The Accuracy of the Papanicolaou Smear in the Screening and Diagnostic Settings

Marylou Cárdenas-Turanzas, MD, DrPH,[1] Michele Follen, MD, PhD,[2] Graciela M. Nogueras-Gonzalez, MPH,[1] J.L. Benedet, MD,[3] J. Robert Beck, MD,[4] and Scott B. Cantor, PhD[1]

[1]*Departments of Biostatistics and* [2]*Gynecologic Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, TX,* [3]*Division of Gynecologic Oncology, Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, Canada, and* [4]*Division of Population Science, Fox Chase Cancer Center, Philadelphia, PA*

■ **Abstract**

Objective. We evaluated the performance of the Papanicolaou smear in screening and diagnostic settings.

Study Design. We analyzed Papanicolaou smear results of 1,850 women recruited into a clinical trial to evaluate an emerging technology for the detection of cervical cancer. Screening and diagnosis groups were based on the history of previous Papanicolaou smear results. We calculated sensitivities, specificities, positive and negative likelihood ratios (LR+ and LR−), receiver operating characteristic curves, and areas under the receiver operating characteristic curve (AUC).

Results. In the screening group, by defining disease as cervical intraepithelial neoplasia (CIN) 2,3/cancer or worse and using high-grade squamous intraepithelial lesion (HSIL) as the test cutpoint, the AUC was 0.689, and the LR+ and LR− were 39.25 and 0.67, respectively. In the diagnosis group, the AUC was 0.764, and the LR+ and LR− were 3.79 and 0.56, respectively. By defining disease as human papillomavirus/CIN 1 or worse and HSIL as the test cutpoint, the AUC was 0.586, and the LR+ and LR− were 17.01 and 0.92 in the screening group; in the diagnosis group, the AUC was 0.686, and the LR+ and LR− were 2.77 and 0.75, respectively.

Conclusions. In a screening setting, a Papanicolaou smear result of HSIL or worse is 39 times more likely in a patient with CIN 2,3/cancer than in a patient without it. This compares to 4 times more likely in the diagnostic setting. The magnitude of the positive likelihood ratio observed in the screening group indicated that abnormal Papanicolaou smear results obtained in the screening setting should have more impact on clinical decision making than those from results obtained in the diagnostic setting. ■

Key Words: cervical intraepithelial neoplasia, cervix neoplasms, screening, diagnosis, sensitivity and specificity

Patients referred to a colposcopy clinic will often have a repeat Papanicolaou smear at the time of their colposcopy, even though they were referred because of an abnormal smear. During the diagnostic process, clinical decisions are made based on the outcomes of screening and detection tests. By the end of the diagnostic process, a patient may have undergone 1 or 2 Papanicolaou smears, human papillomavirus (HPV) testing, an endocervical curettage, as well as a colposcopic examination with directed biopsies.

Reprint requests to: Scott B. Cantor, PhD, Section of Health Services Research, Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Unit 447, 1515 Holcombe Blvd, Houston, TX 77030-4009. E-mail: sbcantor@mdanderson.org

This study was conducted by the Departments of Biostatistics and Gynecologic Oncology, The University of Texas M. D. Anderson Cancer Center; Department of Obstetrics and Gynaecology, University of British Columbia; and Fox Chase Cancer Center Division of Population Science. This study was conducted at sites in Houston, TX, and Vancouver, British Columbia, Canada.

Financial support for this study was provided by grant number CA82710 from the National Cancer Institute. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report.

Journal of Lower Genital Tract Disease, Volume 12, Number 4, 2008, 269–275

In most clinical practices, the results of the Papanicolaou smear will affect the decision to treat. If either the referral smear or the smear obtained at the colposcopy show high-grade squamous intraepithelial lesions (HSIL) or worse, the patient is often treated with a wire loop electrosurgical excision procedure.

