred for colposcopy, while colposcopy referral is no longer recommended for women with atypical cervigram diagnoses. The earlier "suspicious" category included atypical results and was not a good indicator for achieving high specificity because an unacceptable proportion of women without disease were being identified and referred for colposcopy.

The "unsatisfactory" category of cervigram classification is no longer used. Instead, a negative cervigram is subcategorized according to whether or not components of the transformation zone are visible. As a normal process of aging, a woman's transformation zone moves cephalad into the endocervical canal. Earlier analyses often grouped unsatisfactory and technically defective results together. Consequently, women in the older age groups with a normal cervix may have been misclassified as "unsatisfactory" using the earlier classification scheme, whereas they would currently be classified as negative, albeit with the limitation that the full transformation zone can not be assessed using cervicography. In older women it is recommended that an endocervical smear be performed in conjunction with the cervigram. Table 6 shows the current cervigram classification scheme.

In his 1981 study, Stafl[3] evaluated 422 women ages 15–50 seeking routine screening services, in addition to a group of women referred for a previously abnormal Papanicolaou smear. In the routine screening group, he detected 293 women (72.5%) with negative cervigrams, 76 (18.8%) with unsatisfactory cervigrams, and 35 (8.7%) with suspicious findings. The 35 women with suspicious cervigrams were referred for colposcopy-directed biopsy. Ten of these were found to have some abnormality,

including one carcinoma *in situ*, two cases of severe dysplasia, three cases of moderate dysplasia, and ten cases of mild dysplasia. When Stafl compared the cervicography results with cytology results collected during the same screening examination, he found that cervicography picked up nine abnormalities not detected by cytology, including one case of severe dysplasia, two cases of moderate dysplasia, and six cases of mild dysplasia. Sensitivity and specificity were not calculated in his report because women with negative Papanicolaou smear and cervigram results were not referred for colposcopy-directed biopsy and could not be validated. Stafl's study was further limited by the relatively small number of screening participants, and the non-specific, earlier classification system for cervigrams. Stafl's early classification scheme likely overcalled minor or insignificant lesions by grouping them into the "suspicious" category. This classification scheme probably misclassified negative cervigrams by reporting cervigrams for women with a nonvisible squamocolumnar junction as "unsatisfactory". Despite these limitations, the results proved promising when compared to cytologic screening.

Following Stafl's initial report, Blythe[105] conducted a study of cervicography among 578 women with negative Papanicolaou smears. He found 188 women (32.5%) with suspicious cervigrams, 79 women (13.6%) with normal results, 152 (26.4%) with unsatisfactory cervigrams, and 159 (27.5%) with technically defective cervigrams. Of the women with suspicious lesions, 1 (0.5%) had biopsy-confirmed invasive carcinoma, 4 (2.1%) had CIN 2, 7 (3.7%) had CIN 1, and 14 (7.4%) had koilocytotic atypia. More than half of the technically defective cervigrams were due to human error. The high unsatisfactory rate is indicative of Stafl's classification scheme that allows for the likely

misclassification of the cervigrams of older women whose entire transformation zone was not visible. Of great interest is the fact that cervicography picked up one case of invasive cancer and four cases of CIN 2 that were missed by cytology. Even though the sample of women was not representative of the general screening population, the findings showed cervicography to be a promising technique that may have saved at least one life. In 1986, Spitzer and colleagues[107] supported this finding when they published a case report of a patient in their clinic whose invasive cervical cancer was detected by cervicography but missed by a repeat Papanicolaou smear following an atypical Papanicolaou smear result.

Tawa and others[6] evaluated the performance of cervicography for 3271 women ages 18–50 attending a clinic for routine cervical screening and who had not had an abnormal Papanicolaou smear within the previous six months. The investigators used a modification of Stafl's diagnostic classification system, which, as discussed above, is now considered less than optimal. Cervicography led to the referral of 373 women (11.4%) for colposcopy because of a suspicious cervigram, whereas conventional cytology led to the referral of 39 women (1.2%) because of a positive Papanicolaou smear. Of the 81 cases of biopsy-confirmed CIN, cervicography detected 88% of CIN 1, 94.4% of CIN 2, and 84.6% of CIN 3. In contrast, the Papanicolaou smear detected 14.0% of CIN 1, 11.1% of CIN 2, and 38.5% of CIN 3. However, cervicography resulted in an unacceptably high false positive rate. Cervicography yielded a suspicious result in 301 of the 316 women (95.3%) who had a biopsy diagnosis of "no dysplasia". In contrast, only 7.9% of these women had a positive Papanicolaou smear. It is important to note that at the time this study was conducted, koilocytotic atypia was thought to be an

HPV-induced lesion, whereas CIN 1 was not, and therefore these two categories were analyzed separately. In this study, the "HPV lesions" were included in the no dysplasia category, which could account for some of the false positive cervigram results. These researchers were unable to calculate measures of validity because they could not confirm that women with negative results by both cervicography and cytology were disease free.

Tawa and his colleagues used a matched pairs analysis to assess discordant screening results for women with positive histology results. This analysis showed that significantly more women with a confirmed positive biopsy had a positive cervigram and a negative Papanicolaou smear than had a positive Papanicolaou smear and a negative cervigram (McNemar's $\chi^2 = 44.3$, $p < 0.001$). Conversely, of the women with confirmed negative biopsy, significantly more women had negative Papanicolaou smears than negative cervigrams (McNemar's $\chi^2 = 248.9$, $p < 0.001$). These analyses are not hampered by the likely exclusion of women with false negative results by both screening tests because only cases with discordant test results are used for the McNemar's statistical test.[108] This relatively large study contributed favorable findings about the use of cervicography as a screening test for cervical neoplasia. However, the poor specificity of cervicography in this study produced questions about its feasibility as a mass screening test and influenced the later rethinking of the cervigram classification scheme.

Rehder and Blythe[104] conducted a study of 673 women attending a clinic for prenatal care or routine gynecologic screening. Their analysis produced confusing results because knowledge of the relationship between HPV and cervical neoplasia was in its infancy. Papanicolaou smear results interpreted as positive for HPV infection were

defined as koilocytosis, koilocytotic atypia, atypia suggestive of condyloma, or suspect condyloma. Cervigram results corresponding to dysplasia were interpreted as negative for HPV. Biopsies were interpreted as positive for HPV if koilocytosis was present. Their report did not allow for an assessment of the performance of cervicography for detecting the entire spectrum of cervical dysplasia.

Gundersen and colleagues[103] compared cervigrams and Papanicolaou smears for 250 women ages 20–40 during a screening visit at a Midwestern private practice. They reported high false negative rates for the Papanicolaou smear (84.6% of the women with histologically-confirmed condyloma or CIN) and high false positive rates for cervicography (14.3% overall). They failed to address concerns associated with women who may have had missed lesions by the screening protocol. Their study was also limited by the restricted age range, the relatively small sample size, the exclusion of women with technically defective or unsatisfactory cervigrams, and the nongeneralizable sample of women seeking private services in a Midwestern town.

Han and Lee[109] conducted a study in which 257 patients visiting an early detection clinic in Korea received cervicography and Papanicolaou smear testing, colposcopy, and tissue biopsy during a single examination. In their study, cervicography yielded a sensitivity of 85.2% and a specificity of 83.4% for the detection of any dysplasia. In contrast, the Papanicolaou smear yielded a sensitivity of 54.5% and a specificity of 78.1%. Since all women in this study underwent colposcopic examination, the estimates of sensitivity and specificity are statistically valid.

Coibion and others[110] described their study of 1834 women ages of 20–86 who visited a screening clinic in Belgium. All patients were tested with cervicography and a Papanicolaou smear, and were referred for colposcopy-directed biopsy if either test was positive. Of the 71 women who had colposcopy performed, 59 had an abnormal histologic diagnosis characterized by koilocytotic atypia, dysplasia, or neoplasia. Cervicography detected 55 of these 59 cases (93.2%), while cytology detected 11 (18.6%). Among the referred women, cervicography yielded 12 false positive results and cytology yielded no false positives, based on the broad histologic classification stated above. Cervicography identified all seven cases of CIN 3, whereas cytology detected two of these serious lesions. Cervicography classified three histologically-negative cases as cancer, while cytology classified one of these as cancer. CIN 2 was grouped with koilocytotic atypia and CIN 1; therefore the extent of detection of CIN 2 cannot be determined. The authors reported estimates of sensitivity, specificity, and negative predictive value, even though women with negative results by both screening tests were not verified.

Soutter and colleagues[13] evaluated cervicography in a high risk sample of 211 women with abnormal Papanicolaou smears who had subsequent colposcopic examinations and biopsies taken of visualized lesions. Assessment of cervicography was based on cervigram classification as "suspicious", including lesions that were called trivial or more severe. Cervicography correctly detected CIN in 46 of the 63 (73%) women with this diagnosis. There were 53 false positive readings among 148 women (36%) without CIN. Technically defective cervigrams accounted for 10.8% of the

sample, largely due to the cervix being obscured by blood. This study is limited by its small sample size, its restricted sample composition, and its failure to distinguish between grades of CIN.

Szarewski, et al.[12] evaluated the performance of cervicography among 1162 women who were self-referred to a primary screening service. Screening was performed with cervicography and only an endocervical smear. Because the authors did not report their complete data, it is difficult to interpret the results. However, they did report that of the 112 women who had a biopsy following a positive cervigram, 19 (17%) were histologically negative, 67 (59.8%) had koilocytotic atypia or CIN 1, and 26 (23.2%) had CIN 2 or worse. Additionally, 13 women had negative colposcopic examinations in which no biopsy was taken. A limitation of this study is that colposcopic examinations and subsequent biopsies were performed by each participant's personal physician. Interobserver variability may affect the consistency of colposcopy and histology results. Since only an endocervical cytologic smear was taken, it was assumed that cervicography would have picked up all abnormalities on the ectocervix. It is possible, however, that a complete Papanicolaou smear (containing endocervical and ectocervical components) would have detected additional women who were not picked up by cervicography.

In a study conducted by Reid and colleagues,[5] 1012 women ages 18–35 were screened using cervicography, Papanicolaou smear, and HPV testing using Southern blot hybridization. Of these women, 298 (29.4%) were referred for colposcopic examination and directed biopsy after having one or more abnormal screening test. Three screening methods were used in this study, and, although not conclusively verified, the investigators

assumed that women with negative results on all three tests were true negatives. Some assurance was provided to the cytology diagnosis, because these were evaluated by two independent cytopathologists, and discordant results were adjudicated by a third cytopathologist. This was one of the first studies of cervicography performance that compared the findings with the presence of cancer-associated HPV types. The Southern blot hybridization test was an earlier method used for HPV testing, which has since been superseded by more sensitive methods.[111] These investigators looked at detection of CIN 2/3 by cervicography and cytology separately from CIN 1. Cervicography identified 14 of 23 cases of CIN 2/3 (60.9%), whereas the Papanicolaou smear identified 12 (52.2%) of these important lesions, a difference which was not statistically significant. The authors reported that the combination of cervicography and cytology used for the detection of CIN 2/3 led to an unacceptably high recall rate (i.e., low specificity). This study was limited by the self-selected samples of women in a narrow age range. Women were selected from an urban sexually transmitted disease clinic (n=672) and from suburban gynecologic clinics (n=340). Results for these two samples were grouped together, although they represent different populations.

Cecchini and others[112] evaluated cervicography among 606 women referred for colposcopy because of an abnormal Papanicolaou smear. Two evaluators interpreted cervigrams separately. As assessment of interobserver variability on two categories of diagnosis (negative vs. suspicious) yielded a *kappa* statistic of 0.61, indicating moderately good agreement. The *kappa* statistic indicates the extent of observer agreement beyond that expected by chance alone. *Kappa* ranges from -1 to 1, with 1.0

corresponding to perfect agreement, zero corresponding to no agreement, and negative values corresponding to agreement worse than that expected by chance.[113] Suspicious results were detected by cervicography in 50% of the 606 cases by evaluator A and in 58.6% of the cases by evaluator B. Cervicography yielded false positive results of 170/303 (56.1%) and 222/355 (62.5%) by evaluators A and B respectively, indicating its low specificity. Of HPV/CIN 1 lesions, 79.4% (Evaluator A) and 80.8% (Evaluator B) were detected by cervicography. Of lesions classified as CIN 2 or more severe, 95.2% (evaluator A) and 90.5% (evaluator B) were identified by cervicography. The results of this study are not generalizable to a primary screening population because all women in this study were inherently at high risk due to their prior abnormal Papanicolaou smears.

A multicenter study of 1449 women ages 12–90 presenting for either routine cervical screening or followup evaluation for an abnormal Papanicolaou smear was conducted by Ferris and others.[8] Cervicography correctly identified 165 of 326 women (50.6%) with dysplasia, including 122 of 265 (37.4%) with mild dysplasia, 28 of 42 (66.7%) with moderate dysplasia, 14 of 18 (77.8%) with severe dysplasia, and 1 of 1 (100%) with cancer. False positive cervigram results were common, even in this high-risk population. Of the 106 women with negative or atypical histology, 47 (44.3%) had positive cervigrams. In this sample, 79.2% of women with dysplasia were identified with a Papanicolaou smear result of atypical or more severe, and only 8% received false positive cytology results. These authors also assessed the results of combined testing with cervicography and the Papanicolaou smear, where a result was considered positive if either screening test was positive (atypical Papanicolaou smears were not included in the

positive group for this analysis). Combined testing identified 63% of women with any dysplasia, including 59% with mild dysplasia, 80% with moderate dysplasia, 81% with severe dysplasia, and 100% with cancer. However, this testing modality resulted in 52.5% of women with no dysplasia being misclassified as positive by the screening tests. It would have been optimal for these investigators to analyze the screening and referral women separately as well, because they represent populations with different levels of risk.

A study in Australia compared cervicography to cytology in a sample of 245 women ages 18–53 attending an STD clinic.[114] In this high-risk population, cervicography correctly identified 68 of the 101 women (67.3%) with dysplasia (koilocytotic atypia and CIN). Cervicography yielded false positive results in 4 women (3.9%) and false negative results in 33 of the 101 women (32.7%) with dysplasia. Cytology correctly identified 47 women (46.5%) with dysplasia, with one false positive Papanicolaou smear and 54 (53.5%) false negatives. The authors did not distinguish between severity of CIN, and the results are not generalizable to a routine screening population.

In another study by Cecchini and her colleagues[11] 2105 women ages 17–83 attending a Papanicolaou smear clinic also received cervicography and a visual inspection examination with acetic acid and no magnification (VIA). Women were referred for colposcopic examination if they had an abnormal Papanicolaou smear, a suspicious cervigram, or an acetowhite area detected by VIA. These screening tests were compared with a gold standard based on colposcopy-directed biopsy, with a positive result

characterized by CIN 2 or CIN 3. An interesting finding was that of the four cases of CIN 3 detected, three (75%) were detected by VIA, two of which (50%) were also identified by cervicography. Two of these serious cases (50%) were detected by cytology (one in which the other two screening tests were negative). Of the four cases of CIN 2, three (75%) were identified by all three screening methods, and one was picked up only by VIA. Histology results that were negative, koilocytotic atypia, or CIN 1 were reported together. Of the 2028 women in this group (including those not referred for colposcopy), 280 (13.8%) received suspicious cervigram results, 334 (16.5%) received positive VIA results, and 69 (3.4%) received positive cytology results. High false positive rates such as those reported here are expected when the cutoff for a positive screening test is lower than it is for histologic confirmation.

A small study by Kesic and colleagues[15] addressed some of the limitations of previous studies by performing a colposcopic examination of all 418 women in their study. Therefore, all women had the same opportunity for a colposcopy-directed biopsy. In this study, cervicography yielded a sensitivity of 89.5 % (17/19 cases of CIN 2, CIN3, microinvasive cancer, or invasive cancer correctly identified) and a specificity of 90.2% (339/376 individuals diagnosed as normal, trivial changes, or CIN 1 correctly identified) for the detection of histologically-confirmed lesions classified as CIN 2 or more severe, although numbers were small. Cervicography performed significantly better than cytology, which yielded a sensitivity of 52.6% (10/19 cases correctly identified) and a specificity of 93.1% (350/376 cases correctly identified). Papanicolaou smears Class III and higher were considered positive; therefore, some smears corresponding to

koilocytotic atypia or ASCUS may have been misclassified as negative. This study appeared to use a cervigram classification system which more closely resembles the classifications currently approved by NTL, in which negative cervigrams and those showing only trivial lesions are grouped with negative results for analytic purposes. This study is limited by its small sample of clinic attendees, but it does overcome some of the methodological problems encountered by earlier studies. This study also demonstrates that grouping atypical cervigrams with the negative results rather than with positive cervigrams helps improve the specificity of the test.

One of the largest studies comparing cervicography with cytology was conducted by Coibion and others who studied 4015 women ages 20–79 attending a screening clinic.[7] Women with a positive result by either cervicography or cytology were referred for colposcopy. Cervigram and cytology results classified as atypical were considered negative for analytic purposes. These investigators found that cervicography detected significantly more lesions classified as any CIN or carcinoma than did cytology. Cervicography correctly identified 106 of 123 (86.2%) cases of CIN or carcinoma, while cytology identified 27 (22.0%). Of these 123 women, 96 (78%) had positive cervigrams and negative or atypical cytology results, while only 17 of women in this group (13.8%) were detected by cytology and not by cervicography (McNemar's $\chi^2 = 53.8$, $p < 0.001$). It is possible that some of this difference could be accounted for by women with CIN or cancer who received atypical cytology results, which were not included in the positive cytology group in this study. When considering the relative performance of these screening tests for the detection of CIN 2/3 and cancer, no significant difference was

found between the two tests. Of the 24 cases of CIN 2/3 or cancer, 10 (41.7%) were correctly identified by cervicography and not by cytology, and 8 (33.3%) were correctly identified by cytology only (McNemar's $\chi^2 = 0.05$, $p = 0.99$). Cytology was significantly more specific than cervicography. Of the 3991 with either normal histology or CIN 1, 120 (3.0%) were classified as negative or atypical by cytology and not by cervicography and 15 (0.38%) were classified as negative or atypical by cervicography only (McNemar's $\chi^2 = 80.1$, $p < 0.001$). Cytology yielded a positive predictive value for the detection of CIN 2/3 or carcinoma of 42.4%, while cervicography yielded a positive predictive value of 11.4%. Coibion and colleagues also showed that, as expected, cervicography detected more CIN 2/3 and cancer among women in the younger age groups, and it did not identify any of these lesions among women ages 55 or older. Such differential in the performance of cervicography by age groups is expected, because cervicography cannot detect lesions inside the endocervical canal, which is more common in older women.

The study by Coibion and others[7] was followed by a multicenter evaluation of cervicography in clinics pertaining to three hospitals and three cancer screening centers (one of which is the same location where the Coibion study was conducted).[115] The screening and histologic diagnoses were made separately in each clinical center, and therefore, the investigators could not account for interobserver variability in the screening and histologic diagnoses. In the Coibion study[7] histology was assessed independently by two pathologists with a third pathologist arbitrating discordant results. In the later study,[115] a single reading was performed because the arbitration process used in the first

study was not found to alter the results significantly. In this multicenter study, cervicography detected 18 of the 33 cases (54.5%) of CIN 2/3 and 72 of 116 cases (62.1%) of koilocytotic atypia and CIN combined. Of the 5159 women with koilocytotic atypia, CIN 1, or a negative diagnosis (including those with negative screening tests), 5013 (97.2%) had negative or atypical cervigrams. Cytology identified 29 of the 33 cases (87.9%) of CIN 2/3 and 64 of the 116 cases (55.2%) of koilocytotic atypia and CIN combined. Of the women without CIN 2/3, 5093 (98.7%) had negative cytology results. Among the cases of CIN 2/3, 4 (12.1%) were detected by cervicography and not by cytology, and 15 (45.5%) were detected by cytology only.

The investigators addressed some of the differences in the findings of these two Belgian studies.[115] They suggested that the addition of new cervigram evaluators, some with less experience, may have contributed to the lower detection capacity of cervicography in the later study. This suggestion is supported by their reporting of greater interobserver variability among cervigram evaluators in the later study (data not reported). In the Coibion study,[7] cervigrams were assessed by two evaluators working collaboratively, whereas in the multicenter study, cervigrams were evaluated by evaluators working independently. The collaborative process may have impacted on the improved performance of cervicography in the earlier study.

Baldauf and others[9] conducted a study in a sample of 1539 women ages 15–82 undergoing routine cervical screening or prenatal examination. Women with positive or atypical cytology or cervigram results, and a random 10% sample of women with negative screening test results, were referred for colposcopy. The results of the random

10% subsample were extrapolated to the larger subgroup of women with negative screening tests, though the authors did not report the results. For analytic purposes, a positive cervigram was defined as Positive 1, Positive 2, or Positive 3, and a positive cytology result was defined as having atypical cells, koilocytotic atypia, CIN, or cancer. Cervicography correctly identified 13 of the 23 women (56.5%) with CIN 2/3 and 33 of 62 women (53.2%) with any CIN. Cervicography yielded false positive results in 39 women (no CIN found), representing 3.0% of the 1281 women with a negative diagnosis. Atypical cervigrams corresponded to 96 of the 262 women (36.6%) with negative histology, 11 of the 39 women (28.2%) with CIN 1, and 7 of the 23 women (30.4%) with CIN 2 or CIN 3. It is again evident from these results that grouping atypical and negative cervigrams together shifts cervicography performance toward greater specificity, with a corresponding reduction in sensitivity. Cytology also identified 13 of the 23 cases of CIN 2/3 (56.5%) and 33 of 62 women with any CIN (53.2%), while yielding 24 false positive results (1.9% of women with a negative diagnosis). Of the 62 women with confirmed lesions, 28 (45.2%) had discordant screening test results by the two methods. The two screening methods combined yielded a sensitivity of 77%, a specificity of 95%, a positive predictive value of 44%, and a negative predictive value of 98% for all CIN (extrapolated values for negative screening tests were incorporated into this analysis). The performance estimates of the combined testing for the detection of CIN 2/3 were not provided.

Using an interesting approach, a recently published study [116] compared a screening regimen of combined cytology and cervicography with cytology alone in a randomized

trial of 5550 women designed to assess whether a difference exists in the number of women with CIN in a subsequent screening one year later. Their results did not show a significant difference in the number of cases of CIN among women initially screened with either cytology alone or the combined testing. Implicit in their alternate hypothesis was a suggestion that initial screening by cytology alone would result in undetected CIN, which would be identified during screening one year later. However, this approach did not account for regression of CIN or development of new lesions during this time period. A very interesting finding, though, was that no cases of CIN 2/3 were detected by both cervicography and cytology, and only one case of CIN 1 was identified by both tests. Thus, the two screening methods identified different women with abnormal cervices. In the combined screening arm at the initial screening, equal numbers of cases of CIN 2/3 were identified by each method. At the second screening, 7 (63.6%) and 4 (36.4%) of 11 cases of CIN 2/3 were identified by cytology and cervicography respectively. The total number of false positives was much higher for cervicography (n=152) than for cytology (n=12), even though all women received cytology during the initial screening, and only one study arm received cervicography initially.

This historical review of the literature provides some insight into the changes made in the cervigram classification scheme since Stafl first began using the method in 1981. These studies also reflect gains in the knowledge of the natural history of HPV infection and cervical neoplasia. In the early studies, cervigrams classified as "suspicious" included those with trivial changes that are now classified as "atypical" (to be distinguished from cervigrams currently classified as "positive") in recognition that

they are less often associated with clinically significant abnormality. The referral of any woman with a suspicious cervigram frequently led to an inappropriately high proportion of women with no true neoplasia being referred for colposcopic examination. An additional shortcoming of the initial classification scheme was the inclusion of women for whom the squamocolumnar junction was not fully visible in an "unsatisfactory" category This led to a high proportion of unevaluable cervigrams, many of which corresponded to older women with presumably normal cervices. Revisions made to the cervigram classification scheme resulted in a great improvement in the specificity of cervicography, at the expense of reduced sensitivity. However, the higher specificity of cervicography made the technique more feasible as a method for mass screening for cervical neoplasia, if only the sensitivity for high grade neoplasia could also be improved.

When cervicography was first introduced, the association of HPV infection with cervical cancer was not known. When the association of HPV with cervical abnormalities was initially established, its association with the process leading to CIN and cancer was not well understood.[104] In some of these earlier studies,[104] "HPV effect" was the term used for koilocytotic atypia, a category now reported in the Bethesda cytologic classification as LSIL (along with CIN 1). It is now known that most cases of koilocytotic atypia and CIN 1 resolve without treatment, while CIN 2/3 is much more likely to progress to cancer (though it can also spontaneously regress).[20] Many clinicians now feel that CIN 2/3 and cancer are more important outcomes and therefore the appropriate targets for screening. This change in ideology is apparent in the more recent studies of cervicography where the investigators have reported results of the performance

of cervicography and cytology for the detection of CIN 2/3 separately from CIN 1 and koilocytotic atypia.[5,7,9,11,12,15,115] Several of these studies have shown that, as the definition of "disease" has become more restrictive by including only CIN 2/3 and cancer, the sensitivity of cervicography has weakened compared to its performance for the detection of koilocytotic atypia, all CIN, and cancer combined.[5,7,115]

All of the previously published studies suffer from their inclusion of selected populations. Studies of the use of cervicography as a primary screening method are limited by their inclusion of women selected from among clinic attendees. Clinic populations may not be representative of the general population in the region where the studies are conducted, but they are representative of the population to be served at the particular clinic. Studies of women receiving gynecologic services for routine screening involve self-selection for seeking the screening service. Women in the general population who choose to receive routine screening services tend to be different from women who receive gynecologic services in private medical offices, from women who do not seek screening services, or from those who do not know that they should. The studies of women referred for and attending a colposcopy examination are limited because the women with a recent abnormal Papanicolaou smear represent a high risk population, which is different from the primary screening population or the general population for cervical cancer risk. The evaluation of a screening method in targeted, high-risk populations typically leads to falsely high estimates of positive predictive value because of the likely higher prevalence of disease.[46]

An additional use of cervicography is as a triage method to help sort out a management plan for women with atypical Papanicolaou smears (ASCUS). Since triage is different from primary screening, studies of the evaluation of cervicography as a triage method are not presented in here in detail. Several investigators[14,117-120] have found that cervicography performed more sensitively than a repeat Papanicolaou smear for detecting koilocytotic atypia or CIN in women with a previous ASCUS Papanicolaou smear, but it detected fewer true abnormalities than did colposcopy.

Other Visual Screening Methods

Direct visual inspection

Some public health professionals and researchers believe that widespread cytologic screening is not possible in many developing countries because of the prohibitively high cost of maintaining quality laboratories and well-trained cytotechnologists and cytopathologists. Cervicography may be able to perform a valuable role in such settings. Another alternate method of screening is direct visual inspection. Direct visual inspection is a low cost, low technology screening method that has been proposed for use in developing countries. This technique involves the visual inspection of the cervix through a speculum, either with or without application of acetic acid. Direct visual inspection can be performed by trained, non-physician health workers. Nurses or other health workers learn to distinguish visible cervical abnormalities, and refer them for further evaluation by a physician. Direct visual inspection does not require the expense of a specialized camera nor the expertise of a highly trained and certified cervigram

evaluator. If cervicography is to be a feasible screening method in low resource settings, it would have to prove itself better than its "appropriate technology" counterparts. There is scant research available on the performance of visual inspection as a screening tool. The results of some of the published studies are presented here.

In a research protocol designed to detect invasive cancer among women attending maternal and child health centers in Delhi, India, Singh and colleagues[121] found that direct inspection alone would have led to the referral of 11.4% of women and detection of 62.6% of the invasive cancers. Detection was determined by assignment into a high-risk category defined by cervical erosions that bled on touch, small growth, or a suspicious looking cervix. The authors did not mention applying acetic acid, and the visual examination appeared to be performed by physicians. True estimates of sensitivity and specificity could not be determined, because no information was provided on women who were initially screened as negative. These investigators did not assess the performance of visual inspection for detecting cancer precursors (CIN 2/3) or minor abnormalities (koilocytotic atypia and CIN 1).