This leads to questions regarding the accuracy of the Papanicolaou smear in the context of the setting in which the test is performed. One would expect a second or repeat test in the diagnostic setting (e.g., the colposcopy clinic) to build on the accuracy of the initial test in the screening setting (e.g., the general practice clinic). However, there is concern that a second Papanicolaou smear, when performed in the diagnostic setting, does not provide substantial additional diagnostic value. The purpose of this study is to describe and compare the accuracy of the Papanicolaou smear taken by study investigators and read by expert pathologists in the screening and diagnostic settings.

MATERIALS AND METHODS

This study is based on data collected as part of clinical trials to evaluate optical spectroscopy, an emerging technology for the screening and diagnosis of cervical squamous intraepithelial lesions. Nonpregnant women 18 years and older were recruited and allocated into either a screening group if they had no history of abnormal Papanicolaou smears and a diagnosis group if they had a previous abnormal Papanicolaou smear result. Details of the sociodemographic characteristics of these groups can be found in Pham et al. [1]. Approximately 82% of the women in the screening group were recruited from the community. The remaining women were recruited at a colposcopy clinic. The study was accomplished in 3 clinical locations: 2 comprehensive cancer centers in the United States and Canada and a community hospital in the United States. Details on the recruitment strategies have previously been described [2].

As part of the research protocol, the women were given several tests and clinical examinations associated with cervical cancer screening and detection. Each woman provided a complete medical history and was given a physical and pelvic examination. The pelvic examination included a conventional Papanicolaou smear (using a wooden spatula and a cervical brush), bacterial cultures to test for chlamydia and gonorrhea, viral specimens for HPV testing, and a colposcopic examination of the vulva, vagina, and cervix.

The conventional Papanicolaou smears were placed onto slides, fixed, and stained. Depending on the site of recruitment, the slides were processed at the Department of Pathology at either The University of Texas M. D. Anderson Cancer Center or the British Columbia Cancer Agency. During the colposcopic examination, 6% acetic acid was used, and the squamous columnar junction and the transformation zone were identified. Then the colposcopist took 1 or 2 colposcopically directed biopsies of the area with the worst overall colposcopic impression. The colposcopist also took 1 or 2 biopsies of squamous and columnar epithelium from an area of normal appearance. If the overall colposcopic impression was normal, biopsies were obtained from 1 or 2 normal sites and included both types of cervical epitheliums.

Each biopsy sample and conventional Papanicolaou smear was evaluated twice. The first evaluation was done by a gynecological pathologist at the study site, and a second blinded evaluation was done by one of the pathologists participating in the study. In cases of disagreement between the 2 evaluations, the sample was read a third time by another study pathologist to determine the final diagnosis. The process to evaluate pathological samples, as well as the range of agreement among study sites, has been detailed elsewhere [2, 3].

Cytological diagnosis was categorized according to the Bethesda classification (National Cancer Institute 2001) [4] as normal (including infection and reactive repair), atypical squamous cells of undetermined significance (ASCUS), low-grade squamous intraepithelial lesions (LSIL), HSIL, or cancer. For consistency, histological diagnosis was initially reported using the World Health Organization classification but then recategorized as normal, LSIL (HPV, cervical intraepithelial neoplasia [CIN] 1), HSIL (CIN 2, CIN 3, CIS, adenocarcinoma in situ), or cancer (invasive squamous or adenocarcinoma). In this article, we will refer to the World Health Organization histological classification.

All participants were informed of the purposes of the study and were provided written consent. The study was approved by the Institutional Review Boards at The University of Texas M. D. Anderson Cancer Center, The University of Texas Health Science Center at Houston, the Lyndon Baines Johnson Hospital Health District, the British Columbia Cancer Agency, and the University of British Columbia. Follow-up and treatment were available to all participants.