To evaluate the performance of unaided visual inspection on a general screening population, Nene and others[122] conducted a population-based study in rural India. Two threshold levels for visual screening were established based on descriptive visual characteristics of the cervix. A Papanicolaou smear was taken at the time of visual examination. The gold standard was described as cytology/histology. Colposcopy was not available at the clinic at the time of the study, and the authors did not report the tissue collection method used for histologic diagnosis. Performance estimates of visual

inspection (sensitivity and specificity) were reported, but are unreliable due to the poor sensitivity of cytology and the lack of colposcopy to help determine the biopsy site. However, the proportion of women referred by visual inspection provides important information about the usefulness of the method as a screening technique. Depending on the referral threshold used, 57.3% (low threshold) and 6.0% (high threshold) were detected. The high referral rate using the low threshold would likely result in an inappropriate number of woman with normal cervices being referred for further testing. Restricting referrals appeared to result in the failure to detect women with high grade lesions or cancer, although this assessment is limited as described above.

A study of 2843 married women ages 30 and older in Kerala, India, was carried out by Wesley and colleagues.[123] Visual inspection without magnification or acetic acid was performed by a trained cytotechnician supervised by a medical doctor. This study suffered from the same limitation as that experienced by Nene and colleagues.[122] Colposcopy was not available and the gold standard was determined by a combination of cytology and histology, without the benefit of colposcopy to direct the biopsy placement. Additionally, only women with cytology results of severe dysplasia, carcinoma *in situ*, and invasive cancer were referred for further gynecological examination, biopsy, and treatment. This design did not allow for the possibility of misclassification on the degree of abnormality on the cytology result. Visual inspection led to the referral of 1279 women (45.0%) using the low threshold and 179 women (6.3%) using the high threshold. Reported estimates of sensitivity and specificity are unreliable for the reasons stated above.

The performance of visual inspection of the cervix after application of acetic acid was evaluated by Megevand and colleagues[124] in a study of 2426 women ages 20–83 in Cape Town, South Africa. Women were referred for colposcopy if they had a cytology result of squamous intraepithelial lesions or if their visual examination with acetic acid revealed acetowhite areas. The results for women with ASCUS or negative cytology and with no acetowhitening were assumed negative and not confirmed by any other method. Valid estimates of sensitivity and specificity therefore cannot be calculated. Histology results were determined by colposcopy-directed biopsy or from excised tissue of suspected CIN 2/3 as determined by colposcopic examination. Visual inspection with acetic acid led to the referral of 76 women (3.1%). Of the 31 women with a diagnosis of CIN 2/3, 20 (64.5%) were detected by visual inspection with acetic acid. Of the 284 women with a diagnosis of koilocytotic atypia or more severe, 55 (19.4%) were detected by visual inspection. Of the 2142 women with assumed normal cervices, 2121 (99.0%) had no acetowhitening on visual examination, suggestive of high specificity. In sum, these results seemed promising.

In a small study of 95 college students, Frisch and others[125] evaluated the predictive values of visual inspection of the cervix with acetic acid in combination with the Papanicolaou smear. Women were referred for colposcopy and directed biopsy, if applicable, if they had a cytology result of atypical or more severe, or a positive cervigram (versus negative, atypical, or technically defective). All three women with confirmed CIN 2/3 were detected by visual inspection, whereas cytology missed two of these serious lesions. Of the 95 study subjects, 71(74.7%) had acetowhitening visible on

examination. Positive predictive values were reported as 82% for cytology alone and 57% for cytology plus visual inspection with acetic acid. Negative predictive values were reported as 67% for cytology alone and 91% for cytology plus visual inspection (i.e., the numerator is equal to the number of women with both tests negative). Therefore, if a screening modality were to refer cases for a positive Papanicolaou smear or a positive visual examination, many false positive results would occur due to the overreferral of women by visual inspection. However, negative results by both tests would confirm the negative disease status more than a negative Papanicolaou smear alone. These results should be interpreted with caution because of the small, restricted sample and because women with a visual acetowhite lesion but negative results by other methods were not referred for colposcopy.

The recently published results of a well-designed study conducted in Zimbabwe[126] brings out several important issues surrounding visual screening for cervical cancer. During the first phase of this study, 8731 women for a primary care clinic underwent VIA and Papanicolaou smear screening, and women with positive tests were referred for colposcopy with directed biopsy if indicated. Positive VIA results were defined as abnormal (white plaques, ulcer, or acetowhite epithelium) or cervical cancer (cauliflower-like growths or fungating mass). Positive Papanicolaou smears were defined as LSIL or more severe, and cytology results of ASCUS or atypical glandular cells of undetermined significance were considered test negative. Though a random selection of women with normal or atypical VIA results were scheduled for colposcopy as a control group, there was a differential response among the control women (i.e., women with negative VIA

results who were referred for colposcopy) who attended and did not attend their colposcopic examination. Therefore, sensitivity and specificity were not reported.

A second phase of the Zimbabwe study recruited 2203 women for screening and colposcopy examination for all women. Sensitivities of 76.7% and 44.3% were reported for VIA and cytology respectively for the detection of HSIL. Specificities of 64.1% and 90.6% were reported for VIA and cytology respectively for the detection of HSIL. The results of the Zimbabwe study may not be generalizable to all age groups, because the mean age of the women in the study was 32.2.

The low specificity of VIA brings into question its usefulness as a screening technique in a country with limited resources. While the low cost of direct visual inspection makes the technique accessible in resource-poor populations, the high false positive rate generates an increase in costs associated with unnecessary colposcopic examinations, if provided, along with the emotional burden of 36% of women without abnormal cervices receiving positive screening test results. To date, most studies of visual inspection of the cervix as a screening method have shown high rates of referral for colposcopy, corresponding to low specificity. Though the costs associated with the technique itself are low, the referral of large numbers of women without cervical abnormalities would result in excessive financial and emotional costs associated with followup and management of these women.

Aided visual inspection devices

Speculoscopy

Speculoscopy is a visual screening method which uses a chemiluminescent blue-white light called a Speculite® (Trylon Corporation. Torrance, CA) attached to the upper dilator blade of the speculum and a low-power magnification loupe. The cervix is washed with 5% acetic acid, and visualization with the speculite allows for magnified observation of lesions staining white.

Several studies have been conducted to evaluate speculoscopy as an adjunctive method used with conventional Papanicolaou smear screening.[127-131] Others have assessed the validity of speculoscopy used alone as a primary screening method[131,132] or as a technique to follow women with abnormal Papanicolaou smears.[133] These studies generally have shown an increase in sensitivity when combined Papanicolaou smear plus speculoscopy was used, compared with Papanicolaou smear alone. This is a necessary outcome when two screening tests are combined and a "positive" result is defined by either test being positive. These studies are limited by methodological considerations, the most important of which are the failure to adequately address loss of specificity, differing reference tests used as a gold standard to define "diseased", and/or failure to mask the investigators to previous screening test results.

One recently published study,[131] however, did report on the specificity of speculoscopy. For the detection of HSIL only, they found that speculoscopy yielded a sensitivity and specificity of 84% and 21% respectively. For the detection of LSIL and HSIL combined, speculoscopy yielded a sensitivity and specificity of 82% and 23%

respectively. Stated in another way, 79% of all women in the study were referred for colposcopic examination. Such high referral and false positive rates (79% for HSIL and 77% for all SIL) would be unacceptable for general screening, since a very large proportion of healthy women would be referred for colposcopic examination.

Aviscope

A device similar to the Speculite is currently being evaluated by the Program for Appropriate Technology in Health (PATH, Seattle, WA). This optical device, called the AviScope™, uses green and white LED lights to provide 4x magnification to highlight acetowhite tissue. There are no published, peer-reviewed studies available on the current prototype of the AviScope at the time of this writing.

Optoelectronic Devices

Some of the newer techniques proposed for cervical cancer screening are optoelectronic devices, such as the Polarprobe® [Polartechnics Ltd (Truscan®), Australia].[134] These devices are portable optoelectronic instruments that rely on low level electrical impulses and light pulses at various frequencies to classify tissue response. A probe is placed in contact with the cervix and the energy response signals and spectroscopic pattern are measured and relayed to a computer. The pattern is compared to a computerized catalogue of preprogrammed signal patterns, permitting immediate feedback as to the tissue type.[135] Optoelectronic devices have not yet undergone rigorous testing, although trials are currently being planned.

HPV Testing

HPV testing involves the use of one or more nucleic acid-based tests for detecting the presence of HPV DNA or RNA and, if present, determining its type. These tests, called hybridization or probe methods, involve the specific recognition and pairing of target nucleic acid sequences by complementary, or homologous, probe nucleic acid sequences.[111] Probes are developed using sequences corresponding to known HPV types. Three broad categories of molecular methods used to describe methods for HPV detection and typing are non-amplified hybridization, target-amplified hybridization, and signal-amplified hybridization.[111]

Non-amplified methods include Southern blot, dot blot, filter in situ hybridization, and in situ hybridization, among a few others. These methods are commonly used as "back-end" procedures for detecting target-amplified polymerase chain reaction (PCR) products. Southern blot and dot blot procedures use nylon or nitrocellulose filters to bind target DNA before hybridization, thus providing a localized reaction area where manipulations (e.g., washing and changing buffer components) can be performed without loss of target DNA.[111] Southern blot provides extra steps to add specifically fragmented targeted DNA before immobilization to allow for subtype discrimination, whereas dot blot does not provide information on HPV type unless type-specific probes are employed.[111] Southern blot is considered a reasonably sensitive and specific method for distinguishing DNA fragment patterns for rapid mapping of new DNA types, studying HPV integration, or as a confirmatory assay for amplified tests. However, the Southern blot technique is considered too complex for routine clinical use. [111]

Polymerase chain reaction is a target-amplified test that employs repeated cycles of denaturation, primer hybridization, and primer extension to create numerous (more than one million) copies of the desired target DNA sequence. After amplification, detection is achieved by applying a back-end procedure to the amplified DNA products (called amplicons). Appropriate detection techniques include conventional hybridization methods (Southern blot or dot blot), bromide staining and visualization following gel electrophoresis, or oligonucleotide probes in enzyme-linked immunosorbent assay (ELISA) format.[111] A specific advantage of PCR is that it can utilize a minute amount of DNA from Papanicolaou smears or tissue sections and subsequently produce numerous amplicons. PCR methods, especially those using oligonucleotide probe-based ELISA lend themselves to fairly accurate HPV typing.[111] PCR's high biologic sensitivity is also a disadvantage because the method is susceptible to contamination, which may produce analytical false-positives, or to the detection of clinically insignificant amounts of HPV DNA.[111]

The Hybrid Capture method (Digene, Silver Spring, MD) is a signal-amplified test for detecting DNA or RNA targets.[111] For HPV detection, Hybrid Capture utilizes immobilized RNA probes on a microplate to target HPV DNA, followed by an immunologically-based back-end test similar to ELISA. Hybrid Capture employs the immobilized single-stranded RNA probes to hybridize with all 8000 DNA nucleotides of the entire HPV genome. A washing step eliminates all but the target DNA. A second antibody conjugated to molecules of alkaline phosphatase is added, which results in the coating of the captured DNA-RNA hybrids with thousands of antibodies. Detection is

accomplished by the addition of a chemiluminescent dioxetene-based substrate which reacts with the alkaline phosphatase enzymes. This reaction results in a light emission, which is measured by a luminometer in relative light units. The intensity of the light is proportional to the quantity of target DNA in the specimen, and is expressed as a ratio of the signal to the positive control.[111] The specific advantage of the Hybrid Capture method, therefore, is its ability to quantify the amount of HPV DNA in a specimen.

With the first generation Hybrid Capture System (Hybrid Capture tube test), processed samples were hybridized in tubes and two mixtures of RNA probes were used to recognize high risk HPV types 16, 18, 31, 33, 35, 45, 51, 52, and 56 and the low risk types 6, 11, 42, 43, and 44. With the second generation Hybrid Capture II (HC II), analytical sensitivity was improved by the addition of new probes for high and intermediate risk HPV types 39, 58, 59, and 68, and by the reformulation of the hybridization reagents. Additionally, the tube test kit was replaced by microtiter plates, making the test easier to perform and therefore, making it more appropriate for clinical use.[136] Hybrid Capture II recently received approval from the U.S. Food and Drug Administration (FDA). A recent study of Hybrid Capture for cervical cancer screening indeed found improved sensitivity for HC II compared to Hybrid Capture tube test.[137] The Hybrid Capture tube test detected around 70% of HSIL and cancer (sensitivity) at the 10 pg/ml level to define positivity, whereas HC II achieved a sensitivity of 88.6% while referring 12.5% of women for colposcopy using a threshold of 1.0 pg/ml to define positivity. Further decreasing the threshold for positivity yielded little gain in sensitivity, but sharp reductions in specificity occurred.

As knowledge of the association of HPV with cervical neoplasia developed, it has been suggested that HPV testing be used clinically to screen for cervical cancer. It is now understood that most lesions associated with HPV infection are benign and self-limiting, especially in younger women. Therefore, mass HPV testing of the general population is not feasible, because it would lead to the overmanagement and overtreatment of women with lesions that are likely to spontaneously regress.[92]

Current opinion considers that the key to HPV testing is to use it as a triage method to help predict which women with mild lesions (koilocytotic atypia and CIN 1) are most likely to progress to more serious states of dysplasia (CIN 2/3) or cancer.[92] Some proposed uses of HPV testing include 1) clarification of inconclusive cytologic diagnoses (i.e., ASCUS), 2) general screening of older women, and 3) triage of koilocytotic atypia and CIN 1.[92] As a method to aid in the interpretation of inconclusive cytologic diagnoses, HPV testing might be used to provide quality control to cytopathology laboratories, to triage individual inconclusive cytology results, and to resolve non-diagnostic colposcopy or histology following cytologic evidence of koilocytotic atypia or CIN. HPV testing might be used to supplement Papanicolaou smears to help identify older women at high risk of developing cervical cancer. HPV prevalence in cytologically normal women declines with age, while it remains high in women with cervical neoplasia, regardless of age. Therefore, HPV testing may permit more selective screening of postmenopausal women. As a triage method for women with cytologic evidence of a low grade lesion, HPV testing may help determine which women are likely to progress to more advanced dysplasia based on the type and persistence of

HPV infection.[92,93] A large, multicenter trial is currently underway to evaluate various triage protocols (including HPV testing) for women with ASCUS and low grade cytology results.[45] HPV data from this trial are not yet available.

Summary

Cervicography is one of several available screening methods for detecting women at risk of developing cervical cancer. Cervicography was developed at a time in which clinicians were becoming aware of the limitations of the existing standard of Papanicolaou smear screening. While cytologic screening has been the standard of care since the mid-1950s, cervicography has been scrutinized more critically. In earlier studies, the performance of cervicography has been mixed. Its low specificity during the early years was reduced by modifying the classification system for colposcopic referral. More recent studies of cervicography are limited by their small sample size and inclusion of selected populations. For cervicography to become a feasible primary screening option, its success must be proven not only against the Papanicolaou smear, but also against new cytologic technologies, HPV testing, and other visual screening methods.

The present research project addresses many of the limitations of previous studies. The participants were randomly selected from a population in an area with a high prevalence of cervical cancer. Participation rates in the study were above 90%. Cervicography and three cytologic screening methods were used to identify women for colposcopic examination. An additional random sample of women was referred for colposcopy to validate the screening protocol. Results of histology and HPV testing are

available to further assess screening test performance. Additional reviews of cervigrams and diagnostic materials help us to estimate the potential, optimal performance of cervicography screening and to further explain the results achieved.

References

1. Ferris DG, Greenberg MD. Reid's colposcopic index. J Fam Pract 1994; 39:65-70.

2. Burghardt E, Pickel H, Girardi F. Colposcopy and cervical pathology: Textbook and atlas. 3rd ed. New York: Thieme, 1998.

3. Stafl A. Cervicography: A new method for cervical cancer detection. Am J Obstet Gynecol 1981; 139:815-825.

4. Personal communication with Fred Kostecki, National Testing Laboratories Worldwide, 1999.

5. Reid R, Greenberg M, Lorincz A, Jenson AB, Laverty CR, Husain M, et al. Should cervical cytologic testing be augmented by Cervicography or human papillomavirus deoxyribonucleic acid detection? Am J Obstet Gynecol 1991; 164:1461-1471.

6. Tawa K, Forsythe A, Cove JK, Saltz A, Peters HW, Watring WG. A comparison of the Papanicolaou smear and the Cervigram: sensitivity, specificity, and cost analysis. Obstet Gynecol 1988; 71:229-235.

7. Coibion M, Autier P, Vandam P, Delobelle A, Huet F, Hertens D, et al. Is there a role for Cervicography in the detection of premalignant lesions of the cervix uteri? Br J Cancer 1994; 70:125-128.

8. Ferris DG, Payne P, Frisch LE, Milner FH, diPaola FM, Petry LJ. Cervicography: Adjunctive cervical cancer screening by primary care clinicians. J Fam Pract 1993; 37:158-164.

9. Baldauf J, Dreyfus M, Lehmann M, Ritter J, Philippe E. Cervical cancer screening with Cervicography and cytology. Eur J Obstet Gynecol Reprod Biol 1995; 58:33-39.

10. Schneider A, Zahm DM, Kirchmayr R, Schneider VL. Screening for cervical intraepithelial neoplasia grade 2/3: Validity of cytologic study, Cervicography, and human papillomavirus detection. Am J Obstet Gynecol 1996; 174:1534-1541.

11. Cecchini S, Bonardi R, Mazzotta A, Grazzini G, Iossa A, Ciatto S. Testing Cervicography and cervicoscopy as screening tests for cervical cancer. Tumori 1993; 79:22-25.

12. Szarewski A, Cuzick J, Edwards R, Butler B, Singer A. The use of Cervicography in a primary screening service. Br J Obstet Gynaecol 1991; 98:313-317.

13. Soutter WP, Chaves J, Gleeson R, Lim K, Segall S, Skehan M. Cervicography in a colposcopy clinic. J Obstet Gynaecol 1991; 11:218-220.

14. August N. Cervicography for evaluating the "atypical" Papanicolaou smear. J Reprod Med 1991; 36:89-94.

15. Kesic VI, Soutter WP, Sulovic V, Juznic N, Aleksic M, Ljubic A. A comparison of cytology and cervicography in cervical screening. Int J Gynecol Cancer 1993; 3:395-398.

16. Urdaneta H. Detección de patologia cervical no diagnosticada por citologia vaginal mediante cervigrafia. Rev Colomb Obstet Ginecol 1993; 44:220-222.

17. Herrero R, Schiffman MH, Bratti C, Hildesheim A, Balmaceda I, Sherman ME, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: The Guanacaste Project. Pan Am J Public Health 1997; 362-375.

18. Kurman RJ, Soloman D. The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses: Definitions, Criteria, and Explanatory Notes for Terminology and Specimen Adequacy. New York: Springer-Verlag, 1994.

19. Sherman ME, Schiffman MH, Lorincz AT, Manos MM, Scott DR, Kurman RJ, et al. Toward objective quality assurance in cervical cytopathology. Am J Clin Pathol 1994; 102:182-187.

20. Schiffman MH, Liaw KL, Herrero R, Sherman ME, Hildesheim A. Epidemiologic support for a simplified view of cervical carcinogenesis. Eurogin Bulletin 1998; No. 1.

21. International Agency for Research on Cancer. The epidemiology of cervical cancer and human papillomavirus. IARC Scientific Publications no. 119. Lyon, France: International Agency for Research on Cancer, 1992.

22. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Human Papillomaviruses. Volume 64, Lyon, France: International Agency for Research on Cancer Scientific Publications; 1995.

23. Schiffman MH, Brinton L. The epidemiology of cervical carcinogenesis. Cancer Supplement 1995; 76:1888-901.

24. Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J Natl Cancer Inst 1995; 87:796-802.

25. Koss LG. The Papanicolaou test for cervical cancer detection: A triumph and a tragedy. JAMA 1989; 261:737-743.

26. Babes A. Diagnosis of cancer of the uterine cervix by smears. La Presse Médicale 1928; 36:451-454.

27. Douglass LE. A further comment on the contributions of Aurel Babes to cytology and pathology. Acta Cytol 1967; 11:217.

28. Campion MJ, Reid R. Screening for gynecologic cancer. Obstet Gynecol Clin North Am 1990; 17:695-727.

29. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. Int J Cancer 1999; 80:827-841.

30. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide frequency of eighteen major cancers in 1985. Int J Cancer 1993; 54:594-606.

31. Coleman MP, Esteve J, Damiecki P, Arslan A, Renard H. IARC, editor. Trends in cancer incidence and mortality. 1993; Lyon, France: International Agency for Research on Cancer.

32. Beral V, Hermon C, Munoz N, Devesa SS. Cervical Cancer. Trends in Cancer Incidence and Mortality. 1994; 19/20, 265-285.

33. Gustafsson L, Pontén J, Zack M, Adami H-O. International incidence rates of invasive cervical cancer after introduction of cytological screening. Cancer Causes Control 1997; 8:755-763.

34. Gustafsson L, Pontén J, Bergström R, Adami H-O. International incidence rates of cervical cancer before cytologic screening. Int J Cancer 1997; 71:159-165.

35. Herrero R, Brinton L, Reeves WC, Brenes MM, de Britton RC, Gaitan E, et al. Screening for cervical cancer in Latin America: A case-control study. Int J Epidemiol 1992; 21:1050-1056.

36. Hernández-Avila M, Lazcano-Ponce EC, Alonso de Ruiz P, Romieu I. Evaluation of the cervical cancer screening programme in Mexico: A population-based case-control study. Int J Epidemiol 1998; 27:370-376.

37. Sherlaw-Johnson C, Gallivan S, Jenkins D. Evaluating cervical cancer screening programs for developing countries. Int J Cancer 1997; 72:210-216.

38. Maddox P., Szarewski A, Dyson J, Cuzick J. Cytokeratin expression and acetowhite change in cervical epithelium. J Clin Pathol 1994; 47:15-17.

39. van Niekerk WA, Dunton CJ, Richart RM, Hilgarth M, Kato H, Kaufman RH, et al. Colposcopy, cervicography, speculoscopy, and endoscopy. IAC Task Force summary. Acta Cytol 1998; 42:33-49.

40. Cecchini S, Bonardi R, Iossa A, Zappa M, Ciatto S. Colposcopy as a primary screening test for cervical cancer. Tumori 1997; 83:810-813.

41. Lozowski MS, Mishriki Y, Talebian F, Solitare G. The combined use of cytology and colposcopy in enhancing diagnostic accuracy in preclinical lesions of the uterine cervix. Acta Cytol 1982; 26:285-291.

42. Pete I, Tóth V, Bósze P. The value of colposcopy in screening cervical carcinoma. Eur J Gynaecol Oncol 1998; 19:120-122.

43. Kurman RJ, Henson DE, Herbst AL, Noller KL, Schiffman MH. Interim guidelines for management of abnormal cervical cytology. JAMA 1994; 271:(23)1866-1869.

44. Schiffman M, Soloman D, Sherman M. Why, how, and when the cytologic diagnosis of ASCUS should be eliminated. J Lower Genital Tract Disease 1998; 2:165-169.

45. The ALTS group. HPV testing as a triage strategy for women with cytologic evidence of low grade squamous intraepithelial lesions: Early results from a randomized trial on the clinical management of ASCUS and LSIL cytologic abnormalities. J Natl Cancer Inst; In press.

46. Gordis L. Epidemiology. Philadelphia: W.B. Saunders Company, 1996.

47. Miller AB. Fundamental issues in screening for cancer. In: Schottenfeld D, Fraumeni JF, editors. Cancer Epidemiology and Prevention. Second ed. New York: Oxford University Press, 1996:1433-1452.

48. Roland PY, Naumann RW, Alvarez RD, Kilgore LC, Partridge EE. A decision analysis of practice patterns used in evaluating and treating abnormal Pap smears. Gynecol Oncol 1995; 59:75-80.

49. Jaeschke R, Guyatt GH, Sackett DL. Users' guide to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients. JAMA 1994; 271:703-707.

50. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical decision making. Boston: Butterworth-Heinemann, 1988.

51. Choi BCK. Slopes of a receiver operating characteristic curve and likelihood ratios for a diagnostic test. Am J Epidemiol 1998; 148:1127-1132.

52. Morell ND, Taylor JR, Snyder RN, Ziel HK, Slatz A, Willie S. False-negative cytology rates in patients in whom invasive cervical cancer subsequently developed. Obstet Gynecol 1982; 60:41-45.

53. Möbius G. Cytological early detection of cervical carcinoma: Possibilities and limitations. Analysis of failures. Cancer Res Clin Oncol 1993; 119:513-521.

54. DiBonito L, Falconieri G, Tomasic G, Colautti I, Bonifacio D, Dudine S. Cervical cytopathology: An evaluation of its accuracy based on cytohistologic comparison. Cancer 1993; 72:3002-3006.

55. Quantin C, Dusserre L, Montaud AM, Mottot C, Feldman JP. A model for quality assessment in cervical cytology used as a screening test. Qual Assur Health Care 1992; 4:105-113.

56. Parkin DM. Cancer in developing countries. Cancer Surv 1994; 19/20:519-561.

57 Soost H-J, Lange H-J, Lehmacher W, Ruffing-Kullmann B. The validation of cervical cytology: Sensitivity, specificity, and predictive values. Acta Cytol 1991; 35:8-14.

58. DeMay RM. Common problems in Papanicolaou smear interpretation. Arch Pathol Lab Med 1997; 121:229-238.

59. Barton SE, Jenkins D, Hollingworth A, Cuzick J, Singer A. An explanation for the problem of false-negative cervical smears. Br J Obstet Gynaecol 1989; 96:482-485.

60. Schneider DL, Herrero R, Bratti C, Greenberg MD, Hildesheim A, Sherman ME, et al. Cervicography screening for cervical cancer among 8,460 women in a high-risk population. Am J Obstet Gynecol 1999; 180:290-298.

61. Lazcano-Ponce EC, Alonso de Ruiz P, López Carrillo L, Nájera Aguilar P, Avila Ceniceros R, Escandón Romero C, et al. Validity and reproducibility of cytologic diagnosis in a sample of cervical cancer screening centers in Mexico. Acta Cytol 1997; 41:277-284.

62. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. Am J Epidemiol 1995; 141:680-689.

63. Dewar MA, Hall K, Perchalski J. Cervical cancer screening: Past success and future challenge. Prim Care 1992; 19:589-606.

64. Schiffman MH, Brinton L, Devesa SS, Fraumeni Jr. JF. Cervical cancer. In: Schottenfeld D, Fraumeni JF, editors. Cancer Epidemiology and Prevention. Second ed. New York: Oxford University Press, 1996:1090-1116.

65. Boronow RC. Death of the Papanicolaou smear? A tale of three reasons. Am J Obstet Gynecol 1998; 179:391-396.

66. Davey DD, Nielsen ML, Frable WJ, Rosenstock W, Lowell DM, Kraemer BB. Improving accuracy in gynecologic cytology. Arch Pathol Lab Med 1993; 117:1193-1198.

67. Klinkhamer PJJM, Vooijs GP, de Haan AFJ. Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears. Acta Cytol 1988; 32:794-800.

68. de Ruíz PA, Lazcano Ponce EC, Duarte Torres R, Ruíz Juárez I, Martínez Cortez I. Diagnostic reproducibility of Pap testing in two regions of Mexico: The need for quality control mechanisms. Bull Pan Am Health Organ 1996; 30:330-338.

69. Duca P, Braga M, Chiappa L, Piffer R, Turolla E. Intralaboratory reproducibility of interpretation of Pap smears: Results of an experiment. Tumori 1988; 74:737-734.

70. Confortini M, Biggeri A, Cariaggi MP, Carozzi FM, Minuti PA, Russo A, et al. Intralaboratory reproducibility in cervical cytology: results of the application of a 100-slide set. Acta Cytol 1993; 37:49-54.

71. Cocchi V, Carretti D, Fanti S, Baldazzi P, Casotti MT, Piazzi R, et al. Intralaboratory quality assurance in cervical/vaginal cytology: Evaluation of intercytologist diagnostic reproducibility. Diagn Cytopathol 1997; 16:87-92.

72. Cocchi V, Sintoni C, Carretti D, Sama D, Chiari U, Segala V, et al. External quality assurance in cervical/vaginal cytology: Interlaboratory agreement in the Emilia-Romagna region of Italy. Acta Cytol 1996; 40:480-488.

73. Joste NE, Rushing L, Granados R, Zitz JC, Genest DR, Crum CP, et al. Bethesda classification of cervicovaginal smears: Reproducibility and viral correlates. Hum Pathol 1996; 27:581-585.

74. Kern WH, Zivolich MR. The accuracy and consistency of the cytologic classification of squamous lesions of the uterine cervix. Acta Cytol 1977; 21:519-523.

75. Jones S, Thomas GDH, Williamson P. Observer variation in the assessment of adequacy and neoplasia in cervical cytology. Acta Cytol 1996; 40:226-234.

76. Curry H, Thompson DW, Dietrich M, Lipa M, Massarella GR, Taves IR, et al. Proficiency testing in cytology laboratories in Ontario, Canada: A decade of experience. III. A precision study of consistency and reproducibility in cytology reporting. Acta Cytol 1987; 31:220-225.

77. Sherman ME, Kurman RJ. The role of exfoliative cytology and histopathology in screening and triage. Obstet Gynecol Clin North Am 1996; 23:(3)641-655.

78 Davey DD. Quality and liability issues with the Papanicolaou smear: The problem of definition of errors and false-negative smears. Arch Pathol Lab Med 1997; 121:267-269.

79 Genak LJ, Schumann GB. Chain of events in Papanicolaou testing: Impact on quality assurance. Diagn Cytopathol 1989; 5:221-227.

80 Jones B. Rescreening in gynecologic cytology. Arch Pathol Lab Med 1996; 120:519-522.

81 Henry JA. Influence of smear quality on the rate of detecting significant cervical cytologic abnormalities. Acta Cytol 1996; 40:529-535.

82 Sherman ME, Schiffman MH, Herrero R, Kelly D, Bratti C, Mango L, et al. Performance of a semiautomated Papanicolaou smear screening system: results of a population-based study conducted in Guanacaste, Costa Rica. Cancer 1998; 84:273-280.

83 Sherman ME, Schiffman MH, Lorincz AT, Herrero R, Hutchinson ML, Bratti C, et al. Cervical specimens collected in liquid buffer are suitable for both cytologic screening and ancillary human papillomavirus testing. Cancer Cytopathol 1997; 81:89-97.

84 Lieu D. The Papanicolaou smear: Its value and limitations. J Fam Pract 1996; 42:391-399.

85 Spitzer M. Cervical screening adjuncts: Recent advances. Am J Obstet Gynecol 1998; 179:544-556.

86. Franco E, Syrjänen K, de Wolf C, Patnick J, Ferenczy A, McGoogan E, et al. New developments in cervical cancer screening and prevention. Cancer Epidemiology, Biomarkers, and Prevention 1996; 5:853-856.

87. Singer A. Cervical cancer screening: state of the art. Bailliere's Clinical Obstetrics and Gynaecology 1995; 9:39-64.

88. O'Leary TJ, Tellado M, Buckner SB, Ali IS, Stevens A, Ollayos CW. PAPNET-Assisted rescreening of cervical smears: Cost and accuracy compared with a 100% manual rescreening strategy. JAMA 1998; 279:235-237.

89. Raab SS, Zaleski MS, Silverman JF. The cost-effectiveness of the cytology laboratory and new cytology technologies in cervical cancer prevention. Am J Clin Pathol 1999; 111:259-266.

90. Brown AD, Garber AM. Cost-effectiveness of 3 methods to enhance the sensitivity of Papanicolaou testing. JAMA 1999; 281:347-353.

91. Wright TC, Kurman RJ. A critical review of the morphologic classification systems of preinvasive lesions of the cervix: the scientific basis for shifting the paradigm. Papillomavirus Report 1994; 5:175-182.

92. Schiffman MH, Sherman ME. HPV testing to improve cervical cancer screening. In: Srivastava S, Lippman SM, Hong WK, Mulshine JL, editors. Early detection of cancer: Molecular markers. Armonk: Futura Publishing Co., 1994:265-277.

93. Cox JT. Clinical role of HPV testing. Obstet Gynecol Clin North Am 1996; 23:811-851.

94 Greenberg M, Sedlacek TV, Campion MJ. Cervical neoplasia: Are adjunctive tests to cervical cytology worthwhile? Clin Obstet Gynecol 1995; 38:600-609.

95. Greenberg MD, Campion MJ, Rutledge LH. Cervicography as an adjunct to cytologic screening. Obstet Gynecol Clin North Am 1993; 20:13-29.

96. Jaeschke R, Guyatt G, Sackett DL. Users' guide to the medical literature: III. How to use an article about a diagnostic test. A. Are the results of the study valid? JAMA 1994; 271:389-391.

97 Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299:926-930.

98. Greenes RA, Begg CB. Assessment of diagnostic technologies: Methodology for unbiased estimation from samples of selectively verified patients. Invest Radiol 1985; 20:751-756.

99. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39:207-215.

100. Zhou X-H. Effect of verification bias on positive and negative predictive values. Stat Med 1994; 13:1737-1745.

101. Limburg H. Comparison between cytology and colposcopy in the diagnosis of early cervical carcinoma. Am J Obstet Gynecol 1958; 75:1298-1301.

102. Stafl A. Cervicography. Clin Obstet Gynecol 1983; 26:1007-1016.

103. Gundersen JH, Schauberger CW, Rowe NR. The Papanicolaou smear and the Cervigram: A preliminary report. J Reprod Med 1988; 33:46-48.

104. Rehder KE, Blythe JG. Cytology and Cervicography compared to cytology alone for human papillomavirus detection. Colposcopy and Gynecologic Laser Surgery 1988; 4.1-7.

105. Blythe JG. Cervicography: A preliminary report. Am J Obstet Gynecol 1985; 152:192-197.

106. Mitchell H. Is there a role for screening by cervicography in Australia? Med J Aust 1989; 150:441-444.

107. Spitzer M, Krumholz BA, Seltzer V, Molho L. Cervical cancer detected by cervicography in a patient with negative cervical cytology. Obstet Gynecol 1986; 68:68S-70S.

108. Schatzkin A, Connor RJ, Taylor PR, Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. Am J Epidemiol 1987; 125:672-678.

109. Han SK, Lee JW. Significance of Cervicography in the early diagnosis of cervical neoplasia. The Journal of Catholic Medical College 1988; 41:298-307.

110. Coibion M, Bercham G, Hertens D, Miloiu A, Andry M; Guenier C, et al. Belgian multicentric study on Pap smear and Cervicography. 1990; ESSO 90 Satellite Symposium, Belgium.

111. Lörincz AT. Molecular methods for the detection of human papillomavirus infection. Obstet Gynecol Clin North Am 1996; 23:707-730.

112. Cecchini S, Iossa A, Bonardi R, Gustavino C, Ciatto S. Evaluation of the sensitivity of cervicography in a consecutive colposcopic series. Tumori 1992; 78:211-213.

113. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991.

114. McKinnon KJ, Ford RM, Hunter JC. Comparison of cytology and cervicography in screening a high risk Australian population for cervical human papillomavirus and cervical intraepithelial neoplasia. Aust NZ J Obstet Gynaecol 1993; 33:176-179.

115. De Sutter P, Coibion M, Vosse M, Hertens D, Huet F, Wesling F, et al. A multicentre study comparing cervicography and cytology in the detection of cervical intraepithelial neoplasia. Br J Obstet Gynaecol 1998; 105:613-620.

116. Autier P, Coibion M, De Sutter P, Wayemberg M. Cytology alone versus cytology and cervicography for cervical cancer screening: A randomized study. Obstet Gynecol 1999; 93:353-358.

117. Jones DED, Creasman WT, Dombroski RA, Lentz SS, Waeltz JL. Evaluation of the atypical pap smear. Am J Obstet Gynecol 1987; 157:544-549.

118. Spitzer M, Krumholz BA, Chernys AE, Seltzer V, Lightman AR. Comparative utility of repeat Papanicolaou smears, Cervicography, and colposcopy in the evaluation of atypical Papanicolaou smears. Obstet Gynecol 1987; 69:731-735.

119. Ferris DG, Payne P, Frisch LE. Cervicography: An intermediate triage test for the evaluation of cervical atypia. J Fam Pract 1993; 37:463-468.

120. Eskridge C, Begneaud WP, Landwehr C. Cervicography combined with repeat Papanicolaou test as a triage for low-grade cytologic abnormalities. Obstet Gynecol 1998; 92:351-355.

121. Singh V, Sehgal A, Luthra UK. Screening for cervical cancer by direct inspection. BMJ 1992; 534-535.

122. Nene BM, Deshpande S, Jayant K, Budukh AM, Dale PS, Deshpande DA, et al. Early detection of cervical cancer by visual inspection: A population-based study in rural India. Int J Cancer 1996; 68:770-773.

123. Wesley R, Sankaranarayanan R, Mathew B, Chandralekha B, Aysha Beegum A, Amma NS, et al. Evaluation of visual inspection as a screening test for cervical cancer. Br J Cancer 1997; 75:436-440.

124. Megevand E, Denny L, Dehaeck K, Soeters R, Bloch B. Acetic acid visualization of the cervix: An alternative to cytologic screening. Obstet Gynecol 1996; 88:383-386.

125. Frisch LE, Milner FH, Ferris DG. Naked-eye insepction of the cervix after acetic acid application may improve the predictive value of negative cytologic screening. J Fam Pract 1994; 39:457-460.

126. University of Zimbabwe/JHPIEGO Cervical Cancer Project. Visual inspection with acetic acid for cervical-cancer screening: Test qualities in a primary-care setting. Lancet 1999; 353:869-873.

127. Wertlake PT, Francus K, Newkirk GR, Parham GP. Effectiveness of the Papanicolaou smear and speculoscopy as compared with the Papanicolaou smear alone: A community-based clinical trial. Obstet Gynecol 1997; 90:421-427.

128. Edwards G, Rutowski C, Palmer C. Cervical cancer screening with Papanicolaou smear plus Speculoscopy by nurse practitioners in a Health Maintenance Organization. J Lower Genital Tract Disease 1997; 1:141-147.

129. Mann W, Lonky N, Massad S, Scotti R, Blanco J, Vasilev S. Papanicolaou smear screening augmented by a magnified chemiluminescent exam. Int J Gynecol Obstet 1993; 43:289-296.

130. Massad S, Lonky NM, Mutch DG, Blanco JS, Vasilev SA, Finan MA, et al. Use of Speculoscopy in the evaluation of women with atypical Papanicolaou smears: Improved cost effectiveness by selective colposcopy. J Reprod Med 1993; 38:163-169.

131. Loiudice L, Abbiati R, Boselli F, Cecchini G, Costa S, Grossi E, et al. Improvement of Pap smear sensitivity using a visual adjunctive procedure: A co-operative Italian study on speculoscopy (GISPE). Eur J Cancer Prev 1998; 7:295-304.

132. Lonky NM, Edwards G. Comparison of chemiluminescent light versus incandescent light in the visualization of acetowhite epithelium. The American Journal of Gynecologic Health 1992; 6:11-15.

133. Lonky NM, Mann WJ, Massad LS, Mutch DG, Blanco JS, Vasilev SA, et al. Ability of visual tests to predict underlying cervical neoplasia: Colposcopy and Speculoscopy. J Reprod Med 1995; 40:530-536.

134. Schneider A, Zahm DM. New adjunctive methods for cervical cancer screening. Obstet Gynecol Clin North Am 1996; 23:657-673.

135. Mould TAJ, Singer A. Adjuvant tests to cytology: Cervicography and the Polarprobe. In: Franco E, Monsonego J, editors. New Developments in Cervical Cancer Screening and Prevention. Oxford: Blackwell Science Ltd., 1997:406-410.

136. Poljak M, Brencic A, Seme K, Vince A, Marin IJ. Comparative evaluation of first- and second-generation Digene hybrid capture assays for detection of human papillomaviruses associated with high or intermediate risk for cervical cancer. J Clin Microbiol 1999; 37:796-797.

137. Schiffman M, Herrero R, Hildesheim A, Sherman ME, Bratti C, Wacholder S, et al. HPV DNA testing for cervical cancer screening: Results from the 9,000 women in the NCI Guanacaste Project. JAMA 2000; 283:87-93.

Table 1

Terminology used to describe cervical neoplasia						
Normal	Low grade squamous intraepithelial lesions (LSIL)		High grade squamous intraepithelial lesions (HSIL)			Invasive cancer
	Koilocytotic / condylomatous atypia	Cervical Intraepithelial Neoplasia (CIN) 1	CIN 2	CIN 3		
		Very mild dysplasia / Mild dysplasia	Moderate dysplasia	Severe dysplasia	Carcinoma *in situ*	

Source: Schiffman MH, Brinton L, Devesa SS, Fraumeni JR JF. Cervical cancer. In Schottenfeld D, Fraumeni JF, editors. Cancer Epidemiology and Prevention. Second Ed. New York: Oxford University Press 1996:1090–1116.

Table 2

Contingency Table for Evaluating the Performance of a Screening Test

		Referent Diagnosis		
		Disease	**No Disease**	
Screening Test	+	a	b	a+b
	–	c	d	c+d
		a+c	b+d	a+b+c+d

Sensitivity = a/a+c = True Positive Rate

Specificity = d/b+d = True Negative Rate

False Negative Rate = c/a+c = 1–Sensitivity

False Positive Rate = b/b+d = 1–Specificity

Positive Predictive Value = a/a+b

Negative Predictive Value = d/c+d

Proportion Referred = a+b/a+b+c+d

* These estimates are often expressed as a percent by multiplying the estimate by 100.

Table 3

How test referral bias affects estimates of sensitivity, specificity, and negative predictive value

Test referral bias present, where subjects who test positive by the screening test receive the confirmatory test and all others are assumed not to have the disease:

		Referent Diagnosis		
		Disease	**No Disease**	
Screening Test	+	a	b	a+b
	–	c	d+x	c+(d+x)
		a+c	b+(d+x)	

Where

(+) indicates positive by the initial screening test;

(-) indicates negative by the initial screening test;

a, b, c, and d are frequencies; and

x = the number of cases with true disease who were misclassified as negative by the initial screening test, subsequently not included in the referral sample, and therefore misclassified as "not diseased"

No test referral bias

		Referent Diagnosis		
		Disease	**No Disease**	
Screening Test	+	a	b	a+b
	–	c+x	d	(c+x)+d
		a+(c+x)	b+d	

Table 4

Consequences of test referral bias*

	With test referral bias	True estimate
Sensitivity	a/(a+c)	a/[a+(c+x)]
Specificity	(d+x)/[b+(d+x)]	d/(b+d)
False Positive Rate	b/[b+(d+x)]	b/(b+d)
Negative Predictive Value	(d+x)/[c+(d+x)]	d/(c+x)+d

* Estimates are based on Table 3

Where

a, b, c, and d are frequencies; and

x = the number of cases with true disease who were misclassified as negative by the initial screening test, subsequently not included in the referral sample, and therefore misclassified as "not diseased"

Table 5

Example of the effects of test referral bias on estimates of the performance of a screening test in a hypothetical cohort of 1100 subjects,
assuming the "true" sensitivity and specificity as indicated in the x=0 condition

x = 0

		Disease (Gold Standard Test)				
		+	-	Total		
Screening	+	100	100	200	Sensitivity = 100/200 =	50.0%
Test	-	100	800	900	Specificity = 800/900 =	88.9%
Total		200	900	1100	Predictive value (-) = 800/900 =	88.9%

x = 1

		Disease (Gold Standard Test)				
		+	-	Total		
Screening	+	100	100	200	Sensitivity = 100/199 =	50.3%
Test	-	99	801	900	Specificity = 801/901 =	88.9%
Total		199	901	1100	Predictive value (-) = 801/900 =	89.0%

x = 50

		Disease (Gold Standard Test)				
		+	-	Total		
Screening	+	100	100	200	Sensitivity = 100/150 =	66.7%
Test	-	50	850	900	Specificity = 850/950 =	89.5%
Total		150	950	1100	Predictive value (-) = 850/900 =	94.4%

x = 100

		Disease (Gold Standard Test)				
		+	-	Total		
Screening	+	100	100	200	Sensitivity = 100/100 =	100.0%
Test	-	0	900	900	Specificity = 900/1000 =	90.0%
Total		100	1000	1100	Predictive value (-) = 900/900 =	100.0%

x = the number of cases with true disease who were misclassified as negative by the initial screening test, subsequently not included in the referral sample, and therefore misclassified as "not diseased"

(+) indicates a positive test

(-) indicates a negative test

Table 6

Cervigram Classification Scheme

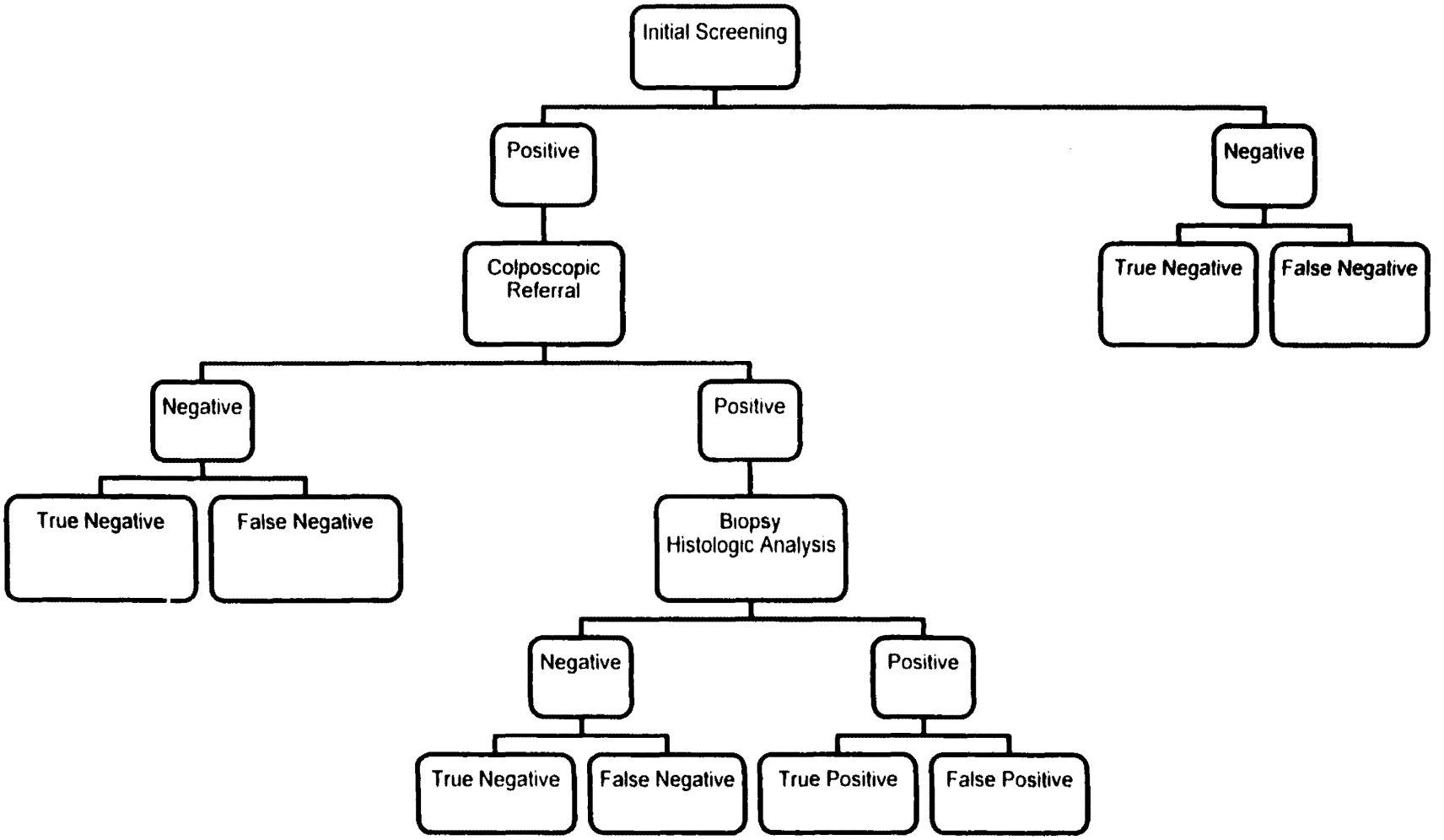
Do not refer for colposcopy:

Negative:	No lesion seen
Atypical 1 (A1):	A trivial lesion inside the transformation zone is visible, but colposcopy not recommended because of the benign appearance or site of the lesion
Atypical 2 (A2):	A trivial lesion outside the transformation zone is visible, but colposcopy not recommended because of the benign appearance or site of the lesion
Technically defective:	Unable to be properly evaluated

Refer for colposcopy:

Positive 0 (P0):	Probably normal, but colposcopy preferable to rule out serious neoplasia
Positive 1A (P1A):	Compatible with trivial disease but colposcopy recommended because part of the lesion extends into the canal
Positive 1B (P1B):	Compatible with a low-grade lesion/ flat condyloma/ and exophytic condyloma
Positive 2 (P2):	Compatible with a high grade lesion
Positive 3 (P3):	Compatible with cancer

Figure 1
Screening, Referral, and Diagnosis Protocol



Appendix A

Calculation of Magnification Measurements[a]

A. Magnification of the cervigram slide on the screen

W * 0.7/3.6 = M

Where:

W = Width of the projected image on the screen in centimeters
M = Magnification on the screen
0.7 = the image reduction on the slide
3.6 = the width of the slide

B. Apparent magnification of the cervigram slide on the screen observed from a distance

(W * 0.7/3.6) * 30/D = M

Where:

W = Width of the screen in centimeters
D = Distance from the screen in centimeters
M = Magnification on the screen
30 = Distance between the eye and the object

C. Calculation of the actual size of the lesion

S/M = A

Where:

S = Size in cm as measured on the screen
M = Magnification as determined by calculation A above
A = Actual size

a. Calculations are made assuming that the room setup meets the specifications provided by National Testing Laboratories Worldwide. Under these specifications, the projector and lens must project an image that can fill the entire width of a screen which can accommodate at least a 5-foot wide projection. The screen must be flat white without granularity; the screen must not be lenticular. The evaluator chair is usually 3 to 6 feet from the screen, though he or she may occasionally walk closer toward the screen.

Source: National Testing Laboratories Worldwide

First Manuscript

Cervicography Screening for Cervical Cancer among 8460 Women in a High-Risk Population*

Diana L. Schneider

This project was funded by the National Cancer Institute (NCI), contract nos. N01-CP-21081 and N01-CP-31061.

* A shorter version of this manuscript was published in the *American Journal of Obstetrics and Gynecology* 1999;180:290-298.

Abstract

OBJECTIVE: Cervicography was evaluated as a primary screening method for cervical cancer.

STUDY DESIGN: Cervigrams of 8460 women were taken upon enrollment into a population-based study of cervical neoplasia. Cervicography results were compared with a referent diagnosis determined by histology and three cytologic tests, conventional cytology, and presence of cancer-associated human papillomavirus types.

RESULTS: Cervicography identified all 11 cancers, whereas cytology missed one. Cervicography yielded sensitivities for detecting high grade squamous intraepithelial lesions or cancer of 49.3% overall (specificity = 95.0%), 54.6% in women younger than age 50, and 26.9% in women ages 50 and older. Cytology yielded sensitivities for detecting high grade squamous intraepithelial lesions or cancer of 77.2% overall (specificity = 94.2%), 75.5% in women younger than age 50, and 84.6% in women ages 50 and older.

CONCLUSIONS: Cervicography performed marginally better than cytology for the detection of invasive cervical cancer. Cytology performed better than cervicography for the detection of high grade squamous intraepithelial lesions. Cervicography might best be performed in premenopausal women.

Introduction

Cervical cancer incidence and mortality have markedly declined over recent decades in countries implementing widespread cytologic screening using the Papanicolaou smear.[1,2] However, the Papanicolaou smear, like any other screening method, is not a perfect tool. Cytologic screening failures may reflect inadequate sampling, poor slide preparation, or errors in microscopic screening or classification.[3-7] Concerns about errors in conventional Papanicolaou smears have motivated some researchers to evaluate alternate and/or adjunctive screening methods.[8-12]

The present analysis was conducted to provide a rigorous and independent evaluation of Cervicography™ (National Testing Laboratories Worldwide, Fenton, MO) as a primary screening method for early identification and prevention of cervical cancer. Cervicography is a visual screening method which evaluates the macroscopic appearance of the cervix rather than the cytologic appearance of exfoliated cells. In cervicography, a trained health care provider or technician takes two high resolution photographic images (Cervigrams™) of the cervix after applying a 5% acetic acid wash. The cervigrams are interpreted by expert evaluators who classify the images using modified colposcopic criteria as the basis for referral for colposcopy. Magnification is achieved by projection of the 35mm slide onto a screen.

Cervicography may have particularly important applications as a primary screening tool for early detection of cervical cancer in countries where specialized medical and laboratory expertise is concentrated in few urban areas and/or where

Papanicolaou smear reliability is problematic. Cervicography allows a wide variety of health professionals to take photographic images of the cervix in a standardized, controlled fashion. The film may then be sent elsewhere for developing and rapid, expert interpretation. Previous studies of cervicography showed mixed results and were limited by their inclusion of selected populations; small sample size; and earlier, non-specific criteria for diagnostic classification.[13-24]

This study of cervicography was conducted as part of a large, population-based cohort study of the natural history of cervical neoplasia conducted in the province of Guanacaste, Costa Rica, sponsored by the National Cancer Institute. The Guanacaste site was selected because of its consistently high rates of cervical cancer despite existing Papanicolaou smear screening services.[25] In this analysis, 8460 cervigram results are compared with a referent standard "referent diagnosis" based on histology, three cytologic methods, and cervicography. Cervicography is also compared with conventional cytologic screening and HPV DNA testing by Hybrid Capture.

Materials and Methods

The study design is described in greater detail elsewhere.[25] The protocol for this study was approved by the Institutional Review Boards of the National Institutes of Health, the Costa Rican Social Security Institute, and the Uniformed Services University of the Health Sciences.

Subject Selection and Participation Rates

The cohort was enumerated in February and March 1993 and enrolled between June 1993 and December 1994. The population was recruited to be a true, door-to-door census of adult women in Guanacaste. A total of 11,742 women was selected who met the initial eligibility criterion of being 18 years of age or older by July 1, 1993. Of those, 10,738 (91.5%) were eligible for an interview. Women who were not Guanacaste residents, or were mentally incompetent, physically incapacitated, unable to understand Spanish, or deceased were excluded.

Study clinics were established at regional hospitals or temporarily at dozens of local health outposts. Potential participants provided informed consent and participated in a standardized interview before examination. Pregnant women were deferred until three months after giving birth. A total of 10,049 of the 10,738 eligible women (93.6%) was interviewed. Pelvic examinations were not performed on the 583 virgins who completed the enrollment interview. The remaining 9466 (94.2% of those interviewed) were eligible for an enrollment pelvic examination. Of those, 291 women either refused or were physically unable to undergo the pelvic examination. The pelvic examination was performed on 9175 women and a referent diagnosis was established for all of these. One woman did not have screening tests performed at the time of her pelvic examination because of known cervical cancer. Cervigrams were obtained for 9062 women, corresponding to 98.8% of women completing the pelvic examination. There were 602 women with cervigrams (6.6%) who reported having had a hysterectomy who were

subsequently excluded from the analyses, leaving 8460 participants available for the present study.

Clinical Specimens

After obtaining informed consent, patients participated in a detailed interview assessing risk factors for cervical neoplasia and then underwent a routine pelvic examination. Cervical cytologic and human papillomavirus (HPV) DNA specimens were collected during the pelvic examinations using a Cervex Brush (Unimar, CT). The quality of cytology was optimized by careful clinician training, a strict fixation and staining protocol, and site visits by U.S. experts in cytotechnology and cytopathology. Two types of cytologic preparations were made for each participant, including a conventional Papanicolaou smear (read by local pathologists and by a computer-assisted method, see below) and a ThinPrep (Cytyc, MA). Conventional Papanicolaou smears were fixed with Pap Perfect (Medscand, FL), and stained by the Papanicolaou method in Costa Rica. After the smear was made, the Cervex brush was rinsed in 20 ml of PreservCyt (Cytyc, MA). The vials containing the PreservCyt solution were sent to the United States where ThinPreps were made.