Study Sample and Statistical Analysis

For this study, 1,000 women were recruited into the screening group and 850 into the diagnosis group. Figure 1 presents a flowchart of study participants. Of

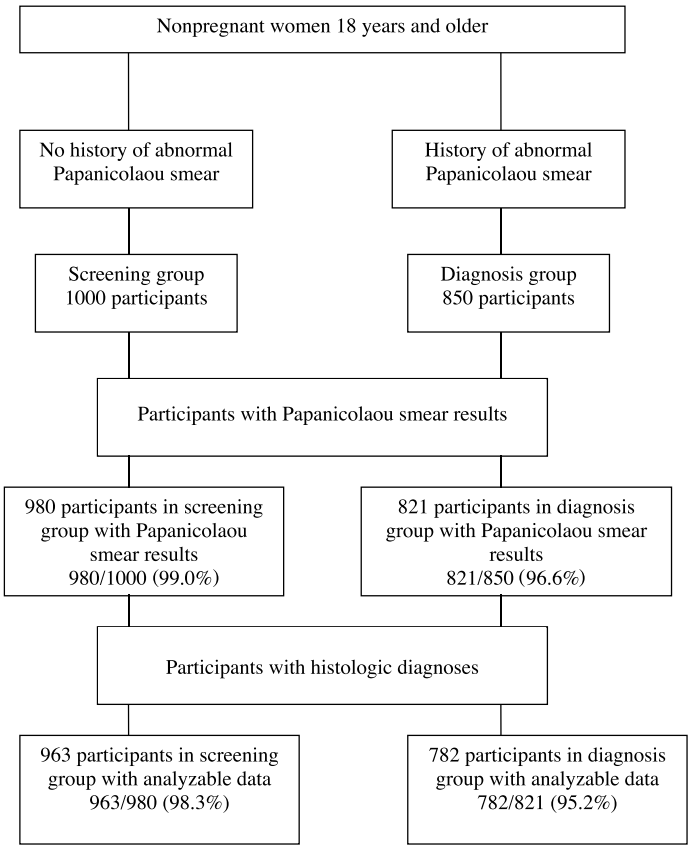

Figure 1. Flow diagram for patient recruitment and data analysis.

all women recruited, 1,745 were entered into the final analysis, 963 from the screening group and 782 from the diagnosis group. The primary reasons for excluding a woman from the final analysis were if she refused a colposcopy or biopsy, or if her Papanicolaou smear or biopsy results were missing. The percentages of missing results were higher in the diagnosis group compared with the screening group (8.7% vs 3.8%, respectively, $p < .001$). The unit of analysis was the participant, and the gold standard of diagnosis was defined as the worst histological result among all biopsies obtained from the same participant. We used HPV/CIN 1 and CIN 2,3/cancer as the thresholds of disease to reflect the different standards of practice for clinical decision making.

We compared the sociodemographic characteristics (age, race, education, marital status, and employment), clinical characteristics (menopausal status and gravidity), and the risk factors (smoking and HPV infection) of the women between both study groups. All variables except age were categorical. Pearson χ^2 test was used to determine differences in sociodemographic characteristics, clinical characteristics, and risk factors by study groups. Student *t* test was used to determine differences in age by study groups.

We determined the sensitivity, specificity, and the respective 95% CIs for the Papanicolaou smear for each disease threshold (e.g., HPV/CIN 1 and CIN 2,3 or cancer) and for each test cutpoint in the study groups. We then computed the positive likelihood ratio (LR+) and negative likelihood ratio (LR−) for each cutpoint of each disease threshold. We constructed a receiver operating characteristic (ROC) curve for each disease threshold of each study group and compared the areas under the curves by using a χ^2 test with nonparametric approach as suggested by DeLong et al. [5].

We used SPSS 12.0 for Windows (SPSS, Inc, Chicago, IL) and Stata version 9 statistical software (StataCorp LP, College Station, TX) for the statistical analysis.

RESULTS

Age, race, marital status, employment, gravidity, and menopausal status were found to be significantly

different ($p < .001$) between the screening and diagnosis groups. A significantly higher percentage of women in the diagnosis group reported a current smoking habit when compared with the screening group (22.8% vs 11.8%, respectively; $p < .001$).

The proportion of women with a positive HPV test (high-risk, low-risk, or both viral types) was significantly greater in the diagnosis group than in the screening group (51.0% vs 11.5%; $p < .001$). Table 1 details the characteristics of the study groups.

Table 1. Demographic and Clinical Characteristics of Participants

Characteristic	Screening group, N (%)	Diagnosis group, N (%)	*p*
Total	963	782	
Site of recruitment			<.001
Vancouver	144 (14.9)	317 (40.5)	
Houston	819 (85.1)	465 (59.5)	
Age (mean), y	43.95	36.42	<.001
Race			<.001
Non-Hispanic white	470 (48.8)	498 (63.7)	
African American	151 (15.7)	84 (10.7)	
Hispanic	267 (27.7)	101 (12.9)	
Asian	63 (6.5)	63 (8.1)	
Native American	3 (0.3)	7 (0.9)	
Other	9 (0.9)	29 (3.7)	
Education			.06
Less than high school	67 (7.0)	77 (9.8)	
High school or equivalent	162 (16.8)	135 (17.3)	
Some college	385 (40.0)	290 (37.1)	
Bachelor	216 (22.4)	180 (23.0)	
Some graduate education	22 (2.3)	30 (3.8)	
Graduate degree	110 (11.4)	69 (8.8)	
Unknown/Refused	1 (0.1)	1 (0.1)	
Marital status			<.001
Married	534 (55.5)	302 (38.6)	
Never married	189 (19.6)	227 (29.0)	
Living in a married-like situation	40 (4.2)	75 (9.6)	
Divorced or separated	175 (18.2)	162 (20.7)	
Widowed	25 (2.6)	14 (1.8)	
Unknown/Refused	0	2 (0.3)	
Employment			<.001
Employed (full-time or part-time)	660 (68.5)	532 (68.0)	
Unemployed	69 (7.2)	57 (7.3)	
Retired	82 (8.5)	29 (3.7)	
Housewife or student	149 (15.5)	160 (20.5)	
Unknown/Refused	3 (0.3)	4 (0.5)	
Number of pregnancies			<.001
None	201 (20.9)	239 (30.6)	
1–3	537 (55.8)	400 (51.2)	
>3	225 (23.4)	143 (18.3)	
Menopause			<.001
No	551 (57.2)	653 (83.5)	
Yes, I have gone through it	285 (29.6)	106 (13.6)	
Yes, I am now going through it	126 (13.1)	23 (2.9)	
Unknown	1 (0.1)	0	
Current smoker			<.001
Yes	97 (11.1)	178 (22.8)	
No	779 (80.9)	604 (77.2)	
Unknown/Refused	87 (9.0)	0	
HPV diagnosis (Hybrid Capture)			<.001
Test positive			
High-risk types	82 (8.5)	314 (40.2)	
Low-risk types	15 (1.6)	24 (3.1)	
Both types	13 (1.3)	59 (7.5)	
Test negative	847 (88.0)	381 (48.7)	
Unknown	6 (0.6)	4 (0.5)	