Additional cervical cells were collected using a Dacron swab which was placed in Specimen Transport Medium (Digene Corporation, MD), frozen, and shipped to the United States for HPV DNA testing using the first generation Hybrid Capture tube test (Digene Corporation, MD). Samples were tested for HPV positivity at the 10 pg/ml level[26] using probes for HPV types 16, 18, 45, and 56 (high-risk, cancer-associated

types): 31, 33, 35, 39, 51, 52, and 58 (intermediate-risk, cancer-associated types); and 6, 11, 42, 43, and 44 (low-risk types).[27,28]

The cervix was then rinsed with 5% acetic acid and two photographic images of the cervix (Cervigrams) were taken using a Cerviscope™ (National Testing Laboratories Worldwide, MO). The undeveloped film was sent to the United States for developing, processing (National Testing Laboratories Worldwide, MO), and evaluation.

Cervigram Review

Cervigrams were interpreted by certified evaluators (Mitchell Greenberg, Thomas Sedlacek, Michael Campion) and classified according to the diagnostic criteria approved by NTL (Table 1).

Cytologic Diagnosis

Cytologic classification was made using three methods: Conventional Papanicolaou smear; the PapNet system [Neuromedical Systems, Inc., Suffern, NY (now TriPath, Elon, NC)], which uses the same slide as the Papanicolaou smear; and ThinPrep. Conventional Papanicolaou smears were interpreted in Costa Rica (Mario Alfaro). ThinPreps were prepared and interpreted in the United States (Martha Hutchinson). After interpretation in Costa Rica, all available conventional Papanicolaou smear slides were sent to Neuromedical Systems, Inc. for repeat screening using the PapNet system, a neural network-based, semi-automated device. The resulting images of each slide were stored on digital tapes which were reviewed by an experienced cytotechnologist. The cytotechnologist rescreened all possibly abnormal smears microscopically and referred

slides that were not considered negative to a pathologist for assignment of the referent diagnosis (Mark Sherman). Further details of the PapNet review are described elsewhere.[10] Conventional Papanicolaou smear, ThinPrep, and PapNet results were classified according to The Bethesda System as negative (within normal limits or reactive cellular changes), atypical squamous cells of undetermined significance, low grade squamous intraepithelial lesion, high grade squamous intraepithelial lesion, or carcinoma.[29]

Colposcopic Referral, Biopsy, and Management

Participants were referred for colposcopy if 1) physical examination was suspicious for cancer, 2) there was an abnormal (atypical squamous cells of undetermined significance or more severe) cytologic result by any of the three methods, or 3) there was a positive cervigram (positive 0, positive 1, positive 2, or positive 3). Colposcopy was performed by a single gynecologist (Jorge Morales). Biopsies of lesions visualized on colposcopy were fixed in 10% buffered formalin, embedded in paraffin, stained with Haematoxylon and eosin, and diagnosed in Costa Rica for clinical purposes. All histologic material, including biopsies and subsequent cone biopsies and hysterectomies were sent to the United States for review and assignment of a final study diagnosis (Mark Sherman).

Participants with histologically-confirmed high grade squamous intraepithelial lesions or cancer or with a highly suspect diagnosis of high grade squamous intraepithelial lesions (by at least two cytologic methods) were referred for treatment by

large loop excision of the transformation zone (LLETZ or LEEP), cold knife cone, or hysterectomy as appropriate.[25] Treatment was provided through the Costa Rican Social Security system.

As a quality control measure, a random sample of women with all negative screening results was referred to colposcopy to validate the screening protocol compared with colposcopy results. All 144 of these referral controls had a referent diagnosis of normal, indicating 100% sensitivity of the screening protocol.

Referent Diagnosis

Referent diagnoses were made based on histology, cytology, and cervicography results and classified as indicated in Table 2, with subsequent combinations as required for the data analysis.

Data Analysis

All analyses were limited to the 8460 non-hysterectomized women with available cervigram results. Sensitivity and specificity of cervicography were calculated with the referent diagnosis as the referent standard. Sensitivity is defined as the proportion of participants with "disease" who are diagnosed by the screening method under evaluation. Specificity is defined as the proportion of participants without "disease" who are correctly identified as disease-free using the screening method under evaluation. Two different definitions of disease, based on referent diagnoses, were used as targets for screening: 1) disease is high grade squamous intraepithelial lesions or cancer (versus normal, equivocal, or low grade squamous intraepithelial lesions) and 2) disease is low grade

squamous intraepithelial lesions, high grade squamous intraepithelial lesions, or cancer (versus normal or equivocal). The definition of a "positive screen" using cervicography also requires a choice of thresholds. Three different thresholds to define a positive cervigram result were examined for analytical purposes, including 1) positive is atypical or more severe (versus negative), 2) positive is positive 0, positive 1, positive 2, or positive 3 (versus negative or atypical), and 3) positive is positive 0, positive 2, or positive 3 (versus negative, atypical, or positive 1).

The analysis focuses on detection of high grade squamous intraepithelial lesions or cancer because of the greater clinical importance of these lesions. Analyses considering a referent diagnosis of low grade squamous intraepithelial lesions as "diseased" are presented, but not discussed in detail. When presenting detection of low grade squamous intraepithelial lesions, we refer to "percent of patients referred to colposcopy" rather than "specificity" because, given what is currently known about the dynamic nature of HPV infection,[30] diagnostic techniques are not available to adequately confirm that no low grade squamous intraepithelial lesion is present. Percent referred provides us with a proportion which will be correlated with the false positive rate and will permit trade-offs with sensitivity.

Analyses of sensitivity and specificity were conducted using standard contingency table methods. Contingency tables were stratified by age and factors related to pregnancy history, menopausal status, smoking history, and oral contraceptive use to assess whether any of these factors affected the performance of cervicography.

Additionally, cervicography was directly compared with conventional cytologic screening based on the threshold level for colposcopic referral (any positive cervigram, versus atypical squamous cells of undetermined significance or more severe for cytology). Discordant results are described by the referent diagnosis and by an independent measure, the presence of cancer-associated human papillomavirus types (high- and intermediate-risk). Differences between the proportions of women referred for colposcopy by cervicography and cytology among those with a referent diagnosis of high grade squamous intraepithelial lesions or cancer, and among women testing positive for cancer-associated HPV types, were assessed for statistical significance using the McNemar's test for paired data.

Technically defective cervigrams are presented as a separate category of cervigram result in most analyses. In analyses of sensitivity and specificity, technically defective cervigrams were classified as negative in order to provide the most conservative estimates of sensitivity. Similarly, unsatisfactory cytology results were recoded as negative for analyses of sensitivity and specificity to provide an appropriate comparison. Results of analyses comparing discordant diagnoses by cervicography and conventional cytology, are presented with technically defective cervigrams and unsatisfactory cytology excluded.

Results

Cervicography screening compared with the referent diagnosis

The distribution of referent diagnoses by cervigram results are presented in Table 3. Of the 8460 women who had a cervigram, 484 (5.7%) had a positive cervigram result and were referred to colposcopy (although colposcopic referrals from the three cytologic techniques overlapped with cervicography and each other). Of the remaining women, 7010 (82.9%) had a negative cervigram, 853 (10.1%) had an atypical cervigram result, and 113 (1.3%) had a technically defective cervigram.

All 11 (100%) of the 11 cases of invasive cancer were detected by cervicography. Of these critical cases, 3 (27.3%) had a cervigram result of positive 0, 3 (27.3%) had a cervigram result of positive 2, and 5 (45.5%) had a cervigram result of positive 3. No cancers in this population were detected with a cervigram result of positive 1. In 6 of the 11 (54.5%) women with cancer, a grossly visible lesion was identified on physical examination which in itself prompted colposcopy referral.

Of the 125 cases of high grade squamous intraepithelial lesions on referent diagnosis, 117 were histologically-confirmed and 8 were confirmed by two cytologic diagnoses only (HSIL2). Of the histologically-confirmed cases of high grade squamous intraepithelial lesions, 56 (47.9%) were detected by cervicography, including 2 (1.7%) with a positive 0 cervigram, 5 (4.3%) with a positive 1A cervigram, 32 (27.4%) with a positive 1B cervigram, 12 (10.3%) with a positive 2 cervigram, and 5 (4.3%) with a positive 3 cervigram. Of the 56 histologically-confirmed cases of high grade squamous

intraepithelial lesions detected by cervicography, 8 (14.3%) were missed by all three cytologic screening methods.

There were 8 cases of high grade squamous intraepithelial lesions not confirmed by histology, none of which were detected by cervicography. This low positivity of cervicography for the cytology-only cases of high grade squamous intraepithelial lesions was significantly lower than the 47.9% positivity among the histologically-confirmed cases (Fisher's Exact Test, $p = 0.008$). Colposcopy examination revealed that the transformation zone was not visible in 6 of these 8 women (75.0%).

Of the 191 cases of low grade squamous intraepithelial lesions, 74 were histologically confirmed. As shown in Table 3, the detection of low grade squamous intraepithelial lesions by cervicography was again significantly greater for the histologically confirmed cases (58.1%) than the cases which were not histologically confirmed (20.9%, Fisher's Exact Test, $p < 0.001$). Colposcopy notes indicated that the transformation zone was not visible in 6 of the 24 women (25.0%) with non-histologically confirmed low grade squamous intraepithelial lesions referred by cervicography.

Sensitivity and Specificity of cervicography compared with the referent diagnosis

All calculations of sensitivity and specificity were done using the referent diagnosis as the referent standard of "disease". In Table 4, disease is defined as high grade squamous intraepithelial lesions or cancer.

Sensitivity and specificity of cervicography for high grade squamous intraepithelial lesions and cancer, where a positive cervigram is defined by any positive cervigram

As shown in Table 4, considering the referral threshold for cervicography of any positive (positive 0-positive 3), the overall sensitivity for high grade squamous intraepithelial lesions and cancer was 49.3% (95% C.I.: 40.9%, 57.7%), the specificity was 95.0% (95% C.I.: 94.5%, 95.5%), and the positive predictive value was 13.8% (95% C.I.: 10.8%, 16.9%). The sensitivity fluctuated between 36% and 75% in women 49 years of age and younger. A drop in sensitivity was observed in women ages 50 and above, to 30.0% in women ages 50-64 and 25.0% in women ages 65 and older. When women were categorized as younger than age 50 and ages 50 and older, the sensitivities of cervicography were 54.6% and 26.9% respectively ($\chi^2 = 6.4$, $p = .01$).

The reduction in sensitivity in the older age groups was related to menopausal status. Among all women, cervicography yielded a sensitivity of 30.0% in women no longer having menstrual periods, compared with a sensitivity of 54.7% in women still having menstrual periods. In stratified multivariable analyses, menopausal status appeared to be the major predictor of cervicography sensitivity, to the extent that the closely associated variables of age and menopausal status could be disentangled. The performance of cervicography did not appear to be meaningfully affected by hormonal contraceptive use, smoking, or having a previous abnormal Papanicolaou smear once menopausal status was taken into account.

Because no cases of invasive cancer were detected with a cervicography result of positive 1, a separate analysis to assess the performance of cervicography was conducted using a cervigram threshold of positive 0, positive 2, or positive 3 to indicate a positive result (hence, excluding the subcategories positive 1A and positive 1B). However, the sensitivity of cervicography for high grade squamous intraepithelial lesions dropped considerably, such that cervicography yielded an overall sensitivity of 22.1% and a specificity of 98.0% using this threshold.

Sensitivity and specificity of cervicography for high grade squamous intraepithelial lesions and cancer where a positive cervigram is defined by a cervigram result of atypical or more severe

If a more liberal cervicography threshold were established as atypical or positive, an overall sensitivity of 62.5% and specificity of 85.0% would result, i.e., 15.0% of women without high grade squamous intraepithelial lesions or cancer would be referred to colposcopy. Using this threshold, the lower sensitivity of cervicography among women ages 50 and older and among postmenopausal women was still noted. Sensitivities were 70.0% in women younger than age 50 compared to 30.8% in women ages 50 and older, yielding a statistically significant association between age group (<50 vs 50+) and cervigram sensitivity ($\chi^2 = 13.7$, $p < .001$).

For readers interested in the detection of low grade squamous intraepithelial lesions (as well as high grade squamous intraepithelial lesions and cancer) by cervicography, Table 5 presents the sensitivity and percent referral data where disease is

defined as low grade squamous intraepithelial lesions, high grade squamous intraepithelial lesions, or cancer. The results closely parallel the performance for detection of high grade squamous intraepithelial lesions and cancer previously shown in Table 4.

Conventional cytologic screening compared with the referent diagnosis

As a point of reference to Table 4, sensitivities and specificities were calculated for conventional cytology compared with the referent diagnosis. Using a threshold classification of atypical squamous cells of undetermined significance or more severe to define a positive cytology result and a referent diagnosis of high grade squamous intraepithelial lesions or more severe to define "disease", 586 of the 8449 women (6.9%) with available cytology results (5 were missing and 6 were not done) were referred to colposcopy by a conventional cytology result of atypical squamous cells of undetermined significance or more severe, yielding a sensitivity of 77.2% and a specificity of 94.2%. When women were categorized as younger than age 50 and ages 50 and older, the sensitivities of cytology were 75.5% and 84.6% respectively. The association between age group (<50 versus 50+) and cytology sensitivity for high grade squamous intraepithelial lesions and cancer was not statistically significant ($\chi^2 = 0.995$, $p = 0.3$). When a more stringent threshold of low grade squamous intraepithelial lesions or more severe was used to define a positive cytology result (and a referent diagnosis of high grade squamous intraepithelial lesions or more severe still used to define "disease"), conventional cytology yielded a basically unchanged sensitivity of 73.5% and a

specificity of 96.1% (atypical squamous cells of undetermined significance was not diagnosed frequently in Costa Rica).

Where the disease definition was loosened to a referent diagnosis of low grade squamous intraepithelial lesions, high grade squamous intraepithelial lesions, or cancer, and a positive cytology was defined as atypical squamous cells of undetermined significance or more severe, conventional cytology yielded a sensitivity of 67.7% with 6.9% of women screened being referred to colposcopy. When a positive cytology was determined by a result of low grade squamous intraepithelial lesions or more severe, conventional cytology yielded a sensitivity of 64.0% with 5.0% of women screened being referred to colposcopy.

Cervicography compared with conventional cytology, as adjudicated by the referent diagnosis

Cervicography and conventional cytology results were compared more directly by assessing the referent diagnosis where cervicography and cytology yielded discordant results for referral for colposcopy (Table 6). For this analysis, positive screening results were defined as the threshold for colposcopic referral actually used in the enrollment clinics. Thus, a positive cervigram was defined as any positive result and a positive cytology result was defined by a result of atypical squamous cells of undetermined significance or more severe. Of the 8284 participants who had technically adequate cervigram and conventional cytology results, 884 (10.7%) had discordant referral recommendations. Specifically, 489 women (5.9%) were referred by conventional

cytology but not cervicography, while 395 women (4.8%) were referred by cervicography and not by conventional cytology.

Considering first the 395 women with discrepant results referred by cervicography, 1 (0.3%) had a referent diagnosis of cancer, 15 (3.8%) had histologically-confirmed high grade squamous intraepithelial lesions, 30 (7.6%) had histologically-confirmed low grade squamous intraepithelial lesions, and 13 (3.3%) had non-histologically confirmed low grade squamous intraepithelial lesions. Of the 489 women referred by conventional cytology and not by cervicography, none had a referent diagnosis of cancer, 47 (9.6%) had histologically-confirmed high grade squamous intraepithelial lesions, 5 (1.0%) had non-histologically confirmed high grade squamous intraepithelial lesions, 16 (3.3%) had histologically-confirmed low grade squamous intraepithelial lesions, and 75 (15.3%) had non-histologically confirmed low grade squamous intraepithelial lesions.

Of the 131 women with high grade squamous intraepithelial lesions or cancer in Table 6 (three women with technically defective cervigrams and two with unsatisfactory Papanicolaou smears were excluded from this analysis), 52 (39.7%) were referred to colposcopy by conventional cytology and not by cervicography and 16 (12.2%) were referred by cervicography and not by cytology. The superior sensitivity of cytology was statistically significant (McNemar's $\chi^2 = 19.1$, $p < 0.001$), but only cervicography detected every case of invasive cancer.

Cervigram result compared with HPV DNA detection

Apart from the referent diagnosis, another reference standard is detection of HPV using DNA testing. Cancer-associated HPV types were associated with cervicography classification (Table 7a). High- and intermediate-risk, cancer-associated HPV DNA was found in 440 of 7000 women (6.3%) with a negative cervigram, 99 of 850 women (11.6%) with an atypical cervigram, 8 of 144 women (5.6%) with a positive 0 cervigram, 64 of 290 women (22.1%) with a positive 1 cervigram, 13 of 32 women (40.6%) with a positive 2 cervigram, and 8 of 18 women (44.4%) with a positive 3 cervigram. Among the women with available HPV DNA results, 49 of the 67 women with true positive cervigrams (73.1%) and 44 of the 417 women with false positive cervigrams (10.6%) tested positive for cancer-associated HPV types.

Conventional cytology compared with HPV detection

Cancer-associated HPV types were found in 217 of 4,804 women (4.5%) with a cytology result of normal, 208 of 2,996 women (6.9%) with a cytologic result of reactive changes, 21 of 163 women (12.9%) with a cytologic result of atypical squamous cells of undetermined significance, 84 of 265 women (31.7%) with a cytologic result of low grade squamous intraepithelial lesions, 91 of 133 women (68.4%) with a cytologic result of high grade squamous intraepithelial lesions, and 16 of 24 women (66.7%) with a cytologic result (often not confirmed) of microinvasive or invasive cancer (Table 7b). Among the women with available cytology and HPV DNA results, 86 of 105 women

(81.9%) with true positive cytology results and 126 if 480 (26.3%) with false positive cytology results tested positive for cancer-associated HPV DNA.

Cervicography compared with conventional cytology by HPV status

The enrollment HPV status for participants according to screening results is presented in Table 8. Of the 395 women referred to colposcopy by cervicography but not by cytology, 34 (8.6%) tested positive for a cancer-associated HPV type. Of the 488 women referred to colposcopy for a cytologic result of atypical squamous cells of undetermined significance or more severe, but not referred by cervicography, 149 (30.5%) tested positive for a cancer-associated HPV type. The higher HPV DNA prevalence in the additional cytologic "pick-ups" was statistically significant (McNemar's $\chi^2 = 72.3$, $p < 0.001$).

Discussion

In this high-risk, population-based study of 8,460 women, an abnormal cervigram led to the referral 5.7% of women for colposcopy, resulting in detection of all cancers, and 49.3% of all high grade squamous intraepithelial lesions and invasive cancers combined (sensitivity). The specificity of cervicography was 95.0%, with a positive predictive value of 13.8%. The sensitivity of cervicography was markedly reduced in postmenopausal women.

The importance of different levels of positive cervigrams varied. We observed that positive 2 and especially positive 3 classifications by cervicography had high positive predictive value, in that a majority of women given these results were found to have

underlying squamous intraepithelial lesions. However, these classifications were rather uncommon and insensitive. A screening program could not restrict colposcopic referral for these cases. A more liberal threshold is required. Although no cases of cancer were associated with a positive 1 result (compatible with a low-grade cervical lesion caused by HPV infection), the positive 1 diagnostic category was important to the detection of underlying high grade squamous intraepithelial lesions and can not be ignored. Exclusion of the positive 1 diagnostic category from the referral threshold reduced the sensitivity of cervicography to an unacceptably low level. The positive 0 category also must be taken seriously, although the overwhelming majority of women given this result were normal such that the HPV DNA positivity is even lower than among women with normal cervigrams. The classification of positive 0 identified 3 of the 11 women with referent diagnoses of cancer.

Using an even more liberal diagnostic threshold of atypical or positive to define a positive cervigram would permit cervicography to achieve greater sensitivity, but at the expense of a reduced specificity. This potential for cervicography to be non-specific, referring an unacceptably high percentage of normal women, was seen in earlier screening studies.[5,20-22] Using positive cervigrams as the referral threshold has made the screening system much more viable from a referral point of view as indicated by its high specificity. If sensitivity for high grade squamous intraepithelial lesions could be increased without greatly increasing referrals, the method could become more promising.

Regardless of referral threshold, cervicography may not be beneficial to postmenopausal women. A marked reduction in the sensitivity of cervicography was

seen in women no longer having menstrual periods, and probably can be explained by associated positional changes in the transformation zone. Most cervical neoplasia occurs at the transformation zone, which moves cephalad into the endocervical canal as a woman ages, especially if she is not on postmenopausal hormone replacement therapy. Since cervicography enables the evaluator to visualize a projected image of the cervix, the cervigram evaluator cannot detect lesions far inside the endocervical canal. Our data would strongly suggest restriction of cervicography to premenopausal women or women younger than age 50.

As a point of comparison, optimized conventional cytologic screening resulted in 6.9% of women being referred for colposcopy, when a cytology classification of atypical squamous cells of undetermined significance or more severe was established for referral. Cytologic screening resulted in 77.2% of women with high grade squamous intraepithelial lesions or cancer being correctly diagnosed as diseased (sensitivity) and 94.2% of women without disease being correctly identified as nondiseased (specificity). Conventional cytologic screening resulted in higher sensitivity than cervicography (with the exception of invasive cancer), with only a very small difference in specificity between the two screening methods. However, conventional cytology in this study had unusually high sensitivity for high grade squamous intraepithelial lesions and cancer compared with much less successful performance in Guanacaste Province in the past.[31] Cervical cancer rates have remained extremely high in Guanacaste for many years despite existing screening services. We believe that the maintenance of a highly accurate conventional cytology screening program is labor-intensive and technically difficult. But

improvement of existent conventional cytology should be considered along with other options for long-term cervical cancer prevention programs.

When cervicography and conventional cytologic diagnoses differed regarding the need for colposcopy, an analysis of the discordant results provided information about the performance of one screening method compared to the other. More than three times as many cases of high grade squamous intraepithelial lesions and cancer were detected with a positive cytologic result and a negative cervigram than were detected by a positive cervigram and negative cytology ($p < .001$). It is important to note that cervicography identified one cancer that cytology missed. However, 6 of 11 cancers (54.5%) were evident on clinical examination and these women would have been referred on that basis alone. Moreover, significantly more women referred only by conventional cytology were DNA positive for cancer-associated HPV types than women referred only by cervicography. This difference suggests that the additional sensitivity provided by cytology represented detection of additional true positive squamous intraepithelial lesions, not false positives.

In a separate analysis, referral for colposcopy was considered for women with either a positive cervigram *or* a cytology result of atypical squamous cells of undetermined significance or more severe. This "cytology plus cervicography" analysis yielded a sensitivity of 89.7% for detection of high grade squamous intraepithelial lesions or cancer and a specificity of 89.6%. Such combined screening would therefore result in the detection of more cases of disease at the expense of incorrectly referring for colposcopy 10.4% of women without disease.

The generalizability of this study is limited by the fact that it was conducted in a population with an age-adjusted rate of cervical cancer incidence averaging around 33 per 100,000. These rates are higher than the average for Costa Rica and 4-5 times higher than in the U.S.[25] The performance of any screening test may differ in a population with a lower prevalence of disease. Cervicography as a screening method for cervical neoplasia is limited by its dependence on an expert, trained evaluator to interpret cervigrams. Cervicography evaluation is much faster than conventional cytologic interpretation, permitting a large volume of screening dependent on fewer expert personnel. Additional limitations of cervicography include the need to maintain a camera and adequate supplies of film, and potential, but mild, burning effect of acetic acid in women with cervicitis.

Referent diagnoses were determined by an algorithm consisting of all pathology results, and four screening test results. This algorithm generally allowed for an appropriate referent standard against which to evaluate cervicography, although cervicography results were not used to define the non-histologically confirmed high grade squamous intraepithelial lesions. Cancer-associated HPV types were detected more frequently in women with referent diagnoses of equivocal than in women with referent diagnoses of normal, indicating that some women with referent diagnoses of equivocal likely had disease. The NCI-sponsored study of the natural history of cervical neoplasia will allow us to prospectively follow women in this cohort for development or progression of disease and to associate later findings with initial screening results.

A major strength of this study is the large, population-based sample in which it was conducted. A total of 9174 women completed the pelvic examination and screening

tests, and 8460 nonhysterectomized women had available cervigram results. This large sample size permitted the unbiased identification of a sufficiently large number of women with high grade cervical intraepithelial lesions or invasive cancer to be able to assess the performance of screening methods in detecting this category of diagnosis separately from low grade squamous intraepithelial lesions.

In summary, cervicography performed marginally better than conventional cytology for detection of invasive cervical cancer, although numbers were small. Cytology performed better than cervicography for the detection of high grade squamous intraepithelial lesions. The performance of cervicography was significantly reduced in post menopausal women and in women older than age 50. Cervicography might best be performed in women younger than age 50.

References

1. Coleman MP, Esteve J, Damiecki P., Arslan A, Renard H. Trends in cancer incidence and mortality. Lyon: IARC, 1993.
2. Beral V, Hermon C, Munoz N, Devesa SS. Cervical Cancer. Cancer Surveys: Trends in Cancer Incidence and Mortality 19/20. 1994; 265-285.
3. Kurman RJ, Henson DE, Herbst AL, Noller KL, Schiffman MH. Interim guidelines for management of abnormal cervical cytology. JAMA 1994; 271(23):1866-1869.
4. Sherman ME, Kurman RJ. The role of exfoliative cytology and histopathology in screening and triage. Obstet Gynecol Clin North Am 1996; 23(3):641-655.
5. Campion MJ. Epidemiology of genital human papillomavirus infection. Curr Probl Obstet Gynecol Fertil 1989; 109-120.
6. Koss LG. The Papanicolaou test for cervical cancer detection: A triumph and a tragedy. JAMA 1989; 261:737-743.
7. Campion MJ, Reid R. Screening for gynecologic cancer. Obstet Gynecol Clin North Am 1990; 17:695-727.
8. Cox JT, Schiffman MH. An evaluation of human papillomavirus testing as part of referral to colposcopy clinics. Obstet Gynecol 1992; 80(3, Part 1):389-395.

9. Sherman ME, Schiffman MH, Lorincz AT, Herrero R, Hutchinson ML, Bratti C, et al. Cervical specimens collected in liquid buffer are suitable for both cytologic screening and ancillary human papillomavirus testing. Cancer Cytopathology 1997; 81:89-97.

10. Sherman ME, Schiffman MH, Herrero R, Alfaro M, Kelly D, Mango L, et al. Performance of a semi-automated Papanicolaou smear screening system: Results of a population-based study conducted in Guanacaste, Costa Rica. Cancer 1998; 84(5):273-280.

11. Greenberg M, Sedlacek TV, Campion MJ. Cervical neoplasia: Are adjunctive tests to cervical cytology worthwhile? Clin Obstet Gynecol 1995; 38:600-609.

12. Greenberg MD, Campion MJ, Rutledge LH. Cervicography as an adjunct to cytologic screening. Obstet Gynecol Clin North Am 1993; 20:13-29.

13. Reid R, Greenberg M, Lorincz A, Jenson AB, Laverty CR, Husain M, et al. Should cervical cytologic testing be augmented by cervicography or human papillomavirus deoxyribonucleic acid detection? Am J Obstet Gynecol 1991; 164:1461-1471.

14. Tawa K, Forsythe A, Cove JK, Saltz A, Peters HW, Watring WG. A comparison of the Papanicolaou smear and the cervigram: sensitivity, specificity, and cost analysis. Obstet Gynecol 1988; 71:229-235.

15. Coibion M, Autier P, Vandam P, Delobelle A, Huet F, Hertens D, et al. Is there a role for cervicography in the detection of premalignant lesions of the cervix uteri? Br J Cancer 1994; 70:125-128.

16. Ferris DG, Payne P, Frisch LE, Milner FH, diPaola FM, Petry LJ. Cervicography: Adjunctive cervical cancer screening by primary care clinicians. J Fam Pract 1993; 37:158-164.

17. Baldauf J, Dreyfus M, Lehmann M, Ritter J, Philippe E. Cervical cancer screening with cervicography and cytology. Eur J Obstet Gynecol Reprod Biol 1995; 58:33-39.

18. Schneider A, Zahm DM, Kirchmayr R, Schneider VL. Screening for cervical intraepithelial neoplasia grade 2/3: Validity of cytologic study, cervicography, and human papillomavirus detection. Am J Obstet Gynecol 1996; 174:1534-1541.

19. Cecchini S, Bonardi R, Mazzotta A, Grazzini G, Iossa A, Clatto S. Testing cervicography and cervicoscopy as screening tests for cervical cancer. Tumori 1993; 79:22-25.

20. Szarewski A, Cuzick J, Edwards R, Butler B, Singer A. The use of cervicography in a primary screening service. Br J Obstet Gynaecol 1991; 98:313-317.

21. Soutter WP, Chaves J, Gleeson R, Lim K, Segall S, Skehan M. Cervicography in a colposcopy clinic. J Obstet Gynaecol 1991; 11:218-220.

22. August N. Cervicography for evaluating the "atypical" Papanicolaou smear. J Reprod Med 1991; 36:89-94.

23. Kesic VI, Soutter WP, Sulovic V, Juznic N, Aleksic M, Ljubic A. A comparison of cytology and cervicography in cervical screening. Int J Gynecol Cancer 1993; 3:395-398.

24. Urdaneta H. Detección de patologia cervical no diagnosticada por citología vaginal mediante cervigrafia. Rev Colomb Obstet Ginecol 1993; 44:220-222.

25. Herrero R, Schiffman MH, Bratti C, Hildesheim A, Balmaceda I, Sherman ME, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: The Guanacaste Project. Pan Am J Public Health 1997;1(5): 362-375.

26. Schiffman MH, Kiviat NB, Burk RD, Shah KV, Daniel RW, Lewis R, et al. Accuracy and interlaboratory reliability of human papillomavirus DNA testing by Hybrid Capture. J Clin Microbiol 1995; 33:545-550.

27. Lorincz AT, Reid R, Jenson AB, Greenberg M, Lancaster W, Kurman RJ. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. Obstet Gynecol 1992; 79:328-337.

28. Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J Natl Cancer Inst 1995; 87:796-802.

29. Kurman RJ, Soloman D. The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses: Definitions, Criteria, and Explanatory Notes for Terminology and Specimen Adequacy. New York: Springer-Verlag, 1994.

30. Schiffman MH, Liaw KL, Herrero R, Sherman ME, Hildesheim A. Epidemiologic support for a simplified view of cervical carcinogenesis. Eurogin Bulletin 1998; No. 1.

31. Herrero R, Hartge P, Brinton L, Reeves WC, Brenes MM, Urcuyo R. Determinants of the geographic variation of invasive cervical cancer in Costa Rica. Bull Pan Am Health Organ 1993; 27:15-25.

Table 1

Cervigram Classification Scheme

Not referred to colposcopy:

Negative:	No lesion seen
Atypical 1 (A1):	A trivial lesion outside the transformation zone is visible, but colposcopy not recommended because of the benign appearance or site of the lesion
Atypical 2 (A2):	A trivial lesion inside the transformation zone is visible, but colposcopy not recommended because of the benign appearance or site of the lesion
Technically defective:	Unable to be properly evaluated

Referred to colposcopy:

Positive 0 (P0):	Probably normal, but colposcopy preferable to rule out serious neoplasia
Positive 1A (P1A):	Compatible with trivial disease but colposcopy recommended because part of the lesion extends into the canal
Positive 1B (P1B):	Compatible with a low-grade squamous intraepithelial lesion/ flat condyloma/ and exophytic condyloma
Positive 2 (P2):	Compatible with a high grade squamous intraepithelial lesion
Positive 3 (P3):	Compatible with cancer

Table 2
Referent Diagnosis

Cancer:	Histologically-confirmed invasive cancer
HSIL:	Histologically-confirmed high grade squamous intraepithelial lesion
HSIL2:	Women with a conventional Papanicolaou smear and/or PapNet result of high grade squamous intraepithelial lesion, plus a ThinPrep result of high grade squamous intraepithelial lesion, but no histologic confirmation
LSIL:	Histologically-confirmed high grade squamous intraepithelial lesion
LSIL2:	Women with no histologic confirmation of a squamous intraepithelial lesion (SIL) and at least two of the following criteria met: 1) a conventional Papanicolaou smear or PapNet result of high grade squamous intraepithelial lesion; 2) a ThinPrep diagnosis of high grade squamous intraepithelial lesion; or 3) a cervigram result of positive 1, positive 2, or positive 3
Equivocal – NL/HSIL:	Women with a differential diagnosis of high grade squamous intraepithelial lesion versus negative on final review (severe atrophy contributed to this diagnostic category)
Equivocal – NL/LSIL:	Women whose overall results were equivocal, even following review by the chief study pathologist
Equivocal / ThinPrep:	Women with a ThinPrep cytologic diagnosis of high grade squamous intraepithelial lesion and all other screening tests normal
Equivocal / Pap:	Women with either a conventional Papanicolaou smear or PapNet diagnosis of high grade squamous intraepithelial lesion and all other screening tests normal
Equivocal / Cervigram:	Women with a cervigram result of positive 0, positive 1, positive 2, or positive 3 and no cytologic or histologic confirmation
Normal2:	Women referred to colposcopy with a cytologic diagnosis of atypical squamous cells of undetermined significance who were normal after review
Normal 1B:	Grossly visible abnormality, screening tests normal
Normal:	Women with all negative screening results

Table 3
Distribution of the Referent Diagnosis by Cervigram Result in 8460 Women in Guanacaste Province, Costa Rica
(Number and column percent)

	Referent Diagnosis (a)							
Cervigram Result (b)	Normal	Equivocal	LSIL2	LSIL	HSIL2	HSIL	Cancer	Total
Technically Defective	103	6	0	1	2	1	0	113
	1.4	0.8	0.0	1.4	25.0	0.9	0.0	1.3
Negative	6560	316	71	15	6	42	0	7010
	88.5	43.7	61.7	20.3	75.0	35.9	0.0	82.9
Atypical 1	147	5	2	1	0	2	0	157
	2.0	0.7	1.7	1.4	0.0	1.7	0.0	1.9
Atypical 2	595	53	18	14	0	16	0	696
	8.0	7.3	15.7	18.9	0.0	13.7	0.0	8.2
Positive 0	1	134	2	2	0	2	3	144
	< 0.1	18.5	1.7	2.7	0.0	1.7	27.3	1.7
Positive 1A	0	63	6	5	0	5	0	79
	0.0	8.7	5.2	6.8	0.0	4.3	0.0	0.9
Positive 1B	2	130	16	31	0	32	0	211
	< 0.1	18.0	13.9	41.9	0.0	27.4	0.0	2.5
Positive 2	1	11	0	5	0	12	3	32
	< 0.1	1.5	0.0	6.8	0.0	10.3	27.3	0.4
Positive 3	2	6	0	0	0	5	5	18
	< 0.1	0.8	0.0	0.0	0.0	4.3	45.5	0.2
Total	7411	724	115	74	8	117	11	8460

a) Definitions of the referent diagnosis are provided in Table 2.
b) The cervigram classification scheme is provided in Table 1.

Table 4
Sensitivity and Specificity of Cervicography for the Detection of Cancer, HSIL, or HSIL2 (a, b)

			Threshold level					
			Referred by Cervicography if Atypical or more severe			Referred by Cervicography if P0 or more severe		
	No. HSIL or cancer	Total	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)
Overall	136	8460	62.5	85.0	6.4	49.3	95.0	13.8
Age group (c)								
18-24	13	1043	92.3	73.7	4.2	53.9	91.8	7.7
25-29	23	1274	52.2	78.8	4.3	39.1	93.1	9.5
30-34	33	1336	81.8	82.4	10.5	69.7	95.5	28.1
35-39	25	1170	52.0	83.6	6.5	36.0	95.6	15.3
40-49	16	1604	81.3	88.5	6.6	75.0	95.7	15.0
50-64	10	1293	30.0	94.7	4.2	30.0	96.8	6.8
65+	16	740	31.3	93.5	9.6	25.0	96.0	12.1
Still having menstrual periods								
Yes	106	6264	69.8	81.7	6.2	54.7	94.5	14.6
No	30	2196	36.7	94.4	8.3	30.0	96.5	10.5

(a) The referent diagnosis [Table 2] is used as the gold standard diagnosis.
(b) Women with technically defective cervigrams were reported with negative cervigram results of negative [n = 113] and are included in this analysis.
(c) Age group was established based on date of enrollment interview.

Table 5
Sensitivity and Percent Referred for Colposcopy by Cervicography,
for the Detection of Cancer, HSIL, HSIL2, LSIL, or LSIL2 (a, b, c)

			Threshold level					
			Referred by Cervicography if Atypical or more severe			Referred by Cervicography if P0 or more severe		
	No. LSIL, HSIL or cancer	Total	Sensitivity (%)	Percent referred	Predictive Value Positive (%)	Sensitivity (%)	Percent referred	Predictive Value Positive (%)
Overall	325	8460	57.5	15.8	14.0	41.2	5.7	27.7
Age group (d)								
18-24	67	1043	70.2	27.1	16.6	43.3	8.7	31.9
25-29	64	1274	54.7	21.7	12.6	43.8	7.5	29.5
30-34	63	1336	63.5	19.2	15.6	47.6	6.1	36.6
35-39	55	1170	49.1	17.2	13.4	32.7	5.0	30.5
40-49	42	1604	64.3	12.2	13.8	47.6	5.0	25.0
50-64	15	1293	40.0	5.5	8.5	33.3	3.4	11.4
65+	19	740	26.3	7.0	9.6	21.1	4.5	12.1
Still having menstrual periods								
Yes	287	6264	60.6	19.2	14.5	42.9	6.4	30.9
No	38	2196	34.2	6.1	9.8	29.0	3.9	12.8

(a) The referent diagnosis [Table 2] is used as the gold standard diagnosis.
(b) Percent referred is the proportion of subjects in each category referred to colposcopy based on their cervigram result.
(c) Women with technically defective cervigrams were reported with negative cervigram results of negative [n = 113] and are included in this analysis.
(d) Age group established based on date of enrollment interview.

Table 6
Category of Agreement on Colposcopic Referral by the Referent Diagnosis

	Category of agreement on colposcopic referral (a, b)								
	Cervicography not referred Cytology not referred		Cervicography not referred Cytology referred		Cervicography referred Cytology not referred		Cervicography referred Cytology referred		
Referent Diagnosis (c)	number	column %	number	column %	number	column %	number	column %	Total
Normal	6571	89.9	0	0.0	0	0.0	0	0.0	6571
Normal 2 (differential diagnosis of ASCUS vs.Normal)	540	7.4	131	26.8	5	1.3	1	1.2	677
Equivocal - Cervigram	0	0.0	0	0.0	321	81.3	9	10.3	330
Equivocal - Pap	8	0.1	171	35.0	0	0.0	1	1.2	180
Equivocal - ThinPrep	132	1.8	9	1.8	1	0.3	0	0.0	142
Equivocal - NL/ LSIL (differential diagnosis of LSIL vs. ASCUS)	15	0.2	32	6.5	7	1.8	2	2.3	56
Equivocal - NL/ HSIL (differential diagnosis of HSIL vs. negative)	4	< 0.1	3	0.6	2	0.5	0	0.0	9
LSIL 2 (not histologically-confirmed)	16	0.2	75	15.3	13	3.3	11	12.6	115
LSIL (histologically-confirmed)	14	0.2	16	3.3	30	7.6	13	14.9	73
HSIL 2 (not histologically-confirmed)	1	< 0.1	5	1.0	0	0.0	0	0.0	6
HSIL (histologically-confirmed)	12	0.2	47	9.6	15	3.8	40	46.0	114
Cancer	0	0.0	0	0.0	1	0.3	10	11.5	11
Total	7313		489		395		87		8284

(a) Colposcopic referral based on cervigram and conventional cytology results using the following criteria:
Women were referred for colposcopy by cervicography if cervigrams were positive [positive 0, positive 1, positive 2, positive 3].
Women were referred for colposcopy by conventional cytology if smears were classified as atypical squamous cells of undetermined significance or more severe.
Data for participants with technically defective cervigrams (n = 113), unsatisfactory conventional cytologies (n = 57), or both (n = 4) were excluded from this analysis.
Cytology results were missing for 5 participants (1 of whom also had a technically defective cervigram) and conventional cytology was not done for 6 participants.
b) The cervigram classification scheme is provided in Table 1.
c) Definitions of the referent diagnosis are provided in Table 2.

Table 7a

Cervigram results by prevalence of
cancer-associated human papillomavirus types

Cervigram Result (b)	Cancer-associated HPV types (c) Frequency	row %	Total
Technically defective	10	8.9	113
Negative	440	6.3	7000
Atypical 1	10	6.4	156
Atypical 2	89	12.8	694
Positive 0	8	5.6	144
Positive 1A	14	17.7	79
Positive 1B	50	23.7	211
Positive 2	13	40.6	32
Positive 3	8	44.4	18
Total	642		8447

(a) Hybrid Capture tube HPV test was not performed for 6 subjects;
enrollment HPV results were missing for 7 subjects.
(b) The cervigram classification system is provided in Table 1.
(c) Cancer-associated HPV types include 16, 18, 31, 33, 35, 39, 45, 51,52, 56 & 58.

Table 7b

Conventional cytology results by prevalence of
cancer-associated human papillomavirus types

Conventional cytology result (b,c)	Cancer-associated HPV types (d) Frequency	row %	Total
Unsatisfactory	5	8.8	57
Normal	217	4.5	4804
Reactive changes	208	6.9	2996
ASCUS	21	12.9	163
LSIL	84	31.7	265
HSIL	91	68.4	133
Cancer	16	66.7	24
Total	642		8442

(a) Enrollment HPV results were missing for 7 participants;
conventional cytology results were missing for 5 participants.
(b) 6 women had neither the Hybrid Capture tube HPV test
nor conventional cytology performed.
(c) Conventional Papanicolaou smear results are defined according to the Bethesda System:
Unsatisfactory: unsatisfactory for evaluation
Normal: Normal
Reactive: reactive cellular changes
ASCUS: atypical squamous cells of undetermined significance
LSIL: low-grade squamous intraepithelial lesion(s)
HSIL: high grade squamous intraepithelial lesion(s)
Cancer: invasive cancer
(d) Cancer-associated HPV types include 16, 18, 31, 33, 35, 39, 45, 51,52, 56 & 58.

Table 8
Category of agreement on colposcopic referral by enrollment HPV status

	Category of agreement on colposcopic referral (a)								
	Cervicography not referred Cytology not referred		Cervicography not referred Cytology referred		Cervicography referred Cytology not referred		Cervicography referred Cytology referred		
Enrollment HPV status (b)	number	column %	number	column %	number	column %	number	column %	Total
Negative	6864	93.9	330	67.6	354	89.6	27	31.0	7575
Low-risk type (c)	58	0.8	9	1.8	7	1.8	1	1.2	75
Cancer-associated types (d)	385	5.3	149	30.5	34	8.6	59	67.8	627
Total	7307		488		395		87		8277

(a) Colposcopic referral based on cervigram and conventional cytology results using the following criteria:
Women were referred to colposcopy by cervicography if cervigrams were positive [positive 0, positive 1, positive 2, positive 3] vs. negative or atypical
Women were referred to colposcopy by conventional cytology if smears were classified as ASCUS or more severe vs. normal or reactive changes.
Data for participants with technically defective cervigrams (n = 113), unsatisfactory conventional cytologies (n = 57), or both (n = 4) were excluded from this analysis
Cytology results were missing for 5 participants (1 of whom also had a technically defective cervigram) and conventional cytology was not done for 6 participants

(b) Enrollment HPV results were missing for 7 subjects, HPV test was not performed for 6 participants, all of whom also had a technically defective cervigram

(c) Low-risk HPV types include 6, 11, 42, 43, and 44

(d) Cancer-associated HPV types include 16, 18, 31,33,35,39, 45, 51, 52, 56, and 58

Second Manuscript

What is the Optimal Performance of Cervicography?

An Evaluation of Cervicography Screening for Cervical Cancer in a High-Risk Population using an Arbitrated Screening Test Result

Diana L. Schneider

This project was funded by the National Cancer Institute (NCI), contract nos. N01-CP-21081 and N01-CP-31061.

Abstract

OBJECTIVE: An arbitrated cervigram classification was compared with a referent diagnosis to assess the optimal performance of cervicography.

STUDY DESIGN: From an initial sample of 8460 women, cervigrams corresponding to 3644 women underwent a second evaluation and subsequent arbitration if needed to achieve an optimized reading. The original sample was reconstituted based on the sampling frame for selection into the review phase. Interobserver agreement of the revised versus the original cervigram result was assessed. The "optimized" cervicography results were compared with a referent diagnosis determined by histology and three cytologic tests. Sensitivity, specificity, predictive values, and likelihood ratios were estimated, and results were stratified by characteristics of the woman and visual characteristics of the cervigram image. Digital colposcopic images were interpreted to evaluate the perceived appropriateness of the decision to biopsy and biopsy placement, and the impact of these on sensitivity and specificity.

RESULTS: Moderate agreement on cervigram classification was observed between the original and secondary reviewers, yielding *kappa* statistics of 0.47 on specific classification and 0.54 on classification of positive versus not (i.e., negative, atypical, or technically defective). For the detection of high grade squamous intraepithelial lesions or cancer, optimized cervicography yielded an overall sensitivity of 55.2% and a specificity of 94.3%. Higher sensitivity was associated with younger age, premenopausal status, the presence of metaplasia, the absence of cervicovaginal atrophy, and improved quality of

the acetic acid effect. A corresponding lowered specificity was associated with these same factors and other visual characteristics of the image.

CONCLUSION: Evaluator agreement with cervicography is moderate. The optimized cervigram classification improved performance only slightly over a single interpretation. Cervicography performs better than cervicography for the detection of high grade squamous intraepithelial lesions, but both tests perform comparably for the detection of invasive cancer.

Introduction

The present research project was conducted to provide a rigorous and independent evaluation of Cervicography™ [National Testing Laboratories Worldwide (NTL), Fenton, MO] as a primary screening method for early identification and prevention of cervical cancer. This evaluation of cervicography was conducted as part of a large, population-based cohort study of the natural history of cervical neoplasia in the province of Guanacaste, Costa Rica, sponsored by the National Cancer Institute. The Guanacaste site was selected because of its consistently high age-adjusted rates of cervical cancer despite existing Papanicolaou smear screening services.[1]

In previous work,[2,3] we reported that findings from the enrollment phase of this study indicated that cervicography was less sensitive, and only marginally more specific, than conventional cytologic testing for the detection of high grade squamous intraepithelial lesions or cancer. However, cervicography was easy to perform and judged potentially important, especially if sensitivity for detecting high grade lesions could be increased without substantially reducing specificity. In the current study, we submitted a subsample (43%) of cervigrams taken during the enrollment study to additional evaluation by an independent evaluator and subsequent arbitration by a third cervigram evaluator where the enrollment and subsequent review classifications were discordant. This process allowed us to achieve a cervicography classification that approaches the optimal result achievable by cervicography. The revised, or optimized, cervicography

result is reported here as compared with the referent diagnosis of disease status, which is based on reviews of all available cytologic and histopathologic material.

Materials and Methods

This study of cervicography was conducted as part of a population-based study of the natural history of cervical neoplasia in Guanacaste Province, Costa Rica. The study design is described in greater detail elsewhere.[1] The protocol for this study was approved by the Costa Rican, National Cancer Institute, and Uniformed Services University of the Health Sciences Institutional Review Boards.

Subject selection and participation rates

The cohort was enrolled between June 1993 and December 1994. The population was recruited through door-to-door visits in randomly selected census segments. A total of 11,742 women was selected who met the initial eligibility criterion of being 18 years of age or older by July 1, 1993. Of those, 10,738 (91.5%) were eligible for an interview. Women who were not Guanacaste residents, or were mentally incompetent, physically incapacitated, unable to understand Spanish, or deceased were excluded.

Study clinics were established at regional hospitals or at dozens of local health outposts. Pregnant women were deferred until three months after giving birth. A total of 10,049 of the 10,738 eligible women (93.6%) was interviewed. Pelvic examinations were not performed on the 583 virgins who completed the enrollment interview. The remaining 9466 (94.2% of those interviewed) were eligible for an enrollment pelvic examination. Of those, 291 women either refused or were physically unable to undergo

the pelvic examination. Pelvic examinations were performed on 9175 women and a referent diagnosis was established for all of these. Cervigrams were obtained for 9062 women, corresponding to 98.8% of women completing the pelvic examination. The 602 hysterectomized women (6.6%) were subsequently excluded from the analyses, leaving 8460 participants available for the enrollment phase of this study.

Clinical Specimens

After obtaining informed consent, patients participated in a detailed interview to assess risk factors for cervical neoplasia and then underwent a routine pelvic examination. Cervical cytologic and human papillomavirus (HPV) deoxyribonucleic acid (DNA) specimens were collected during the pelvic examinations using a Cervex Brush (Unimar, Wilton, CT). Two types of cytologic preparations were made for each participant, including a Papanicolaou smear and a ThinPrep (Cytyc, Boxborough, MA). Papanicolaou smears were prepared according to the Papanicolaou method in Costa Rica. After the smear was made, the Cervex brush was rinsed in 20 mL of PreservCyt (Cytyc). Vials containing the PreservCyt solution were sent to the United States where ThinPreps were made.

Additional cervical cells were collected using a Dacron swab which was placed in Specimen Transport Medium (Digene Corporation, Silver Spring, MD), frozen, and shipped to the United States for HPV DNA testing using the first generation Hybrid Capture tube test (Digene). Samples were tested for HPV positivity at the 10 pg/mL level[4] using probes for cancer-associated types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, and

58) and low-risk types (6, 11, 42, 43, and 44).[5,6] Subsamples of specimens were retested for HPV positivity using polymerase chain reaction and with the more sensitive Hybrid Capture II that includes two additional types (59 and 68).[7]

The cervix was then rinsed with 5% acetic acid and two photographic images of the cervix (Cervigrams™) were taken using a Cerviscope™ [National Testing Laboratories Worldwide (NTL), Fenton, MO]. The undeveloped film was sent to the United States for developing, processing and evaluation.

Assignment of screening test results (enrollment phase)

Cervigrams taken during the enrollment pelvic examination were initially interpreted by certified evaluators (Mitchell Greenberg, Thomas Sedlacek, Michael Campion) and classified according to the diagnostic criteria approved by NTL as noted in Table 1.

Cytologic diagnosis was made using three methods: Papanicolaou smear; PapNet [Neuromedical Systems, Inc., Suffern, NY (now TriPath, Elon, NC)], which uses the same slide as the Papanicolaou smear; and ThinPrep. Papanicolaou smears were interpreted in Costa Rica (Mario Alfaro). ThinPreps were prepared and interpreted in the United States (Martha Hutchinson). After interpretation in Costa Rica, all available Papanicolaou smear slides were sent to Neuromedical Systems (now TriPath) for repeat screening using the PapNet system, a neural network-based, semi-automated device.[8] Papanicolaou smear, ThinPrep, and PapNet results were classified according to The Bethesda System as negative (within normal limits or reactive cellular changes), atypical

squamous cells of undetermined significance, low grade squamous intraepithelial lesion, high grade squamous intraepithelial lesion, or carcinoma.[9] Glandular lesions were very rare and were not classified separately.

Colposcopic referral and examination

Participants were referred for colposcopy if 1) physical examination was suspicious for cancer, 2) there was an abnormal cytologic result by any of the three methods (atypical squamous cells of undetermined significance or more severe), or 3) there was a positive cervigram. Colposcopy was performed by a single gynecologist (Jorge Morales). During the colposcopy examinations, digital images (Denvu Lld., Tucson, AZ) of the cervix were taken for each woman corresponding to 1) with low magnification before application of 5% acetic acid (which provides an acetowhitening effect that highlights lesions), 2) with low magnification after application of 5% acetic acid, 3) with high magnification after application of 5% acetic acid, 4) with high magnification after application of Lugol's solution (a dilute iodine for staining lesions), and 5) of the biopsy site, if applicable.

As a quality control measure, a random sample of 2% of all women was referred for colposcopy to validate the screening protocol. All 144 of the women with all negative screening tests had a referent diagnosis of normal, indicating virtual 100% sensitivity of the screening protocol.

Histologic specimens and patient management

Biopsies of lesions visualized on colposcopy were fixed in 10% buffered formalin, embedded in paraffin, stained with hematoxylin and eosin, and diagnosed in Costa Rica for clinical purposes. Histologic material, including punch biopsies, subsequent cone biopsies, excised tissue, and hysterectomies, was sent to the United States for review and assignment of the referent diagnosis (Mark Sherman). Participants with a histologically-confirmed high grade squamous intraepithelial lesion or cancer, or with highly a suspect diagnosis of high grade squamous intraepithelial lesion by at least two cytologic methods, were referred for treatment through the Costa Rican Social Security system.[1]

Referent diagnosis

The enrollment referent diagnoses were made based on histology, cytology, and cervicography results and classified as indicated in Table 2, with subsequent grouping as required for the data analyses.

Additional reviews of clinical materials

Subsequent cervigram reviews and assignment of revised results

Key cervigrams (Table 3) from the enrollment phase of the study were reviewed by two certified evaluators (different from the evaluators at enrollment) to assess possible error in the initial interpretation. Evaluators were masked from knowing the previous cervigram, cytology, histology, or HPV results or even the composition of the sample. However, they were aware of the general results from the enrollment phase of the study,

which had previously been published. Selection into the cervigram review phase of the study was made based on previous outcomes and/or risk factors. Cervigrams corresponding to 3644 women were selected for the review. The sample composition and corresponding sampling fractions for the cervigram review are shown in Table 3.

Cervigram classification categories were the same as those shown in Table 1. We compared the cervigram results between the initial evaluator at enrollment and the second evaluator (Mark Spitzer). Cervigrams corresponding to women whose classification assignment differed between the two evaluators (n=824, 22.6%) were sent, along with a 10% sample of cervigrams for women with concordant results (n=282), to a third evaluator (Louis Burke) for arbitration. The subsample of women with concordant results was included to mask the sample composition. A revised cervigram result was assigned based on agreement of two of the three evaluators [initial (enrollment, second, and third (arbitration)] on cervigram classification. Women for whom all three evaluators assigned a different category of classification are presented as a separate category (i.e., three-way disagreement) where applicable.

Additional information was collected about the visual image of the cervix in an attempt to explain discordant results and to stratify cervicography performance estimates. Visual characteristics that were recorded include 1) acetowhite epithelium; 2) discoloration; 3) erosion/ulceration; 4) irregular surface contour; 5) mosaicism; 6) punctation; 7) atypical vessels; 8) whether the lesion was seen in its entirety; 9) whether columnar epithelium was visible on the ectocervix; 10) whether metaplasia was visible on the ectocervix; 11) whether the cervigram showed a congenital

transformation zone (which appears acetowhitened after application of acetic acid); 12) whether the acetic acid effect was sufficient; 13) whether the transformation zone was partially obscured by blood, mucus, the position of the cervix, hair, vaginal wall, or speculum; and 14) presence and degree of inflammation.

Digital colposcopic image review

We also assessed the possibility of error in the colposcopy examination during the enrollment study through a review of the digital images taken during the examination (Louis Burke). All available images taken during the initial colposcopy examination were included in the review (n=1983 women, 96.4% of all examinations). Images were evaluated according to the reviewer's agreement or disagreement with the decision to take a biopsy and, if taken, on the reviewer's agreement or disagreement with biopsy placement.