Performance of the Papanicolaou Smear in the Screening Group

In the screening group, the prevalence of CIN 2,3 or cancer was 2.2%, and the prevalence of HPV/CIN 1 or worse was 13.9%. When the disease threshold was defined as CIN 2,3 or cancer and the cutpoint for a positive test was ASCUS or worse, the sensitivity of the Papanicolaou smear was 0.429 [95% CI = 0.218–0.660] and the specificity was 0.928 [95% CI = 0.909–0.944]. When we restricted the definition of a positive test to LSIL or worse, the sensitivity remained at 0.429 [95% CI = 0.218–0.660] because there were no women with a diagnosis of CIN 2,3 or cancer and a Papanicolaou smear result of ASCUS. With our restricted definition of a positive test, the specificity increased slightly to 0.942 [95% CI = 0.925–0.956]. Using the same disease threshold but increasing the cutpoint for a positive test to HSIL or worse, we found that the sensitivity decreased to 0.333 [95% CI = 0.146–0.570] and the specificity increased to 0.992 [95% CI = 0.983–0.996]. Details on the frequencies used to calculate sensitivity and specificity are shown in Table 2. Still using a disease definition of CIN 2,3 or cancer, we then modified the cutpoint for a positive test to ASCUS or worse, LSIL or worse, and HSIL or cancer. The positive likelihood ratios for these levels were 5.94, 7.34, and 39.25, respectively. The corresponding negative likelihood ratios were 0.62, 0.61, and 0.67, respectively.

When we defined disease as HPV/CIN 1 or worse and an abnormal test as ASCUS or worse, the sensitivity and specificity were 0.224 [95% CI = 0.156–0.304] and 0.943 [95% CI = 0.925–0.958], respectively. When we expanded the definition of an abnormal test to LSIL or worse, the sensitivity of the Papanicolaou smear was 0.202 [95% CI = 0.137–0.280] and the specificity was 0.955 [95% CI = 0.939–0.968]. When we considered a result of HSIL or cancer to be an abnormal test, we obtained a sensitivity of 0.082 [95% CI= 0.042–0.142] and a specificity of 0.995 [95% CI = 0.988–0.999] (Table 2). The positive likelihood ratio for this disease threshold was 3.95 for a positive test of ASCUS or worse, 4.51 for an abnormal test of LSIL or worse, and 17.01 for an abnormal test of HSIL or cancer. The corresponding negative likelihood ratios were 0.82, 0.84, and 0.92, respectively.

Performance of the Papanicolaou Smear in the Diagnosis Group

In the diagnosis group, the prevalence of CIN 2,3 or cancer was 29.2%, and the prevalence of HPV/CIN 1 or worse was 54.1%. Defining disease as CIN 2,3 or cancer and considering an abnormal test to be a Papanicolaou smear result of ASCUS or worse, we found a sensitivity of 0.781 [95% CI = 0.721–0.833] and a specificity of 0.695 [95% CI = 0.655–0.733]. When we defined an abnormal test as LSIL or worse, the sensitivity was 0.746 [95% CI = 0.684–0.801] and the specificity was 0.724 [95% CI = 0.685–0.761]. Defining an abnormal test as HSIL or cancer, we found a sensitivity of 0.513 [95% CI = 0.446–0.580] and a specificity of 0.865 [95% CI = 0.833–0.892]. Details on the frequencies used to calculate sensitivity and specificity are shown in Table 3. The positive likelihood ratios for the Papanicolaou smear thresholds were 2.56 for ASCUS or worse, 2.70 for LSIL or worse, and 3.79 for HSIL or cancer. The corresponding negative likelihood ratios were 0.32, 0.35, and 0.56, respectively.

When disease was defined as histology HPV/CIN 1 or worse as the worst biopsy result and an abnormal test was defined as cytology ASCUS or worse, the sensitivity and specificity were 0.608 [95% CI = 0.560–0.654] and 0.749 [95% CI = 0.701–0.793], respectively. Defining an abnormal test as LSIL or worse, we obtained a sensitivity of 0.575 [95% CI = 0.526–0.622] and a specificity of 0.777 [95% CI = 0.731–0.819]. Defining an abnormal test result as HSIL or cancer, we found a sensitivity of 0.348 [95% CI = 0.302–0.395] and a specificity of 0.875 [95% CI = 0.836–0.907]. With a definition of disease as the indication of HPV/CIN 1 or worse in the worst biopsy result, the positive likelihood ratios were 2.42 for ASCUS, 2.58 for LSIL, and 2.77 for HSIL. The corresponding negative likelihood ratios were 0.52, 0.55, and 0.75, respectively. Again, details on the frequencies can be found in Table 3.