Data analysis

The primary definition of disease used as a target for screening, based on the enrollment referent diagnoses, was defined as high grade squamous intraepithelial lesions (HSIL and HSIL2 in Table 2) or cancer (versus normal, equivocal, or low grade squamous intraepithelial lesions). A secondary definition of disease was defined as low grade squamous intraepithelial lesions (LSIL and LSIL2 in Table 2), high grade squamous intraepithelial lesions, or cancer (versus normal or equivocal). Some analyses considered low grade squamous intraepithelial lesions as a diagnostic endpoint separate from the high grade squamous intraepithelial lesions or cancer category. The primary threshold

used to define a positive cervigram includes all categories of positive cervigrams (positive 0, positive 1, positive 2, and positive 3, versus negative or atypical). Women who were selected for inclusion in the review phase for whom both cervigrams were not available (i.e., missing) were excluded from all analyses (n=6). These exclusions included two women with a final diagnosis of high grade squamous intraepithelial lesions at enrollment. One woman for whom the second evaluator was unable to provide a cervigram classification, and one woman for whom the second evaluator was unable to provide a cervigram classification were excluded from relevant analyses

Sensitivity, specificity, percent of women referred for colposcopy, and positive and negative predictive values were initially calculated using standard contingency table methods,[10] comparing the enrollment screening test results with the enrollment referent diagnosis as the gold standard. Percent referred provides an estimate of the overall burden on colposcopy services for a particular screening test. A detailed report of the initial findings from the enrollment phase of the study is available elsewhere.[2,3]

For women whose cervigrams were included in the review sample, we assessed interobserver agreement between the results by the reviewer at enrollment and by the subsequent reviewer for cervicography using *kappa* and weighted *kappa* statistics. Weighted kappas are intended to provide greater weight to pairs of results that have more extreme differences compared with those for which the assigned classifications are more alike. *Kappa* and weighted *kappa* statistics yielded similar results; therefore, only unweighted *kappas* are presented. *Kappa* statistics are interpreted using the scale described by Altman.[11] Additionally, we tested for differences in the dichotomous

cervigram classification by the initial and second evaluator for women with a referent diagnosis of high grade squamous intraepithelial lesions or cancer and, separately, for women with a referent diagnosis of normal, equivocal, or low grade squamous intraepithelial lesions using McNemar's chi square test for paired proportions.

We re-calculated sensitivity, specificity, percent referred, and positive and negative predictive values for cervicography using standard contingency table methods after incorporating revised cervicography results. For these analyses, the original sample of women (from the enrollment phase of the study) was reconstituted according to the sampling fractions by which women were selected into the cervicography reviews (Table 3). Reconstitution was achieved by multiplying the contingency table frequencies from the review sample by the inverse of the sampling fraction corresponding to each category of selection into the review phase. In other words, these analyses were stratified by the categories established for selection into the cervigram reviews. Each stratum was reconstituted, and then stratum-specific results were combined before analyses were performed. By recalculating sensitivity, specificity, percent referred, and predictive values using the revised test results, we assumed that the arbitrated reviews provide optimal values for estimating the performance of cervicography as a primary screening test for cervical cancer. Confidence intervals for these estimates were calculated using the standard formula for confidence interval determination for binomial proportions.[12]

Another statistic employed in the evaluation of a screening test and presented here is the likelihood ratio.[13,14] The likelihood ratio is the ratio of the probability of having a given test result among diseased persons to the probability of having the same test result

among nondiseased persons. We calculated the likelihood ratios for each level of cervigram classification, i.e, normal, atypical, positive 1, positive 2, positive 3, and for women for whom the 3 evaluators disagreed. In other words, we calculated the likelihood ratios for having a given cervigram result among women with a referent diagnosis of high grade squamous intraepithelial lesions or cancer versus women with a referent diagnosis of normal or equivocal. Additionally, we provide the likelihood ratios for having a given cervigram result among women with a referent diagnosis of low grade squamous intraepithelial lesions versus women with a referent diagnosis of normal or equivocal. As a point of comparison, we also show the likelihood ratios for conventional cervical cytology.

We stratified sensitivity and specificity by predictors of error, including age, menopausal status, visual characteristics of the cervigram, agreement on the decision to take a biopsy, and agreement on biopsy placement. Additionally, we assessed whether discordance between the initial and second cervigram results were associated with these same factors. Discordance was defined in two ways, including 1) when cervigrams were classified as technically defective, negative, atypical, positive 0, positive 1, positive 2, or positive 3; and 2) when cervigrams were classified as negative, atypical, or technically defective versus any positive. For these analyses, the visual characteristics were recorded during the second round (i.e., re-review) of cervigram evaluation. We tested for associations between these characteristics and sensitivity and specificity, and between the characteristics and having discordant results using Fisher's Exact test.[12] In a final analysis, utilizing only the subsample of women in the review sample whose cervigram

classifications were arbitrated by the third evaluator, we assessed whether agreement on the cervigram characteristics was associated with agreement on the cervigram classification using Fisher's Exact test.

Results

Observer agreement on cervigram classification

The cervigram classification by the initial evaluator and the second evaluator (i.e., for the review phase) were compared for the 3637 women included in the cervigram review with available results. When cervigram results were categorized on narrow categories (i.e. atypical, negative, positive 0, positive 1, positive 2, positive 3, technically defective), the analysis yielded a *kappa* statistic of 0.47, indicating moderate agreement between the two evaluators beyond that expected by chance alone. Agreement improved only slightly (*kappa* = 0.54) when cervigram results were grouped based on the broader referral categories (i.e., positive vs. not, including negative, atypical, or technically defective). In the subsample of 1102 women whose cervigrams were arbitrated by the third evaluator, agreement between the third evaluator (i.e., arbitrator) and each of the previous two evaluators was similar, yielding a *kappa* statistic of 0.2 for both comparisons, indicating poor agreement beyond that expected by chance alone. These kappa statistics were the same when cervigram results were grouped narrowly and broadly as defined above.

Cervicography screening compared with the referent diagnosis

The distribution of the referent diagnosis by the revised cervigram results for the reconstituted sample is shown in Table 4. An arbitrated cervigram result based on narrow categories was available for 8035 (95.1%) of the 8452 women with available (i.e., nonmissing) cervigrams. Each of the three evaluators assigned a different classification to cervigrams corresponding to the remaining 417 (4.9%) women. Of the 11 cases of invasive cancer, 10 (90.9%) were identified by the arbitrated cervicography process. It is noteworthy that the one case of invasive cancer not detected by the arbitrated cervigram result was initially correctly picked up by the evaluator at enrollment. However this result was changed following interpretation during the two subsequent reviews.

When the cervigram classification categories were defined more broadly (i.e., normal, atypical, positive 1, positive 2, and positive 3 each as separate categories), 127 (30.5%) of the 417 women for whom all three evaluators disagreed on classification were reassigned as positive (when all positives were grouped together), and 290 (69.5%) were reassigned as negative, atypical, or technically defective (when these categories were grouped together). Of the 127 women who were reassigned as positive, 3 (2.4%) had a referent diagnosis of cancer, 5 (3.9%) had high grade squamous intraepithelial lesions, 1 (0.8%) had low grade squamous intraepithelial lesions, 71 (55.9%) had equivocal diagnoses, and 47 (37%) had a referent diagnosis of normal.

Observer agreement by category of the referent diagnosis

We assessed the difference between the proportion of women with high grade squamous intraepithelial lesions or cancer correctly identified with any positive cervigram by the initial and second evaluators. Of the 134 women with a referent diagnosis of high grade squamous intraepithelial lesions or cancer, 57 (42.5%) women had positive cervigrams and 49 (36.6%) had negative cervigrams by these two evaluators, while 9 (6.7%) women had positive cervigram results by the initial evaluator and negative results by the second, and 19 (14.2%) women had positive cervigram results by the second evaluator and negative results by the first. Based on these results, the difference in the proportion of women with high grade lesions or cancer referred for colposcopy by the two evaluators was statistically of borderline significance (McNemar's chi square = 3.6, $p = 0.06$). Similarly, among the 3503 women in the review sample without serious neoplasia (i.e., having a referent diagnosis of normal, equivocal, or low grade squamous intraepithelial lesions), 252 (7.2%) women had positive cervigrams and 2874 (82.0%) women had negative cervigrams by both the initial and second evaluators, while 164 (4.7%) women had positive cervigram results by the initial evaluator and negative results by the second, and 213 (6.1%) women had positive cervigram results by the second evaluator and negative results by the first. These analyses indicate that there was a significant difference between the two evaluators in the proportion of women without serious neoplasia who were identified with a positive cervigram (McNemar's chi square = 6.4, $p = 0.01$).

Performance estimates of cervicography compared with the referent diagnosis

The two referral categories of cervigram classification [positive (i.e., referred for colposcopy) versus normal, atypical, or technically defective (i.e., not referred)] were used to determine percent referred, sensitivity, specificity, and predictive values of cervicography. An arbitrated result was available for all 8452 women in the reconstituted sample with available cervigrams. Among these women, 549 (6.5%) would be referred for colposcopy based on this arbitrated cervigram result. The arbitrated cervigram result yielded a sensitivity of 55.2% (95% CI 46.8, 63.6), a specificity of 94.3% (95% CI 93.8, 94.8), a positive predictive value of 13.5% (95% CI 10.6, 16.4), and a negative predictive value of 99.2% (95% CI 99.0, 99.4).

The likelihood ratio statistic allowed us to assess the predictive ability of each individual level of the optimized cervigram classification relative to the referent diagnosis (Tables 5a and 5b). As shown in Table 5a, the likelihood ratios for high grade squamous intraepithelial lesions or cancer compared to women without cervical neoplasia increase with the severity of the cervigram result. Additionally, women with low grade squamous intraepithelial lesions are more likely to have positive 1 or positive 2 cervigrams than are women with normal or equivocal referent diagnoses. When cervigram results are grouped into the broader, dichotomous categories, the likelihood ratio of a positive cervigram for predicting high grade squamous intraepithelial lesions or cancer drops considerably to 11.1, and the likelihood ratio of a positive cervigram for low grade squamous intraepithelial lesions becomes 7.6 (Table 5b). The likelihood ratio of 2.4 for

the category where the three evaluators disagreed illustrates the potential importance of these hard-to-diagnose cervigrams.

As a point of comparison, Table 5c shows the likelihood ratios for conventional cytology. This table shows that women with cytology results of cancer have a similar likelihood of having a referent diagnosis of high grade squamous intraepithelial lesions or cancer (compared to women without neoplasia) to that of women with a positive 3 cervigram. However, a cytologic result of high grade squamous intraepithelial lesion appears to be more predictive of true high grade squamous intraepithelial lesions or cancer than is a positive 2 cervigram result. Cytology results grouped as atypical squamous cells of undetermined significance or more severe yield higher likelihood ratios for high grade squamous intraepithelial lesions and cancer and for low grade squamous intraepithelial lesion than those attained by all positive cervigrams combined.

The effects of various characteristics on cervigram results

Table 6 shows the sensitivity and specificity for the revised cervigram result stratified by various characteristics of the women and their cervigrams. Stratification was performed in an attempt to explain the poor ability of cervicography to detect high grade lesions as observed in the overall results. Sensitivity is significantly higher among women younger than age 50 (62.0%) compared to women ages 50 and older (26.9%), and among premenopausal women (62.5%) compared to postmenopausal women (30.0%). Several visual characteristics on the cervigram are also associated with an increase in

sensitivity, including metaplasia, the absence of cervicovaginal atrophy, and an increasing quality of the acetic acid effect.

The sensitivity and specificity of cervicography by age group was also affected by the visibility of the acetic acid effect. Among the 1958 women ages 50 and older with a recorded observation on acetic acid effect, 582 (29.7%) had an acetic acid effect visible. Within this subgroup, the sensitivity increased to 50.0% (cervicography detected 7 of the 14 cases of high grade lesions or cancer, specificity = 94.5%), while none of the 12 women in the older age group with high grade lesions or cancer and without a visible acetic acid effect were detected by cervicography. Among the 6477 women younger than age 50 with a recorded observation on acetic acid effect, 5438 (84.0%) had an acetic acid effect visible. Within this subgroup, the sensitivity increased to 62.6% (cervicography detected 67 of the 107 cases of high grade lesions and cancer, specificity = 92.1%). The one woman in the younger age group with a high grade lesion or cancer who did not have an acetic acid effect visible was missed by cervicography. Among the 1390 women younger than age 50 who had an acetic acid effect visible, the sensitivity increased dramatically to 90.4% (specificity = 76.5%) in women with a good acetic acid effect. In contrast, the sensitivity for the 3194 younger women with a fair acetic acid effect was 37.5 (specificity = 97.0%), and for the 848 younger women with a poor acetic acid effect was 28.6%.

Presence of the above characteristics resulted in a corresponding loss of specificity. Statistical significance achieved from some apparently small differences in specificity may be explained by the higher power of these analyses, due to the larger

numbers of women without serious neoplasia (i.e., the denominator of the equation used to calculate specificity). Specificity is considerably reduced in women whose cervigrams show a congenital transformation zone, although few women had this characteristic (n = 40). This may be explained by the acetowhitened appearance of a congenital transformation zone after application of acetic acid

Digital colposcopic images were available for 1983 women. Two key variables were assessed, including 1) agreement on the decision to biopsy, and 2) agreement on biopsy placement within 5 mm (if a biopsy was taken). Of these 1983 women, 320 (16.2%) had a biopsy taken, and 1615 (81.6%) did not. The reviewer of the digital colposcopic images agreed with the decision to biopsy for 224 (70.0%) of the 320 women who had a biopsy; he agreed with the decision not to take a biopsy for 1233 (76.3%) of the 1615 women who did not have a biopsy taken. The image reviewer agreed with biopsy placement for 170 (75.9%) of the 224 women for whom a biopsy was taken and there was agreement on the decision to biopsy.

Additional analyses were designed to assess the effect of the appropriateness of the decision to biopsy and of the adequacy of the biopsy site on the sensitivity and specificity of cervicography. To achieve this, we compared observations of digital colposcopic images with findings from the colposcopy examination during the enrollment phase of the study. Estimates of sensitivity and specificity were stratified by the two key variables defined above. Results of these analyses appear on the bottom of Table 6. Briefly, we found that sensitivity was higher (70.2%) in women for whom there was no agreement on the decision to biopsy, compared to women for whom there was agreement

on the biopsy decision (53.4%). Of the 380 women for whom the digital colposcopic image review revealed lack of agreement on the decision to biopsy, 294 (77.4%) had not had a biopsy taken and the image reviewer indicated that a biopsy should have been taken, while 86 (22.6%) had a biopsy taken and the image reviewer indicated that a biopsy need not have been taken. However, agreement on biopsy placement was not a factor in explaining the low sensitivity of cervicography. Among women who had a biopsy taken and the image was adequate for assessment (n=218), higher sensitivity was achieved in women for the digital colposcopic image review indicated agreement on biopsy placement (72.4%) than women for whom there was no agreement on biopsy placement (60.0%). It merits comment that images corresponding to 55 of the 1983 (2.8%) women with available digital colposcopic images were deemed inadequate to determine whether or not a biopsy was taken. Images corresponding to an additional 98 women (4.9%) were deemed inadequate to determine whether or not a biopsy should have been taken, and images corresponding to an additional 2 women (0.1%) were deemed inadequate to assess the adequacy of biopsy placement.

For the women whose cervigrams were included in the review sample, we examined how these same characteristics affected agreement between the enrollment and subsequent evaluator on cervigram classification (Table 7). Many visual characteristics were significantly associated with agreement between the two evaluators when both narrow (i.e, atypical, negative, positive 0, positive 1, positive 2, positive 3, technically defective) and broad (i.e., positive vs. not) categories of cervigram classification are defined, as indicated in Table 7. Agreement with the decision to take a biopsy was

associated with concordance on cervigram classification, while perceived appropriateness of the biopsy placement was not (Table 7). The numbers used in the assessment of an association between perceived appropriateness of biopsy placement were small because only women having had a biopsy and for whom there was agreement on the decision to take a biopsy were included in the analysis.

It is important to note, however, that for the women whose cervigrams were reviewed twice in the review phase of the study (i.e., in the second and third, i.e., arbitration rounds), the two evaluators generally had poor to moderate agreement on the visual characteristics beyond that expected by chance alone. A comparison of visual characteristics observed by these two evaluators (i.e., second versus third) yielded *kappa* statistics less than 0.2 for eleven characteristics (indicating poor agreement), between 0.21 and 0.40 for ten characteristics (indicating fair agreement), and between 0.41 and 0.60 for five characteristics (indicating moderate agreement) (data not shown). We explored further to assess whether an association exists between agreement on the visual characteristics of the cervigram and agreement on cervigram classification. Agreement on several cervigram characteristics was positively associated with agreement on cervigram classification when narrow categories of cervigram classification were used (data not shown). However, many of these associations disappeared when agreement on cervigram classification was based on the dichotomous categories of positive versus not. The positive associations that remained included the presence of acetowhite epithelium, discoloration, erosion, irregular surface contour, and/or atypical vessels, and the entirety of the lesion visible on the cervigram. All of these characteristics are important to

assigning the cervigram classification itself, whereas many of the associations that dropped out when agreement was based on the dichotomous categories correspond to characteristics that are not specifically used to guide the diagnosis.

Discussion

In our initial evaluation of cervicography as a primary screening test for cervical cancer[2,3] we established that cervicography performed less than optimally. During the enrollment phase of our study, 5.7% of the 8460 women were referred for colposcopic examination because of a positive cervigram. Cervicography resulted in the detection of all 11 cases of invasive cervical cancer and 49.3% of high grade squamous intraepithelial lesions and cancer combined (sensitivity). The specificity of cervicography was 95.0%, with a positive predictive value of 13.8% and a negative predictive value of 99.1%. These results led to the design of the present study in which we assessed the optimal performance of cervicography screening. By submitting cervicography to a second round of interpretation followed by an arbitration process, we established a good estimate of the optimal cervigram classification.

The performance of cervicography screening was assessed in terms of interobserver agreement and also in terms of traditional validity estimates (i.e., sensitivity, specificity, predictive values, and likelihood ratios). Cervicography suffers from imperfect reproducibility. In comparing the cervigram classification between the evaluator at enrollment with the second evaluator in the review phase, results showed moderate agreement beyond that expected by chance alone. Only a small

improvement in observer agreement was found when cervigram results were grouped into two categories based on the referral threshold of any positive cervigram versus negative, atypical, or technically defective. Further, among women with high grade squamous intraepithelial lesions or cancer, these two evaluators differed in which women they would refer for colposcopic examination. Following arbitration by a third evaluator, 417 women (4.9%) had different cervigram results by each of the three evaluators. Of these 417 women with discordant results, 127 were reassigned as positive using broader categories of classification. Only 8 (6.3%) of these 127 women had serious neoplasia, while 71 (55.9%) had a referent diagnosis of equivocal, indicating that many of these hard-to-interpret cervigrams corresponded to women with hard-to-interpret pathology. Among the 134 women in this study with a referent diagnosis of high grade squamous intraepithelial lesions or cancer, 11.2% had discordant cervigram results by all three evaluators. However, cervicography is not the only cervical cancer screening test that suffers from poor reproducibility. Studies of interobserver agreement of the Papanicolaou smear have also shown moderate or even poor reproducibility.[15-26]

The optimal performance of cervicography (i.e., after the arbitration process) was only slightly improved over the initial result based on a single evaluation. The revised cervigram classification resulted in 6.5% of women being referred for colposcopic examination, and it yielded a sensitivity of 55.2%, a specificity of 94.3%, a positive predictive value of 13.5%, and a negative predictive value of 99.2% for the detection of high grade squamous intraepithelial lesions or cancer. As a point of comparison, during the enrollment phase of our study, conventional cytologic screening resulted in 6.9% of

women being referred for colposcopy, and it yielded a sensitivity of 77.2%, a specificity of 94.2%, a positive predictive value of 17.9% and a negative predictive value of 99.6%.[2] The processes involved in conventional cytologic screening in the enrollment phase of our study were optimized by careful technical assistance in preparation for this study. The assistance with cytologic smears prior to the start of the study may partially explain why the performance of conventional cytology in our study is better than it has been in previous studies.[27]

The more favorable performance of conventional cytology compared with cervicography is also demonstrated by the higher likelihood ratios for cytology for high grade squamous intraepithelial lesions or cancer compared to the likelihood ratios for cervicography. Although the likelihood ratios indicate increasing risk associated with increasing severity of cervigram classification, this statistic does not allow for an assessment of the extent of misclassification as is achieved with the measures sensitivity, specificity, and predictive value, thereby limiting its usefulness as a measure of the validity of a screening method. For example, even though the positive 3 category achieved a likelihood ratio of 425 for the detection of high grade squamous intraepithelial lesions or cancer, this category only accounted for 5.2% of these serious lesions (Table 5a).

We stratified sensitivity and specificity by characteristics of the woman and of her cervigram in an attempt to further understand the performance of cervicography. Characteristics of the cervigram were noted during the review phase of evaluation (i.e., the second review following enrollment). Because many of these factors were not

assessed during the enrollment study, we were unable to determine whether the initial evaluator would have agreed with the second evaluator for the review phase on these characteristics. Similar to our initial finding during enrollment, we found that cervicography is significantly more sensitive in women younger than age 50 and in premenopausal women compared to women ages 50 and older and postmenopausal women respectively. However, the stratum-specific sensitivity is only slightly higher in women younger than age 50 and in premenopausal women than the crude sensitivity. The marked reduction in sensitivity in postmenopausal women is not surprising and is probably associated with positional change in the transformation zone. Most cervical neoplasia occurs at the transformation zone, which moves cephalad into the endocervical canal as a woman ages. Because the cervigram evaluator visualizes the projected image of the cervix, the technique does not allow for the detection of lesions far inside the endocervical canal. The usefulness of cervicography in postmenopausal women and/or in women ages 50 and older is therefore very limited.

Of the visual characteristics of the cervigram that were observed, only the presence of metaplasia, cervicovaginal atrophy, appearance of the acetic acid effect, and quality of the acetic acid effect were associated with sensitivity. Sensitivity was significantly higher in women with metaplasia compared to women without metaplasia visible on the ectocervix. Metaplasia is the process by which new squamous epithelial cells replace columnar epithelial cells in the transformation zone. Thus, women with metaplasia present are more likely to be younger and therefore have serious cervical abnormality detected by cervicography because their transformation zone would still be

visible on the ectocervix. In our study, the presence of metaplasia was significantly associated with age group (50+ versus <50) and menopausal status (Fisher's Exact test $p<0.001$ for both comparisons). Women with metaplasia visible on their cervigrams were more likely to be younger than age 50 and/or premenopausal. These associations were replicated when women were stratified by category of referent diagnosis. However, among women with high grade squamous epithelial lesions or cancer, the proportions of women ages 50 and older with and without metaplasia visible on their cervigrams were not as distinct, and the proportions of postmenopausal women with and without metaplasia were the same. A possible explanation for this observation is that visibility of the transformation zone on the ectocervix facilitated detection of the lesions in some older women. The absence of cervicovaginal atrophy was also associated with the sensitivity of cervicography. This is likely explained by the observation in our study that atrophy is more frequently present in older and/or postmenopausal women, yielding statistically significant associations between atrophy and both age group and menopausal status (Fisher's Exact test $p<0.001$ for both comparisons).

As expected, all women with high grade squamous intraepithelial lesions or cancer who also had a positive cervigram were observed with a visible acetic acid effect. Though the presence of an acetic acid effect was statistically associated with sensitivity, only 13 of the 134 (9.7%) women with serious neoplasia had a negative cervigram and no visible acetic acid effect, therefore only explaining 21.7% of the 60 false negative cervigram results. A visible acetic acid effect is necessary for the detection of cervical lesions with cervicography, and therefore the absence of such an effect would result in

higher specificity because women without neoplasia would not be falsely referred by cervicography. Similarly, in women in whom an acetic acid effect was observed, the quality of the effect was inversely associated with specificity. Women with a good acetic acid effect were significantly more likely to yield low specificity (i.e., high false positivity) than women with a fair or poor acetic acid effect. It is important to note that the presence of an acetic acid effect was also associated with younger age; 84.0% of women younger than age 50 had a visible acetic acid effect, while only 29.7% had a visible acetic acid effect. This finding may be explained by the fact that the transformation zone in older women had moved inside the cervical canal, and it is this area where the acetic acid effect would be visualized. The quality of the acetic acid effect was also associated with sensitivity among those women in whom an acetic acid effect was present, with sensitivity increasing with improved quality of the acetic acid effect. This trend was even more pronounced when analyses were limited to women younger than age 50, in whom the sensitivity increased to 90.4% in women with a good quality of acetic acid effect. However, specificity was reduced to 76.5% in this subgroup. These data suggest that it is possible, therefore, that better attention to the application and/or re-application of acetic acid may enhance the performance of the cervicography technique, especially in the younger age group, though the lowered specificity is concerning.

Several of the visual characteristics observed were significantly associated with the specificity of cervicography, including age group (50+ versus <50), post menopausal status, entirety of the lesion visible, metaplasia not visible, columnar epithelium visible, absence of a congenital transformation zone, presence of cervicovaginal atrophy, the

transformation zone partially obscured by mucus, the transformation zone partially obscured by the vaginal wall, the absence or poorer quality of an acetic acid effect, and a visible appearance of friability. Statistical significance achieved from some apparently small differences in specificity may be explained by the higher power of these analyses, due to the larger numbers of women without serious neoplasia (i.e., the denominator of the equation used to calculate specificity). There are possible explanations for each of these associations. Older and postmenopausal women are less likely to be referred for colposcopy because their transformation zone may not be visible on the ectocervix. Therefore, if they truly do not have neoplasia, they are less likely than younger and premenopausal women to be misclassified as positive by cervicography, because such misclassification may be due to acetowhitening of either trivial lesions or normal tissue around the transformation zone. A similar explanation would apply to the observed higher specificity in women with cervicovaginal atrophy present, who are also more likely to be in the older age group.

The variable "entirety of the lesion visible" only applies to women in whom a lesion was detected. In these women, specificity was significantly greater in women in whom the lesion was entirely visible than in women in whom the lesion was only partially visible. This finding is expected because the algorithm for classifying cervigrams errs on the side of caution with assignment in the positive 1A category. A positive 1A result is provided for lesions that are compatible with trivial disease but colposcopy is recommended because part of the lesion extends into the canal and cannot be completely visualized. In this study, a positive 1A classification was provided more

frequently by the second evaluator alone than it was following arbitration. Results for the reconstituted sample indicate that the second evaluator assigned a positive 1A classification to 236 women, whereas only 26 women retained this result after the arbitration process, suggesting poor interrater reliability for this category of cervigram classification. None of these 26 women had their entire lesion visible. For the overall reconstituted sample, only 1 of the 236 women (0.42%) with a positive 1A result by the second evaluator alone was observed with the entirety of the lesion visible. It is possible that misclassification on the positive 1A cervigram result may be due to poor interobserver agreement on whether or not the lesion is seen in its entirety. Since our assessment of visual characteristics of cervigrams is based on the secondary cervigram reviews, these analyses assume accuracy by the second evaluator on these characteristics. Therefore, some caution is warranted, and these findings should be considered suggestive rather than conclusive.

There was no difference in specificity between women in whom unaltered columnar epithelium was visible and in women with no columnar epithelium visible. However, specificity was significantly lower in women with altered columnar epithelium and in women with metaplasia visible on the ectocervix. This could be explained by the acetowhitening effect of metaplasia which could result in normal metaplastic cervices being misclassified as positive by cervicography.

A congenital transformation zone becomes visible as a ring of dense acetowhite tissue on the cervix. The acetowhitening and the dense appearance likely contribute to misclassification of normal cervices as positive, therefore explaining the low specificity

among women with a congenital transformation zone. Nonetheless, only a small proportion of women had a congenital transformation zone observed [n = 40 (0.5%)], so this finding should be interpreted cautiously.

There is no biologic explanation for women having their transformation zone partially obscured by mucus or the vaginal wall to yield a higher specificity by cervicography, except that any portion partially obscured may be more likely to be classified as normal if the visible portions of the cervix are also normal, resulting in a tendency toward true negative results more frequently than false positive results. However the difference in specificity for women with and without mucus partially obscuring the transformation zone and with and without the vaginal wall partially obscuring the transformation zone is small (96.5 versus 94.2 and 96.9 versus 94.2 respectively), and it is also possible that this difference occurred by chance.

Although there was no difference in specificity for women with and without visible characteristics of inflammation on their cervigrams, the visible appearance of friability was significantly associated with specificity. It is possible that friability is characteristic of cervical infection, but sufficiently different from the acetowhite appearance of neoplasia to not result in misclassification of women without serious neoplasia by the referent diagnosis.