Table 2. Frequency of Papanicolaou Smear Results and Histologic Diagnoses of Patients in the Screening Group

Papanicolaou Smear diagnosis	Histologic diagnosis, N (%)			
	Normal	HPV/CIN 1	CIN 2,3/Cancer	Total
Negative	782 (94.3)	92 (81.4)	12 (57.1)	886
ASCUS	10 (1.2)	3 (2.7)	0 (0.0)	13
LSIL	33 (4.0)	14 (12.4)	2 (9.5)	49
HSIL, cancer	4 (0.5)	4 (3.5)	7 (33.3)	15
Total	829	113	21	963

Percentages may not add to exactly 100.0% because of rounding.
HPV, human papillomavirus; CIN, cervical intraepithelial neoplasia; ASCUS, atypical squamous cells of undetermined significance; LSIL, low-grade squamous intraepithelial lesions; HSIL, high-grade squamous intraepithelial lesions.

Table 3. Frequency of Papanicolaou Smear Results and Histologic Diagnoses of Patients in the Diagnostic Group

Papanicolaou Smear diagnosis	Histologic diagnosis, N (%)			
	Normal	HPV/CIN 1	CIN 2,3/Cancer	Total
Negative	269 (74.9)	116 (59.5)	50 (21.9)	435
ASCUS	10 (2.8)	6 (3.1)	8 (3.5)	24
LSIL	35 (9.7)	43 (22.1)	53 (23.2)	131
HSIL, cancer	45 (12.5)	30 (15.4)	117 (51.3)	192
Total	359	195	228	782

Percentages may not add to exactly 100.0% because of rounding.
HPV, human papillomavirus; CIN, cervical intraepithelial neoplasia; ASCUS, atypical squamous cells of undetermined significance; LSIL, low-grade squamous intraepithelial lesions; HSIL, high-grade squamous intraepithelial lesions.

ROC Curve Analysis

We constructed ROC curves for both study groups at 2 thresholds of disease: CIN 2,3 or worse and HPV/CIN 1 or worse. We found that when the disease was defined as the indication of CIN 2,3 or cancer in the worst biopsy result, the area under the ROC curve was 0.689 for the screening group [95% CI = 0.575–0.803] and 0.764 for the diagnosis group [95% CI = 0.729–0.799] (Figure 2A). No significant difference ($p = .21$) was found when the 2 ROC curves were compared. With a definition of disease as the indication of HPV/CIN 1 or worse in the worst biopsy result, the area under the ROC curve was 0.586 for the screening group [95% CI = 0.549–0.622] and 0.686 for the diagnosis group [95% CI = 0.652–0.719] (Figure 2B). The areas under these 2 ROC curves were found significantly different ($p < .001$).

COMMENT

The importance of high sensitivity and specificity cannot be overstated when validating screening and diagnostic tests. In the case of cervical cancer, given the prevalence of disease, which is low in the screening setting and moderate in the diagnostic setting, and considering the sensitivity and specificity of the Papanicolaou smear, one can calculate the positive predictive value and negative predictive value for the test to determine the posttest probability of disease. Clinicians wish for this revised probability of disease to be above a particular threshold before taking further action. This threshold can be determined informally or through formal analytic methods [6]. In the screening group, the positive likelihood ratios ranged from 4.0 to 39.3; however, in the diagnosis group, the positive likelihood ratios ranged