Several visual characteristics were associated with agreement on cervigram classification. Presence of acetowhite epithelium, mosaicism, punctation, and atypical vessels are all criteria used to assign a cervigram result. It is expected, then, that differing opinion on whether or not these characteristics are present would affect agreement on

cervigram classification. Visibility of the entirety of the lesion, metaplasia, columnar epithelium, congenital transformation zone, cervicovaginal atrophy, acetic acid effect, quality of the acetic acid effect, characteristics of inflammation, degree of inflammation, and redness were associated with agreement on cervigram classification. Many of these same characteristics were also associated with sensitivity and specificity. It is therefore likely that the presence (or absence) of one or more of these visual characteristics contributes to difficulty in cervigram interpretation and potential for misclassification.

Some caution must be taken in interpreting the influence of the visual characteristics on cervigrams as noted above. Women who were reviewed in both the review and arbitration phases generally had poor to moderate agreement on the visual characteristics beyond that expected by chance alone. However, where agreement on visual characteristics was associated with agreement on cervigram classification, most of these are criteria used to guide the classification itself. Visibility of the entirety of the lesion also yielded a significant association between agreement on the presence of the characteristic and agreement on cervigram classification. As discussed earlier, this characteristic may also be considered a guiding criterion for classification of cervigrams.

Limitations of this study are that not all cervigrams evaluated during the enrollment phase of the study were reevaluated in the review phase and that not all cervigrams reviewed underwent arbitration. However, 43% of all eligible cervigrams were reviewed for this study, including 100% of all women with at least one abnormal screening test during the enrollment phase, 100% of women at high risk for cervical abnormalities, and a randomly selected sample of 22% of all women with all normal

screening tests and no risk factors. We were able to extrapolate the results of the review phase to the larger sample by multiplying the results for the subsample of women who were randomly selected by the inverse of the sampling fraction for this group. Women with discordant cervigram results by the enrollment and subsequent evaluators, plus a 10% sample of women with concordant results were arbitrated by a third evaluator. The subsample of women with concordant results by the enrollment and subsequent evaluators were included to mask the sample composition. This cervigram review algorithm allowed us to establish a revised cervigram result based on agreement by at least two evaluators. The cost of evaluating all cervigrams in triplicate would have been prohibitive for this study. Nonetheless, we feel that this algorithm has enabled us to develop a reasonable estimate of an optimized cervigram classification for each woman. National Testing Laboratories Worldwide has since put in place an arbitration process for all cervigrams in which two evaluators independently evaluate the cervigrams, a third evaluator arbitrates, and the initial evaluator receives feedback and an opportunity to revise his/her classification.

Cervigrams in this study were reported (by the third evaluator) to show more blood than is usually seen in cervigrams. Excessive bloodiness may have contributed to high false positivity associated with positive 0 cervigrams because bloodiness is one of the guiding criteria for a positive 0 classification. In our study, of the 46 women with a revised cervigram result of positive 0, 17 (37%) had cervigrams reported with the transformation zone partially obscured by blood. Of these, 15 (88.2%) corresponded to women with a referent diagnosis of low grade squamous intraepithelial lesion, equivocal,

or negative. It is possible that bleeding was due to the specimen collection protocol that required cervicography photographs to be taken following cytologic sampling and therefore may have been affected by excessive scraping. In our protocol, cervicography was not performed before cytologic sampling because of a concern that application of acetic acid might interfere with the cytology results. An alternate explanation is that at least some of the bleeding was associated with high rates of cervical infection in this population. This explanation is partially supported by the finding that 30% of all women in this study were observed to have some degree of cervical inflammation based on the appearance of their cervigrams.

Digital colposcopic images were reviewed to assess the perceived validity of the initial colposcopy result. However, it should be noted that the quality of these images was deemed by the reviewer to be less than optimal, so the results corresponding to the digital colposcopic image review should be interpreted with caution. The evaluator of these images noted an apparent deficiency in the application of acetic acid. As a result, mucus was not adequately removed from many of these images, thus impairing the visualization of the cervix. Additionally, documentation was poor, so it was not always clear which part of the cervix was visible in the image. This was more common with low grade lesions than with high grade lesions. These observations are believed to reflect more on shortcomings in application of acetic acid and documentation than on the quality of the technique itself.

Few, if any, referent standard tests are known to be 100% sensitive and 100% specific. When screening tests are evaluated against an imperfect reference test, and the

screening and reference tests are independent of each other, the sensitivity and specificity of the screening method will be underestimated.[28,29] Therefore, some possibility exists that the true sensitivity and specificity could be higher than those reported in this study. It is unlikely, however, that such underestimation would be considerable because of the efforts made to optimize the referent diagnosis. The enrollment referent diagnosis was made after several rounds of pathologic reviews, which should contribute to a more accurate referent standard. In earlier work, Buck and Gart[30] demonstrated that, when an imperfect reference test is used as the standard, as the true prevalence of the disease increases, the reported sensitivity of the screening test increases and the reported specificity of a screening test decreases. When an imperfect reference test is used, the "apparent" total number of individuals with the disease of interest is more accurately the sum of the number of individuals with true disease who are detected by the reference test plus the number of individuals without true disease who are misclassified as positive by the referent test. Similarly, when an imperfect reference test is employed, the apparent total number of individuals without true disease is more accurately the sum of the number of individuals without true disease who are detected by the reference test plus the number of individuals with true disease who are misclassified as negative by the reference test. The "apparent sensitivity" (referred to as co-positivity by Buck and Gart), then, is the proportion of the apparent number of individuals with disease (i.e., which contains some misclassified individuals) who are classified as positive by the screening test. Similarly, the "apparent specificity" (referred to as co-negativity by Buck and Gart) is the proportion of the apparent number of individuals without disease (i.e., which contains some

misclassified individuals) who are classified as negative by the screening test. Buck and Gart show that if different rates of prevalence (i.e., different numbers of individuals with true disease in a population, or among specific age strata) are applied to the above formulation, the resulting estimates of "apparent sensitivity" and "apparent specificity" will vary in such a way that as the true prevalence increases, the apparent sensitivity also increases, while the apparent specificity decreases. Since the present study was conducted in an area with high rates of cervical cancer, it is possible that the reported sensitivity of cervicography in this study may be higher and the reported specificity may be lower than they would be if the study were conducted in a low prevalence area. Similar interpretations may result when assessing the performance of a screening test in different subgroups which experience different category-specific prevalence rates of disease.

A strength of this study is the large, population based sample in which it was conducted. The large sample size permitted identification of a sufficiently large number of women with high grade squamous intraepithelial lesions or cancer to assess the performance of cervicography for these serious lesions separately from low grade squamous intraepithelial lesions. The reviews of cervigrams allowed us to establish an optimal cervigram result for each woman. The reviews of digital colposcopic images allowed us to stratify results for which the referent diagnosis was possibly misclassified because of error in the initial decision to take a biopsy and/or appropriateness of biopsy placement. A final strength of this study was our assessment of how various

characteristics of the woman and her cervigram affected sensitivity and specificity and agreement on cervigram classification.

In summary, this study demonstrates that the optimal performance of cervicography based on arbitrated reviews was only slightly improved over cervicography screening using a single review. Cervicography is subject to fair to moderate interobserver agreement, and it detects significantly fewer high grade squamous intraepithelial lesions than does the Papanicolaou smear in a mass screening setting. However, sensitivity is high for detecting invasive cancer and is similar to that of conventional cytology. Cervicography is of limited use in women ages 50 and older and in postmenopausal women, as the sensitivity drops markedly in these groups.

References

1. Herrero R, Schiffman MH, Bratti C, Hildesheim A, Balmaceda I, Sherman ME, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: The Guanacaste Project. Pan Am J Public Health 1997; 362-375.
2. Schneider DL, Herrero R, Bratti C, Greenberg MD, Hildesheim A, Sherman ME, et al. Cervicography screening for cervical cancer among 8,460 women in a high-risk population. Am J Obstet Gynecol 1999; 180:290-298.
3. Schneider DL, Herrero R, Bratti C, Greenberg MD, Hildesheim A, Sherman ME, et al. Cervicography screening for cervical cancer among 8,460 women in a high-risk population (Correction). Am J Obstet Gynecol 1999; 180:1592-1592.
4. Schiffman MH, Kiviat NB, Burk RD, Shah KV, Daniel RW, Lewis R, et al. Accuracy and interlaboratory reliability of human papillomavirus DNA testing by Hybrid Capture. J Clin Microbiol 1995; 33:545-550.
5. Lörincz AT, Reid R, Jenson AB, Greenberg M, Lancaster W, Kurman RJ. Human papillomavirus infection of the cervix: relative risk associations of 15 common anogenital types. Obstet Gynecol 1992; 79:328-337.
6. Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J Natl Cancer Inst 1995; 87:796-802.

7. Schiffman M, Herrero R, Hildesheim A, Sherman ME, Bratti C, Wacholder S, et al. HPV DNA testing for cervical cancer screening: Results from the 9,000 women in the NCI Guanacaste Project. JAMA 2000; 283:87-93.

8. Sherman ME, Schiffman MH, Herrero R, Kelly D, Bratti C, Mango L, et al. Performance of a semiautomated Papanicolaou smear screening system: results of a population-based study conducted in Guanacaste, Costa Rica. Cancer 1998; 84:273-280.

9. Kurman RJ, Soloman D. The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses: Definitions, Criteria, and Explanatory Notes for Terminology and Specimen Adequacy. New York: Springer-Verlag, 1994.

10. Fleiss JL. Statistical Methods for Rates and Proportions. Second ed. New York: John Wiley and Sons, 1981.

11. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991.

12. Daniel WW Biostatistics: A Foundation for Analysis in the Health Sciences. 6th ed. New York: John Wiley and Sons, Inc., 1995.

13. Jaeschke R, Guyatt GH, Sackett DL. Users' guide to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients. JAMA 1994; 271:703-707.

14. Sox HC, Blatt MA, Higgins MC, Marton KI. Medical decision making. Boston: Butterworth-Heinemann, 1988.

15. Cocchi V, Sintoni C, Carretti D, Sama D, Chiari U, Segala V, et al. External quality assurance in cervical/vaginal cytology: Interlaboratory agreement in the Emilia-Romagna region of Italy. Acta Cytol 1996; 40:480-488.

16. Cocchi V, Carretti D, Fanti S, Baldazzi P, Casotti MT, Piazzi R, et al. Intralaboratory quality assurance in cervical/vaginal cytology: Evaluation of intercytologist diagnostic reproducibility. Diagn Cytopathol 1997; 16:87-92.

17. Confortini M, Biggeri A, Cariaggi MP, Carozzi FM, Minuti PA, Russo A, et al. Intralaboratory reproducibility in cervical cytology: results of the application of a 100-slide set. Acta Cytol 1993; 37:49-54.

18. Curry H, Thompson DW, Dietrich M, Lipa M, Massarella GR, Taves IR, et al. Proficiency testing in cytology laboratories in Ontario, Canada: A decade of experience. III. A precision study of consistency and reproducibility in cytology reporting. Acta Cytol 1987; 31:220-225.

19. de Ruiz PA, Lazcano Ponce EC, Duarte Torres R, Ruiz Juárez I, Martínez Cortez I. Diagnostic reproducibility of Pap testing in two regions of Mexico: The need for quality control mechanisms. Bull Pan Am Health Organ 1996; 30:330-338.

20. Duca P, Braga M, Chiappa L, Piffer R, Turolla E. Intralaboratory reproducibility of interpretation of Pap smears: Results of an experiment. Tumori 1988; 74:737-734.

21. Jones S, Thomas GDH, Williamson P. Observer variation in the assessment of adequacy and neoplasia in cervical cytology. Acta Cytol 1996; 40:226-234.

22. Joste NE, Rushing L, Granados R, Zitz JC, Genest DR, Crum CP, et al. Bethesda classification of cervicovaginal smears: Reproducibility and viral correlates. Hum Pathol 1996; 27:581-585.

23. Kern WH, Zivolich MR. The accuracy and consistency of the cytologic classification of squamous lesions of the uterine cervix. Acta Cytol 1977; 21:519-523.

24. Klinkhamer PJJM, Vooijs GP, de Haan AFJ. Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears. Acta Cytol 1988; 32:794-800.

25. Lazcano-Ponce EC, Alonso de Ruiz P, López Carrillo L, Nájera Aguilar P, Avila Ceniceros R, Escandón Romero C, et al. Validity and reproducibility of cytologic diagnosis in a sample of cervical cancer screening centers in Mexico. Acta Cytol 1997; 41:277-284.

26. Sherman ME, Schiffman MH, Lorincz AT, Manos MM, Scott DR, Kurman RJ, et al. Toward objective quality assurance in cervical cytopathology. Am J Clin Pathol 1994; 102:182-187.

27. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. Am J Epidemiol 1995; 141:680-689.

28. Phelps CE, Hutson A. Estimating diagnostic test accuracy using a "fuzzy gold standard". Med Decis Making 1995; 15:44-57.

29. Valenstein PN. Evaluating diagnostic tests with imperfect standards. Am J Clin Pathol 1990; 93:252-258.

30. Buck AA, Gart JJ. Comparison of a screening test and a reference test in epidemiologic studies: I. Indices of agreement and their relation to prevalence. Am J Epidemiol 1966; 83:586-592.

Table 1
Cervigram Classification Scheme*

Not referred for colposcopy:

Negative:	No lesion seen
Atypical 1 (A1):	A trivial lesion inside the transformation zone is visible, but colposcopy not recommended because of the benign appearance or site of the lesion
Atypical 2 (A2):	A trivial lesion outside the transformation zone is visible, but colposcopy not recommended because of the benign appearance or site of the lesion
Technically defective:	Unable to be properly evaluated

Referred for colposcopy:

Positive 0 (P0):	Probably normal, but colposcopy preferable to rule out serious neoplasia
Positive 1A (P1A):	Compatible with trivial disease but colposcopy recommended because part of the lesion extends into the canal
Positive 1B (P1B):	Compatible with a low-grade squamous intraepithelial lesion/ flat condyloma/ and exophytic condyloma
Positive 2 (P2):	Compatible with a high grade squamous intraepithelial lesion
Positive 3 (P3):	Compatible with cancer

* Note that as of January 1, 1995, National Testing Laboratories Worldwide revised the atypical category. Prior to January 1, 1995 atypical 1 referred to trivial lesions outside the transformation zone and atypical 2 referred to trivial lesions inside the transformation zone. Current terminology is applied to all cervigram classification in this report.

Table 2
Referent Diagnosis

Cancer:	Histologically-confirmed invasive cancer
HSIL:	Histologically-confirmed high grade squamous intraepithelial lesion
HSIL2:	Women with a conventional Papanicolaou smear and/or PapNet result of high grade squamous intraepithelial lesion, plus a ThinPrep result of high grade squamous intraepithelial lesion, but no histologic confirmation
LSIL:	Histologically-confirmed high grade squamous intraepithelial lesion
LSIL2:	Women with no histologic confirmation of a squamous intraepithelial lesion (SIL) and at least two of the following criteria met: 1) a conventional Papanicolaou smear or PapNet result of high grade squamous intraepithelial lesion; 2) a ThinPrep diagnosis of high grade squamous intraepithelial lesion; or 3) a cervigram result of positive 1, positive 2, or positive 3
Equivocal – NL/HSIL:	Women with a differential diagnosis of high grade squamous intraepithelial lesion versus negative on final review (severe atrophy contributed to this diagnostic category)
Equivocal – NL/LSIL:	Women whose overall results were equivocal, even following review by the chief study pathologist
Equivocal / ThinPrep:	Women with a ThinPrep cytologic diagnosis of high grade squamous intraepithelial lesion and all other screening tests normal
Equivocal / Pap:	Women with either a conventional Papanicolaou smear or PapNet diagnosis of high grade squamous intraepithelial lesion and all other screening tests normal
Equivocal / Cervigram:	Women with a cervigram result of positive 0, positive 1, positive 2, or positive 3 and no cytologic or histologic confirmation
Normal2:	Women referred to colposcopy with a cytologic diagnosis of atypical squamous cells of undetermined significance who were normal after review
Normal 1B:	Grossly visible abnormality, screening tests normal
Normal:	Women with all negative screening results

Table 3

Selection categories into cervicography review, sampling fractions, number selected, and multipliers used to reconstitute the original study sample

Selection into cervicography review			
Criterion	**Percent Selected**	**n**	**Multiplier**
HSIL or cancer at enrollment	100%	136	1.0
At least one abnormal screening test at enrollment	100%	1610	1.0
Tested positive for HPV at enrollment	100%	298	1.0
Five or more lifetime sexual partners	100%	388	1.0
Selected as controls for followup study	100%	513	1.0
Did not meet above criteria	12.7%	700	7.879

Table 4
Distribution of the referent diagnosis by revised cervigram result [a]
(Number and column percent)

Revised Cervigram Result [c]	Referent Diagnosis [b]									
	Normal	Equivocal	Equivocal Pap	Equivocal ThinPrep	LSIL2	LSIL	HSIL2	HSIL	Cancer	Total
Negative	5953	681	149	111	64	16	7	35	1	7017
	88 6	62.5	81.4	78.2	56.6	21 9	100 0	30.2	9 1	83.0
Atypical	377	60	16	14	10	11	0	10	0	498
	5 6	5.5	8.7	9.9	8.8	15 1	0 0	8.6	0.0	5.9
Technically Defective	81	14	2	0	0	1	0	0	0	98
	1.2	1.3	1.1	0.0	0.0	1 4	0 0	0.0	0.0	1.2
Total Not Referred	6411	755	167	125	74	28	7	45	1	7613
	95 4	69.3	91.3	88.0	65.5	38 4	100.0	38.8	9.1	90.1
Positive 0	0	42	0	0	1	1	0	2	0	46
	0 0	3.9	0.0	0.0	0.9	1.4	0.0	1.7	0.0	0.5
Positive 1	48	163	2	2	25	36	0	37	0	313
	0.7	15.0	1.1	1.4	22.1	49.3	0.0	31.9	0.0	3.7
Positive 2	9	20	0	0	1	5	0	17	3	55
	0 1	1.8	0.0	0.0	0.9	6.8	0 0	14.7	27.3	0.7
Positive 3	0	1	0	0	0	0	0	3	4	8
	0.0	0.1	0.0	0.0	0.0	0.0	0.0	2.6	36.4	0.1
Total Positive (Referred)	57	226	2	2	27	42	0	59	7	422
	0.8	20.8	1.1	1.4	23.9	57.5	0.0	50.9	63.6	5.0
3-way disagreement	250	108	14	15	12	3	0	12	3	417
	3.7	9.9	7.7	10.6	10.6	4.1	0.0	10.3	27.3	4.9
Total	6718	1089	183	142	113	73	7	116	11	8452

a. 8 women with missing cervigram results are excluded
b. In Table 2: Normal = Normal, Normal 1B Equivocal = Equivocal-NL/HSIL, Equivocal-NL/LSIL, Equivocal/Cervigram, Normal2; all other categories as indicated in Table 2.
c. The cervigram classification scheme is provided in Table 1.

Table 5a

Likelihood ratios using the revised cervicography[a] and enrollment referent diagnosis[b] results for the detection of squamous intraepithelial lesions or cancer[c]

Revised cervigram result	no. HSIL or cancer	no. LSIL or LSIL2	no. normal or equivocal	Likelihood ratio HSIL or cancer	Likelihood ratio LSIL or LSIL2
Positive 3	7	0	1	424.8	0.0
Positive 2	20	6	29	41.9	9.0
Positive 1	37	61	215	10.4	12.4
Positive 0	2	2	42	2.9	2.1
Atypical	10	21	467	1.3	2.0
Negative	43	80	6894	0.4	0.5
3-way Disagreement	15	15	387	2.4	1.7
Technically Defective	0	1	97	0.0	0.5
Total	134	186	8132		

HSIL: High grade squamous intraepithelial lesion, LSIL: Low grade squamous intraepithelial lesion

a. The cervigram classification scheme is provided in Table 1

b. Definitions of the referent diagnosis are provided in Table 2

c. 8 women with missing cervigram results are excluded

Table 5b

Likelihood ratios using the dichotomized, revised cervicography[a] and enrollment referent diagnosis[b] results for the detection of squamous intraepithelial lesions or cancer[c]

Revised cervigram result	no. HSIL or cancer	no. LSIL or LSIL2	no. normal or equivocal	Likelihood ratio HSIL or cancer	Likelihood ratio LSIL or LSIL2
Positive	74	70	405	11.1	7.6
Negative, Atypical or technically defective	60	116	7727	0.5	0.7
Total	134	186	8132		

HSIL: High grade squamous intraepithelial lesion, LSIL: Low grade squamous intraepithelial lesion

a. The cervigram classification scheme is provided in Table 1

b. Definitions of the referent diagnosis are provided in Table 2

c. 8 women with missing cervigram results are excluded

Table 5c

Likelihood ratios using the conventional Papanicolaou smear and referent diagnosis[a] results for the detection of squamous intraepithelial lesions or cancer[b]

Conventional Pap result	no. HSIL or cancer	no. LSIL or LSIL2	no. normal or equivocal	Likelihood ratio HSIL or cancer	Likelihood ratio LSIL or LSIL2
Cancer	21	0	3	418.1	0
HSIL	64	33	36	106.2	39.4
LSIL	15	75	176	5.1	18.3
ASCUS	5	7	151	2.0	2.0
Reactive Changes	16	37	2946	0.3	0.5
Normal	13	37	4757	0.2	0.3
Unsatisfactory	2	0	55	2.2	0.0
Total	136	189	8124		

HSIL: High grade squamous intraepithelial lesion; LSIL: Low grade squamous intraepithelial lesion
ASCUS: Atypical squamous cells of undetermined significance
a. Definitions of the referent diagnosis are provided in Table 2
b. 11 women for whom no Papanicolaou smear was take or for whom the results are missing are excluded

Table 5d

Likelihood ratios using the conventional Papanicolaou smear and referent diagnosis[a] results for the detection of squamous intraepithelial lesions or cancer[b]

Conventional Pap result	no. HSIL or cancer	no. LSIL or LSIL2	no. normal or equivocal	Likelihood ratio HSIL or cancer	Likelihood ratio LSIL or LSIL2
Cancer HSIL LSIL ASCUS	105	115	366	17.1	13.5
Reactive Changes Normal	29	74	7703	0.2	0.4
Unsatisfactory	2	0	55	2.2	0.0
Total	136	189	8124		

HSIL: High grade squamous intraepithelial lesion; LSIL: Low grade squamous intraepithelial lesion
ASCUS: Atypical squamous cells of undetermined significance
a. Definitions of the referent diagnosis are provided in Table 2
b. 11 women for whom no Papanicolaou smear was take or for whom the results are missing are excluded

Table 6

Sensitivity and specificity of revised cervigram results[a] by characteristics of the woman, her cervigram,[b] and digital colposcopic images[b]

Characteristic	Referent diagnosis[c] is high grade lesion or cancer		Referent diagnosis[c] less severe than HSIL or cancer	
	Cervigram Positive No. [Sensitivity (%)]	p-value[d]	Cervigram Negative No. [Specificity (%)]	p-value[d]
Age group				
<50	67 (62.0)		5948 (93.3)	
50+	7 (26.9)	<0.01	1895 (97.5)	<0.01
Still having menstrual periods?				
Yes	65 (62.5)		5764 (93.3)	
No	9 (30.0)	<0.01	2079 (97.1)	<0.01
Entirety of the lesion seen (if applicable)?				
Yes	35 (77.8)		554 (71.0)	
No	33 (84.6)	0.6	239 (59.3)	<0.01
Metaplasia visible on the ectocervix?				
Yes	70 (60.9)		5209 (92.4)	
No	4 (21.1)	<0.01	2620 (98.2)	<0.01
Columnar epithelium visible?				
Yes	4 (100.0)		627 (95.6)	
Yes, altered	37 (55.2)		3194 (92.7)	
No	33 (52.4)	0.2	4018 (95.4)	<0.01
Congenital transformation zone?				
Yes	1 (50.0)		20 (52.6)	
No	73 (55.3)	1.0	7809 (94.5)	<0.01
Cervicovaginal atrophy present?				
Yes	2 (20.0)		868 (98.5)	
No	72 (58.1)	0.04	6968 (93.8)	<0.01
Transformation zone partially obscured by:				
Blood				
Yes	23 (59.0)		1765 (95.2)	
No	51 (53.7)	0.7	6078 (94.0)	0.08

Speculum				
Yes	0 (0.0)		25 (92.6)	
No	74 (56.1)	0.2	7818 (94.3)	0.7
Hair				
Yes	0 (0.0)		0 (0.0)	
No	74 (55.2)	N/A	7843 (94.3)	N/A
Mucus				
Yes	4 (80.0)		363 (96.5)	
No	70 (54.3)	0.4	7480 (94.2)	0.05
Vaginal wall				
Yes	2 (33.3)		346 (96.9)	
No	72 (56.3)	0.4	7497 (94.2)	0.03
Position of the cervix				
Yes	1 (20.0)		268 (94.0)	
No	73 (56.6)	0.2	7575 (94.3)	0.8
Acetic acid effect visible?				
Yes	74 (61.2)		5450 (92.4)	
No	0 (0.0)	<0.01	2377 (99.0)	<0.01
Quality of acetic acid effect				
Good	52 (88.1)		1045 (76.5)	
Fair	20 (37.0)		3268 (96.8)	
Poor	2 (25.0)	<0.01	1140 (98.4)	<0.01
Visual characteristics of				
inflammation present?				
Yes	28 (52.8)	0.7	2511 (94.7)	
No	46 (56.8)		5280 (94.1)	0.2
Degree of inflammation				
Minor	17 (53.1)		1592 (95.1)	
Moderate	10 (52.6)		811 (94.1)	
Severe	1 (50.0)	1.0	134 (94.4)	0.5
Visual characteristics				
of inflammation				
Friability				
Yes	20 (50.0)		2004 (96.5)	
No	54 (57.5)	0.5	5839 (93.5)	<0.01
Redness				
Yes	18 (48.7)		1818 (94.0)	
No	56 (57.7)	0.4	6025 (94.4)	0.5

Mucopus				
Yes			16 (94.1)	
No	74 (55.2)	N/A	7827 (94.3)	1.0
Agreement on decision to biopsy?[e]				
Yes	39 (53.4)		1110 (82.8)	
No	26 (70.2)	0.1	270 (72.0)	<0.001
Agreement on biopsy placement?[f]				
Yes	21 (72.4)		65 (46.8)	
No	3 (60.0)	0.6	16 (38.1)	0.4

a The cervigram classification scheme is provided in Table 1

b Visual cervigram image characteristics were determined during the review round of cervigram evaluation

c The referent diagnosis is classified as high grade squamous intraepithelial lesions or cancer versus normal, equivocal, or low grade squamous intraepithelial lesion.

d Determined using a two-sided Fisher's Exact Test to assess whether an association exists between the characteristic and the cervigram result within each category of the referent diagnosis.

e Analysis limited to women who were referred for colposcopy with available digital colposcopic images

f. Analysis limited to women who were referred for colposcopy with available digital colposcopic images for whom a biopsy was taken and there was agreement on the decision to take a biopsy.

Table 7

Agreement between initial and second cervigram[a] results by characteristics of the woman, her cervigram[b], and digital colposcopic image[c]

Characteristic	Cervigram results discordant on narrow categories[d] No. (Column %)	Cervigram results concordant on narrow categories[d] No. (Column %)	p-value[f]	Cervigram results discordant on broader categories[e] No. (Column %)	Cervigram results concordant on broader categories[e] No. (Column %)	p-value[f]
Age group						
<50	664 (80 7)	2176 (77 2)		330 (80 3)	2510 (77.7)	
50+	159 (19 3)	644 (22 8)	0 04	81 (19 7)	722 (22 3)	0 3
Still having menstrual periods?						
Yes	649 (78 9)	2114 (75 0)		323 (78 6)	2440 (75 5)	
No	174 (21 1)	706 (25 0)	0 02	88 (21 4)	792 (24 5)	0 2
Acetowhite epithelium present?						
Yes	407 (49 5)	287 (10 2)		228 (55 5)	466 (14 4)	
No	416 (50 5)	2533 (89 8)	<0 001	183 (44 5)	2766 (85 6)	<0 001
Discoloration present?						
Yes	1 (0 1)	2 (0 1)		0 (0 0)	3 (0 1)	
No	822 (99 9)	2818 (99 9)	0 5	411 (100 0)	3229 (99 9)	1 0
Erosion/ulceration present?						
Yes	6 (0 7)	7 (0 3)		4 (1 0)	9 (0 3)	
No	817 (99 3)	2813 (99 7)	0 09	407 (99 0)	3223 (99.7)	0 05
Irregular surface contour?						
Yes	9 (1 1)	9 (0 3)		2 (0 5)	16 (0 5)	
No	814 (98 9)	2811 (99 7)	<0 001	409 (99 5)	3216 (99 5)	1 0
Mosaicism present?						
Yes	26 (3 2)	18 (0 6)		12 (2 9)	32 (1 0)	
No	797 (96 8)	2802 (99 4)	<0 001	399 (97 1)	3200 (99 0)	<0 001
Punctation present?						
Yes	33 (4 0)	17 (0 6)		15 (3 7)	35 (1 1)	
No	790 (96 0)	2803 (99 4)	<0 001	396 (96 4)	3197 (98 9)	<0 001
Atypical vessels present?						
Yes	27 (3 3)	19 (0 7)		19 (4 6)	27 (0 8)	
No	796 (96 7)	2801 (99 3)	<0 001	392 (95 4)	3205 (99 2)	<0 001
Entirety of the lesion seen?						
Yes	294 (35 7)	221 (70 6)		160 (58 8)	355 (70 0)	
No	172 (20 9)	92 (29.4)	0 03	112 (41 2)	152 (30 0)	<0 001
Metaplasia visible on the ectocervix?						
Yes	680 (82 6)	1913 (67 9)		356 (86 6)	2237 (69 4)	
No	134 (16 3)	903 (32 1)	<0 001	49 (11.9)	968 (30 6)	<0 001

Columnar epithelium visible?						
Yes	45 (5 5)	224 (7 9)		19 (4 6)	250 (7 7)	
Yes, altered	392 (47 6)	1160 (41 2)		196 (47 7)	1356 (42 0)	
No	381 (46 3)	1433 (50 9)	<0 001	191 (46 5)	1623 (50 3)	0 01
Congenital transformation zone?						
Yes	13 (1 6)	13 (0 5)		6 (1 5)	20 (0 6)	
No	802 (97 4)	2802 (99 5)	<0 001	399 (97 0)	3205 (99 4)	0 06
Cervicovaginal atrophy present?						
Yes	63 (7 7)	279 (9 9)		24 (5 8)	318 (9 9)	
No	751 (91 3)	2536 (90 1)	0 07	380 (92 5)	2907 (90 1)	0 01
Transformation zone partially obscured by:						
Blood						
Yes	210 (25 5)	597 (21 2)		102 (24 8)	705 (21 8)	
No	613 (74 5)	2223 (78 8)	<0 001	309 (75 2)	2527 (78 2)	0 2
Speculum						
Yes	6 (0 7)	16 (0 6)		2 (0 5)	20 (0 6)	
No	817 (99 3)	2804 (99 4)	0 6	409 (99 5)	3212 (99 4)	1 0
Hair						
Yes	0 (0 0)	0 (0 0)		0 (0 0)	0 (0 0)	
No	823 (100 0)	2820 (100 0)	N/A	411 (100 0)	3232 (100 0)	N/A
Mucus						
Yes	29 (3 5)	116 (4 2)		11 (2 7)	136 (4 2)	
No	794 (96 5)	2702 (95 8)	0 4	400 (97 3)	3096 (95 8)	0 2
Vaginal wall						
Yes	37 (4 5)	81 (2 9)		16 (3 9)	102 (3 2)	
No	786 (95 5)	2739 (97 1)	0 03	395 (96 1)	3130 (96 8)	0 5
Position of the cervix						
Yes	48 (5 8)	91 (3 2)		17 (4 1)	122 (3 8)	
No	775 (94 2)	2729 (96 8)	<0 001	394 (95 9)	3110 (96 2)	0 7
Acetic acid effect visible?						
Yes	708 (86 0)	1997 (71 0)		371 (90 3)	2334 (72 2)	
No	106 (12 9)	816 (29 0)	<0 001	33 (8 0)	889 (27 5)	<0 001
Quality of acetic acid effect						
Good	322 (39 1)	484 (24 3)		174 (42 3)	632 (19 6)	
Fair	308 (37 4)	1127 (56 5)		152 (37 0)	1283 (39 7)	
Poor	75 (9 1)	384 (19 2)	<0 001	43 (10 5)	416 (12 9)	<0 001
Visual characteristics of inflammation present?						
Yes	300 (36 4)	877 (31 3)		153 (37 2)	1024 (31 7)	
No	512 (62 2)	1929 (68 7)	<0 001	250 (60 8)	2191 (67 8)	0 02

Degree of inflammation						
minor	152 (18 5)	571 (64 7)		82 (20 0)	641 (19 8)	
moderate	125 (15 2)	275 (31 2)		67 (16 3)	333 (10 3)	
severe	25 (3 0)	36 (4 1)	<0 001	6 (1 5)	55 (1 7)	0 03
Visual characteristics						
of inflammation present						
Friability						
Yes	205 (24 9)	688 (24 4)		104 (25 3)	789 (24 4)	
No	618 (75 1)	2132 (75 6)	0 7	307 (74 7)	2443 (75 6)	0 7
Redness						
Yes	237 (28 8)	614 (21 8)		119 (29 0)	732 (22 7)	
No	586 (71 2)	2206 (78 2)	<0 001	292 (71 1)	2500 (77 3)	<0 001
Mucopus						
Yes	0 (0 0)	3 (0 1)		0 (0 0)	3 (0 1)	
No	823	2817 (99 9)	1 0	411 (100 0)	3229 (99 9)	1 0
Agreement on decision to biopsy?[g]						
Yes	382 (74 0)	939 (80 8)		200 (73 3)	1121 (79 8)	
No	134 (26 0)	223 (19 2)	<0 001	73 (26 7)	284 (20 2)	0 02
Agreement on biopsy placement?[h]						
Yes	79 (82 3)	82 (73 2)		40 (80 0)	121 (76 6)	
No	17 (17 7)	30 (26 8)	0 1	10 (20 0)	37 (23 4)	0 7

a The cervigram classification scheme is provided in Table 1

b Cervigram characteristics were determined during the review round of cervigram evaluation

c Percentages in each category may not total 100 due to the absence of results with missing data, responses of "not applicable", or more than one possible response

d Agreement is based on the following groupings of cervigram results negative, atypical, positive 0, positive 1, positive 2, positive 3, or technically defective as separate categories

e Agreement is based on the following groupings of cervigram results: negative, atypical, or technically defective vs positive 0, positive 1, positive 2, or positive 3

f Determined using a two-sided Fisher's Exact test, with missing and nonapplicable categories excluded

g. Analysis limited to women who were referred for colposcopy with available digital colposcopic images.

h. Analysis limited to women who were referred for colposcopy with available digital colposcopic images for whom a biopsy was taken and there was agreement on the decision to take a biopsy

Overall Discussion

Cervicography was first described in 1981, initially as a potential surrogate for colposcopic examination and later as a first-tier screening test. Early studies showed the potential of cervicography as an alternate and/or complementary screening method to the Papanicolaou smear. However, initial assessment indicated that its low specificity was a concern, and the manufacturer made adjustments to the diagnostic classification scheme. Subsequently, studies in the 1990s suggested that the specificity improved, but at the expense of low sensitivity. Concerns about methodologic flaws in previous studies cast doubt among some clinicians and researchers that cervicography had been fairly evaluated. The present study was designed to assess the performance of cervicography in an unbiased setting that overcame some of the deficiencies of previous studies. The strengths of this study are that it is population based, that four screening methods were employed for referral for colposcopy, that women who screened negative by all methods were validated, and that reviews of clinical materials allowed for refinement and adjustment for possible error in the screening and referral process.

To evaluate the performance of cervicography screening, this study was designed to assess the validity of the method using the conventional measures of sensitivity and specificity. Validity of a screening test is defined as the ability of a test to distinguish between individuals with the disease of interest and those without the disease of interest. As in the present study, validity is usually measured against a gold standard diagnosis which is based on an external, more definitive diagnostic test (i.e., a gold standard test). The gold standard represents the state of the art for diagnosis. Positive and negative

predictive value are complementary measures used for evaluating a screening test which are particularly useful in clinical decision making.[1,2]

Evaluations of the general performance of a screening test, including an assessment of validity, differ from evaluations of the effectiveness and/or efficacy of a screening program. Evaluations of screening program efficacy are usually designed to assess whether mortality or morbidity is reduced among individuals screened. For cervical cancer, incidence would be a feasible endpoint for such designs, because screening is intended to detect precancerous lesions and subsequently direct the patient to early treatment before cancer develops. Evaluations of screening program effectiveness usually require a randomized controlled trial design to test whether a reduction of mortality or morbidity is achieved among a randomly selected group that is invited to attend for screening (whether or not all individuals in the screening-assigned group comply with the invitation) compared with a group that is not invited for screening. Assessment of screening program effectiveness is often referred to as an intent-to-treat analysis.[2] Screening program efficacy may also be evaluated in a randomized controlled trial, but is not limited to such a study design.[2,3]

The present study was designed to assess the overall performance and validity of cervicography screening for cervical cancer. Randomized controlled trials of effectiveness or efficacy for any screening method for cervical cancer have never been conducted because cytology screening became widely accepted before such trials were conducted. It would therefore be viewed as highly unethical to randomize women into

unscreened study arms. However, observational studies of efficacy are possible, and are often carried out using a case-control design.[4,5] Such studies assess whether women with and without cervical cancer differ on whether or not they have been screened within the detectable phase of precancerous lesions.[5-7]

In the present study initial findings showed that cervicography had low sensitivity but reasonably adequate specificity. High specificity is essential in screening for cervical cancer because referral of a large proportion of women with insignificant lesions would overburden colposcopy services, have an emotional impact on many women who do not have serious neoplasia, and result in excessive cost for colposcopy services. Reviews and arbitration of cervigrams led to a 5 point increase in the sensitivity of cervicography, indicating that even under optimal evaluation circumstances, the potential of cervicography for general screening, when high sensitivity is required, is limited. Adjustment of the cutpoint for referral for colposcopic examination did not change this finding. The current recommendation of referring women with any positive cervigram provides the best result for cervicography. Figure 1 displays a receiver operating characteristic (ROC) curve that shows the tradeoff between sensitivity and specificity of the enrollment cervigram results, the conventional Papanicolaou smear, and the combined screening regimen for the detection of high grade squamous intraepithelial lesions or cancer using different cutpoints for colposcopy referral for each screening protocol. Findings from the enrollment phase of the study[8] indicated that reducing the referral threshold to include women with atypical or positive cervigram results improves the

sensitivity to 62% for the detection of high grade squamous intraepithelial lesions or cancer, but reduces the specificity to an unacceptable level of 85% (d, Figure 1). An additional cutpoint for colposcopic referral was established that included only cervigrams classified as positive 0, positive 2, and positive 3. This threshold resulted in a drastic drop in sensitivity, to 22% (a, Figure 1), indicating the importance of positive 1 cervigrams for detecting high grade squamous intraepithelial lesions.

Stratification by age and menopausal status indicate that cervicography alone should not be recommended for women ages 50 and older and/or postmenopausal women. The sensitivity in older and postmenopausal women is unacceptably low, but the sensitivity in younger and premenopausal women was only slightly higher than it was in the overall sample (since women younger than age 50 comprised 76.7% of the overall sample they contribute more to the crude sensitivity). Some visual characteristics of the cervigram also appeared to be associated with sensitivity. However, further examination indicated that several of these factors were also related to age (i.e., metaplasia, cervicovaginal atrophy). The only other visual characteristics of the cervigram that were significantly associated with sensitivity were the quality and degree of the acetic acid effect. It is possible, therefore, that more careful application of acetic acid and/or reapplication of acetic acid may improve the performance of cervicography.

An additional limitation of cervicography is its fair reproducibility. Our comparison of cervigram classification by the evaluator during the enrollment phase of the study and the subsequent evaluator during the review phase showed that only

moderate agreement between the two evaluators was achieved beyond that expected by chance alone. National Testing Laboratories, Worldwide (NTL), the company that holds rights to the cervicography technique, has recognized this limitation. Since the initiation of this study, NTL has begun double evaluation of all cervigrams with adjudication of discrepancies. Nevertheless, our results indicated that multiple testing improved the overall performance of cervicography only slightly.

As a point of comparison, we examined the performance of conventional cytologic screening in the same population. Contrary to published results in other studies,[9] cytologic screening performed quite well in our study, achieving a sensitivity of 77% and a specificity of 94%. Compared with cervicography, the performance of cytologic screening was not hampered by age. A direct comparison of cervicography and cytology during the enrollment phase indicated that significantly more women with high grade squamous intraepithelial lesions or cancer were detected by cytology and not by cervicography than were identified by cervicography and not by cytology (Figure 2).[8] This analysis was repeated using the revised cervigram result and achieved similar results (data not shown).

In additional analyses, we assessed the performance of a combined screening protocol in which women would be referred for colposcopy if they had either a positive cervigram *or* a cytology result of atypical squamous cells of undetermined significance or more severe. As expected, the combined screening regimen improved the overall sensitivity, but at the expense of referring approximately 12% of all women for

colposcopic examination. The addition of multiple tests to the screening protocol would always increase the sensitivity because a greater proportion of women would be referred for colposcopy. However, the referral of so many women without serious neoplasia would be imprudent and would overburden existing colposcopy services. The receiver operating characteristic curves in Figures 1 and 3 illustrate this point for the enrollment (Figure 1) and revised (Figure 3) cervigram results, as compared to the conventional Papanicolaou smear, and the combined screening regimen for the detection of high grade squamous intraepithelial lesions or cancer. It is apparent in these figures that broadening the referral threshold of cervicography to include women with atypical cervigrams would reduce the specificity to an unacceptably low level without achieving a high enough sensitivity to be comparable to conventional cytology. Figure 3 shows that a combined screening regimen that would allow for referral of women with either a positive cervigram or a conventional cytology result of atypical squamous cells of undetermined significance or more severe would yield a desirable sensitivity of 91.2%, but the specificity of 89.1% might be unacceptably low for mass screening. If such a screening protocol were in effect, 1090 referrals to colposcopy would occur for every 10,000 women with either normal cervices or low grade neoplasia. A combined screening regimen could therefore overburden colposcopy services at high financial and emotional costs to the screened women.

The current charge by NTL for cervicography services is $45 (excluding additional office charges for performing the procedure). However, NTL has established a

mechanism by which patients may request discounted charges if they are unable to pay the full amount. Additionally, NTL now has a policy of placing cerviscope camera equipment in medical offices free of charge to promote the use of cervicography. Using the regular fee scale of $45 for cervicography and a published median fee of $428 for colposcopy and biopsy, the financial costs of referring women with true positive cervigram results may be compared with the costs of referring women with false positive cervigram results. In our study, 8460 women were screened with cervicography, which would yield a total of $380,700 in procedure costs. There were 549 women referred for colposcopy by cervicography. Of these 74 had true positive cervigram results for the detection of high grade squamous intraepithelial lesions or cancer, therefore resulting in $3,330 in cervicography charges and $31,672 in colposcopy/biopsy charges corresponding to women with true positive cervigrams. Among the women referred, 475 had false positive results (i.e., did not have true high grade squamous intraepithelial lesions or cancer), yielding $21,375 in cervicography charges and $203,300 in colposcopy/biopsy charges corresponding to women with false positive cervigrams.

Similarly, the estimated median reimbursement cost for the Pap smear is $22,[10] yielding $186,120 in procedure costs for processing of Papanicolaou smears for 8460 women. In our study, 586 women were referred for colposcopy by cytology. Of these, 105 women had true positive results, resulting in $2,310 in cytology costs and $44,940 in colposcopy/biopsy costs. Among the women referred, 481 had false positive results, yielding $10,582 in cytology costs and $205,868 in colposcopy biopsy costs. Therefore,

women with true positive cervigram results would accrue $35,000 in procedure costs, excluding additional clinic fees, and women with false positive cervigram results would accrue $224,675 in procedure costs. Women with true positive cytology results would accrue $47,250 in procedure costs (while also achieving greater detection ability), and women with false positive cytology results would accrue $216,450 in procedure costs. Comparing the cost estimates for two methods, for $12,250, 31 additional women would be identified by cytology over cervicography. The few additional false positives identified by cervicography would cost $8,225 over the procedure charges associated with false positive results by cytology.

Based on the results of this research project, together with previously published studies, cervicography should not be recommended for mass screening for cervical cancer. Alone, the method does not achieve a high enough sensitivity to merit its use as an alternate method to the Papanicolaou smear; in combination with conventional cytologic screening, specificity is reduced to an unacceptably low level. A benefit of cervicography is that the technique provides a permanent, visual record of the cervix that may be useful for clinician recordkeeping and for patient education. Cervigrams may also be useful for the purpose of clinical training.

Though NTL has made provisions for reduced charges for cervicography for those who are not able to pay full price, it is not clear to what extent the company would be able to justify similar payment plans to less developed countries where most potential users are unable to pay the full price for the service. Since the cost of conventional cytologic

screening is lower per woman than cervicography, cost alone is insufficient to promote the use of cervicography screening. However, wide variation in the quality of cytologic laboratories is apparent. While the performance of conventional cytology is better than that of cervicography when well-controlled specimen collection procedures and laboratories are utilized, cytology suffers when poor quality laboratories are utilized. Perhaps resources should be directed at improving the quality cytologic laboratories, training of cytotechnologists, and/or improving coverage rather than at promoting less proven techniques. While cervicography does provide the advantage of not having to be dependent on quality control of numerous laboratories, the centralized processing with cervicography is not compelling if too many women with serious neoplasia remain undetected by cervicography. In resource-poor countries, visual inspection is also a feasible alternative which does not require any processing. In one recent study[11] a sensitivity of 76.7% was observed for the detection of high grade squamous intraepithelial lesions. However, this technique suffers from low specificity (64.1% in that study) Therefore, if visual screening were implemented for widespread screening, more women with serious neoplasia might be detected than would be identified by cervicography, but a greater proportion of women without serious lesions would also risk unnecessary referral and/or treatment.

In addition to the Papanicolaou smear, new cytologic techniques are showing a promising role in cervical screening, and HPV testing may prove useful for adjudication of borderline cytology results. What is certain is that cervical screening and adjudication

of low grade and atypical results will continue to be a complex task and will probably involve several different methods and testing algorithms that are based on age and other risk factors. Just as important as the screening method itself is a need for greater attention and resources to be directed to increasing access to and utilization of cervical screening through health promotion and outreach. In the long term, development of an HPV vaccine is underway which holds promise for control and eventual eradication of cervical cancer.

References

1. Gordis L. Epidemiology. Philadelphia: W.B. Saunders Company, 1996.
2. Miller AB. Fundamental issues in screening for cancer. In: Schottenfeld D, Fraumeni JF, editors. Cancer Epidemiology and Prevention. Second ed. New York: Oxford University Press, 1996:1433-1452.
3. Prorok PC. Evaluation of screening programs for the early detection of cancer. In: Cornell RG, editor. Statistical Methods for Cancer Studies. New York: Marcel Dekker, Inc., 1984:267-328.
4. Weiss NS. Application of the case-control method in the evaluation of screening. Epidemiologic Reviews 1994; 16:102-108.
5. Weiss NS. Case-control studies of the efficacy of screening tests designed to prevent the incidence of cancer. Am J Epidemiol 1999; 149:1-4.
6. Hernández-Avila M, Lazcano-Ponce EC, Alonso de Ruiz P, Romieu I. Evaluation of the cervical cancer screening programme in Mexico: A population-based case-control study. Int J Epidemiol 1998; 27:370-376.
7. Herrero R, Brinton L, Reeves WC, Brenes MM, de Britton RC, Gaitan E, et al. Screening for cervical cancer in Latin America: A case-control study. Int J Epidemiol 1992; 21:1050-1056.

8. Schneider DL, Herrero R, Bratti C, Greenberg MD, Hildesheim A, Sherman ME, et al. Cervicography screening for cervical cancer among 8,460 women in a high-risk population. Am J Obstet Gynecol 1999; 180:290-298.

9. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. Am J Epidemiol 1995; 141:680-689.

10. Roland PY, Naumann RW, Alvarez RD, Kilgore LC, Partridge EE. A decision analysis of practice patterns used in evaluating and treating abnormal Pap smears. Gynecol Oncol 1995; 59:75-80.

11. University of Zimbabwe/JHPIEGO Cervical Cancer Project. Visual inspection with acetic acid for cervical-cancer screening: Test qualities in a primary-care setting. Lancet 1999; 353:869-873.

Figure 1

ROC Curves for Detection of HSIL and Cancer

Cervicography, Pap Smear, Pap or Cervicography

Enrollment Cervigram Result vs. Enrollment Referent Diagnosis

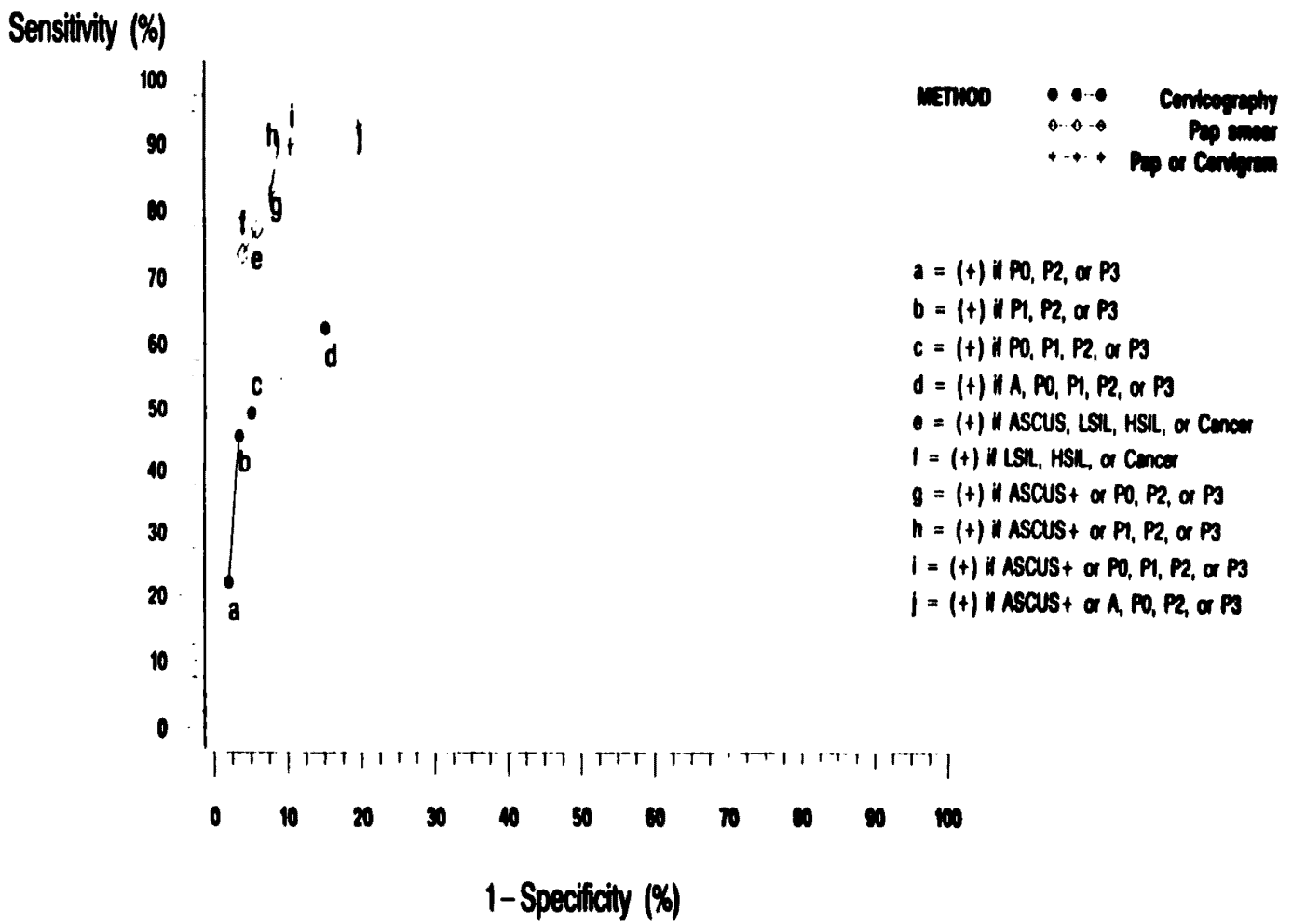


Figure 2
Category of Agreement on Colposcopic Referral by Referent Diagnosis

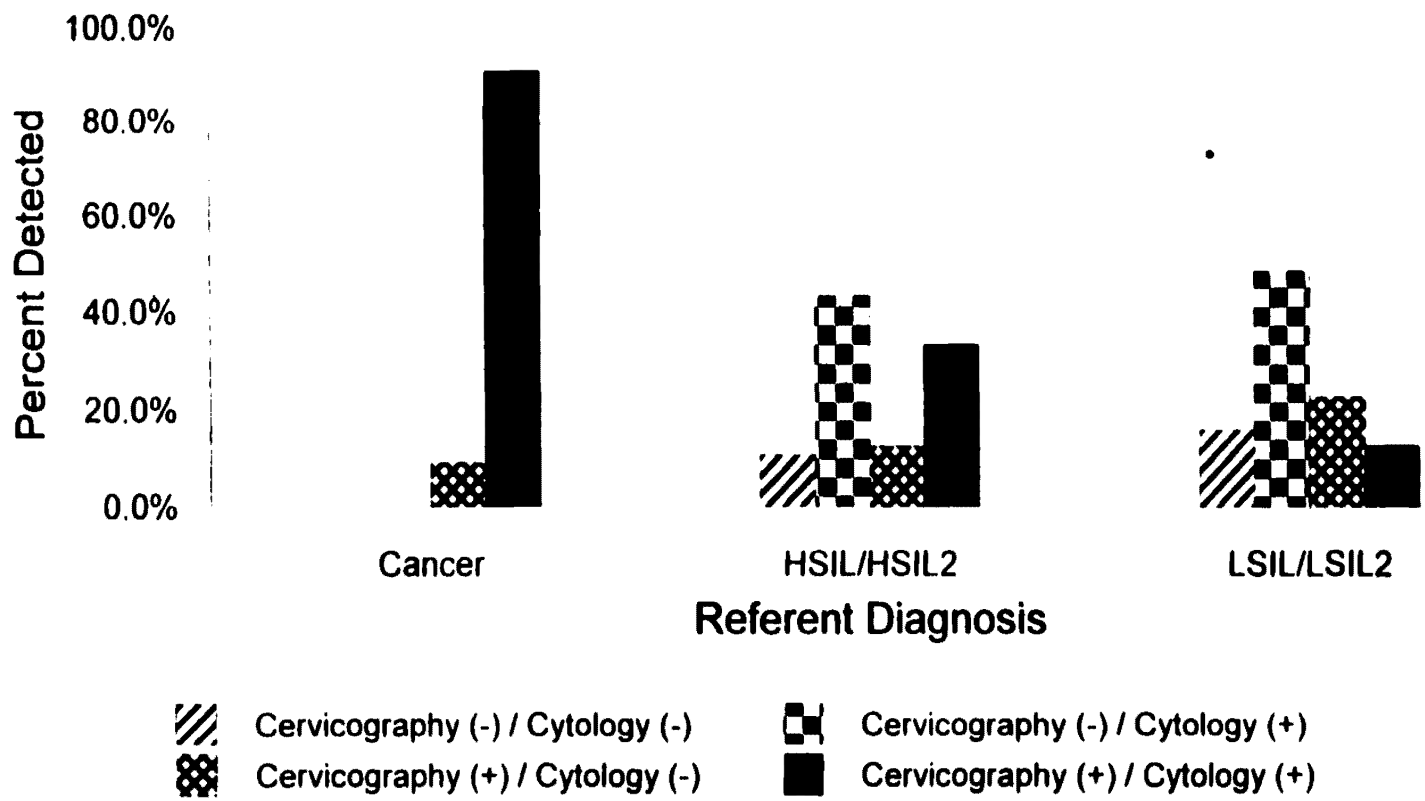


Figure 3

ROC Curves for Detection of HSIL and Cancer

Cervicography, Pap Smear, Pap or Cervicography

Revised Cervigram Result vs. Enrollment Referent Diagnosis