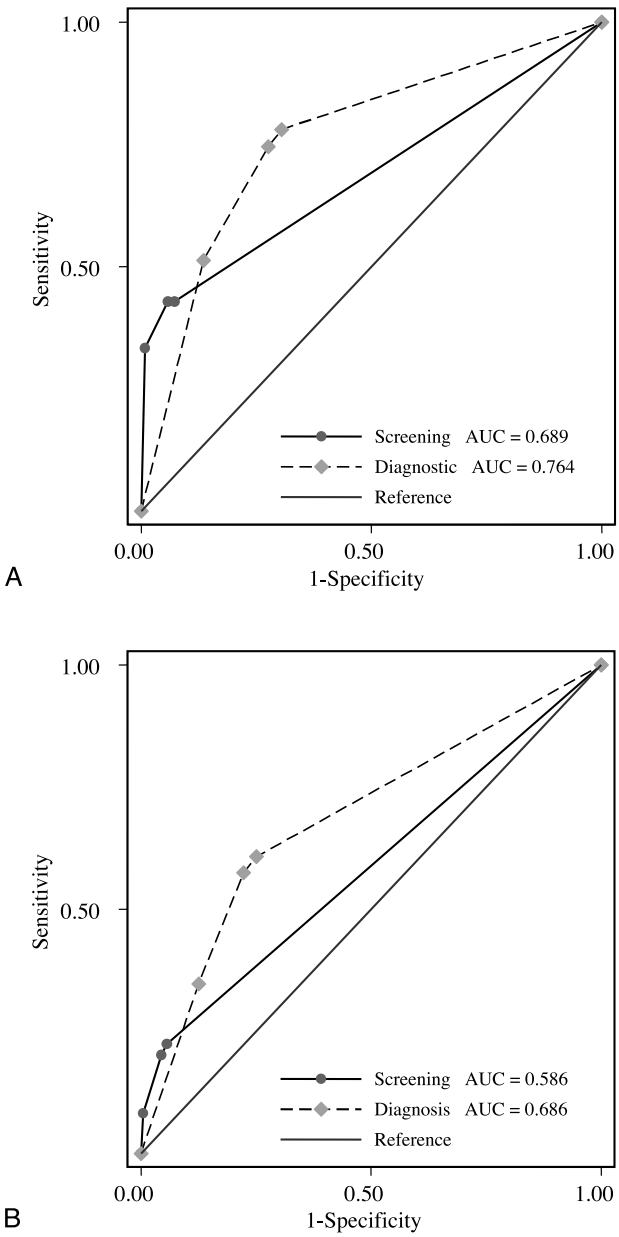

Figure 2. A, Comparison of receiver operating characteristic (ROC) curves for screening and diagnosis groups with definition of disease as CIN 2 or worse. B, Comparison of receiver operating characteristic (ROC) curves for screening and diagnosis groups with definition of disease as HPV/CIN 1 or worse.

from 2.4 to 3.7. None of the positive likelihood ratios in the diagnosis group were better than the recommended "rule of thumb" cutpoint of 5.0 [6]. The negative likelihood ratios were slightly better in the diagnostic setting but were still unsupportive of the Papanicolaou smear test as one of reasonable discrimination in the diagnosis setting.

In a previous meta-analysis of the performance of the Papanicolaou smear that defined disease as HPV/CIN 1 or worse, Nanda et al. [7] reported overall sensitivity ranges from 30% to 87% and overall specificity ranges from 86% to 100%. The systematic review resulted in 12 studies being considered the best evidence available for screening populations. However, the authors recognized that only a few studies had overcome verification bias by obtaining a biopsy or colposcopic examination for at least a fraction of the participants with negative results. In our study, colposcopically directed biopsies were obtained from all of the women, which may explain why our results differ. We also observed that the Papanicolaou smear displayed higher specificity than sensitivity. Similar to the report of Nanda et al. [7], we observed that the specificity tended to improve when the threshold of disease changed from HPV/CIN 1 or worse to CIN 2,3/cancer.

Studies evaluating the benefit of repeating the Papanicolaou smear during the colposcopic examination typically have not performed biopsies on sites of normal colposcopic appearance [8, 9], thus leading to verification bias [10]. If biopsies are only performed on patients with lesions detected by the colposcopist, a false increment in the sensitivity and a false decrease in the specificity occur when evaluating the performance of the repeated Papanicolaou smear [11]. To overcome this verification bias, Simsir et al. [8] used a broader definition of disease, which included results of the follow-up tests such as Papanicolaou smear and colposcopically directed biopsies. The researchers compared the sensitivity and specificity of the repeated smear, biopsy, and the combination of the repeated smear and biopsy, but did not provide the resulting likelihood ratios, ROC analysis, or area under the curve. To evaluate the accuracy of the Papanicolaou smear in 2 settings, we tried to provide a more extensive analysis. Our study is distinctive in that we obtained biopsies from both colposcopically normal and abnormal sites.

The higher positive likelihood ratio of the Papanicolaou smear in the screening group indicated a higher probability of disease among women with positive tests, especially at the cutpoint of HSIL or cancer. Unlike a positive test result from a woman in the diagnosis group, a positive test result from a woman in the screening group indicated an increased probability of correctly diagnosing disease when using colposcopy. This occurred because the posttest odds of disease for the Papanicolaou smear become the pretest odds of disease for the subsequent test used [12].

In our study, both the screening and diagnosis groups had Papanicolaou smears collected and interpreted by

the same group of clinicians. This may not be the case in settings with documented variations regarding the quality of the test. Nevertheless, when quality assurance shows reasonable performance of the Papanicolaou smear in the screening setting, there is a potential for saving resources by avoiding the repetition of unnecessary tests.

Another limitation of our study was shown by the fact that there were no diagnoses of ASCUS that truly had CIN 2,3/cancer histology in the screening group. This may be because of the limited number of cases of CIN 2,3/cancer in the screening group as well as the limited number of pathologists who participated in the study and their particular training or approach to cytological interpretation.

The magnitude of the positive likelihood ratio observed in the screening group indicated that abnormal Papanicolaou smear results obtained in the screening setting should have more impact on clinical decision making than those from results obtained in the diagnostic setting.

Acknowledgment

The authors thank Rebecca Partida for editorial contributions, which enhanced the clarity of the manuscript.

REFERENCES

1. Pham B, Earle N, Rabel K, Follen M, Scheurer M. Maximizing the diversity of participants in a phase II clinical trial of optical technologies to detect cervical neoplasia. *Gynecol Oncol* 2007;107(S1):208–14.

2. Guillaud M, Benedet JL, Cantor SB, Staerkel G, Follen M, MacAulay C. DNA ploidy compared with human papilloma virus testing (Hybrid Capture II) and conventional cervical cytology as a primary screening test for cervical high-grade lesions and cancer in 1555 patients with biopsy confirmation. *Cancer* 2006;107:309–18.

3. Malpica A, Matisic JP, Niekirk DV, Crum CP, Staerkel GA, Yamal JM, et al. Kappa statistics to measure interrater and intrarater agreement for 1790 cervical biopsy specimens among twelve pathologists: qualitative histopathologic analysis and methodologic issues. *Gynecol Oncol* 2005;99:S38–52.

4. Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA* 2002;287:2114–9.

5. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–45.

6. Deeks JJ. Systematic reviews in health care: systematic reviews of evaluations of diagnostic and screening tests. *BMJ* 2001;323:157–62.

7. Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med* 2000; 132:810–9.

8. Simsir A, Ioffe OB, Bourquin P, Brooks SE, Henry M. Repeat cervical cytology at the time of colposcopy. Is there an added benefit? *Acta Cytol* 2001;45:23–7.

9. Zardawi IM, Rode JW. Clinical value of repeat Pap smear at the time of colposcopy. *Acta Cytol* 2002;46:495–8.

10. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med* 1987;6:411–23.

11. Project UoZJCC. Visual inspection with acetic acid for cervical-cancer screening: test qualities in a primary-care setting. *Lancet* 1999;353:869–73.

12. Grimes DA, Schulz KF. Refining clinical diagnosis with likelihood ratios. *Lancet* 2005;365:1500–5.