Papnet-assisted, primary screening of cervico-vaginal smears

M. A. Duggan, *M.D., F.R.C.P.C.*

Dept. of Pathology and Laboratory Medicine, University of Calgary and Cytopathology Division, Calgary Laboratory Services, Calgary, Alberta (Canada)

Summary

Purpose: The Papnet system was initially designed for rescreening negative Pap tests but may also be an effective primary screener.

Methods: A set of 2,200 archival slides diagnosed by conventional, manual screening as 2,000 (90%) WNL, 47 (2.1%) carcinomas, 50 (2.3%), HSIL, 50 (2.3%) LSIL, and 53 (2.4%) ASCUS/AGUS were compared to the results of Papnet-assisted, primary screening. Following Papnet scanning, the digitized images were triaged and classified as abnormal or negative. All abnormals had a full manual screening, whereas negatives had a limited screening. Results by each screening method were compared and discordant cases were peer reviewed for a consensus result. Screening efficacy by each method was measured against a standard result composed of the concordant and consensus results.

Results: There were 101 concordant and 181 discordant abnormal results. The standard result for the slide set was 1,953 (88.9%) WNL, 87 (3.9%) ASCUS/AGUS, 52 (2.4%) LSIL, 62 (2.8%) HSIL, 39 (1.8%) carcinomas, and 5 (0.2%) unsatisfactory. Papnet versus manual sensitivity rates were 87.6% vs 72.3% at the ASCUS/AGUS threshold, 85.6% vs 82.4% at the LSIL threshold, and 89.1% vs 90.1% at the HSIL thereshold.

Conclusions: Papnet-assisted, primary screening equals conventional, manual screening in the detection of a wide range of cell abnormalities and is more effective in the detection of abnormalities at the lower end of the abnormal spectrum.

Key words: Pap test; Primary screening; Papnet system; Efficacy.

Introduction

Cervical premalignancy (SIL; squamous intraepithelial lesion) is a disorder of cell growth and maturation which may progress over time from low to high grade SIL and subsequently to invasive carcinoma [1]. The Papnicolaou stained, cervico-vaginal smear (Pap test) is offered to all eligible women in order to detect the cellular changes of cervical premalignancy. Laboratory examination of the Pap test requires visual examination of the estimated 50,000-300,000 epithelial cells on a glass slide by a cytotechnologist. This labor intensive methodology has remained virtually unchanged since the test's introduction, and is subject to a degree of error [2]. Laboratory error with respect to calling a test normal when an abnormality exists is labeled a false negative result. The converse is labeled a false positive result. The false negative rate of cervical cancer screening averages 20%, although it may range from 5-50% [2]. The variability of the false negative rate, and the inability of conventional, manual screening to eliminate all false negatives highlights a need to change the laboratory methodology of Pap test evaluation.

In recent years, computerized semi-automated Pap test screening devices have been developed for the quality control screening of negative Pap tests [3]. The Papnet system (Neuromedical Systems Inc., Upper Saddle River, New Jersey, U.S.A.) is one such device. The combination of manual screening and Papnet rescreening of all negative tests can reduce the false negative rate of a SIL result by 50% [4]. The system is composed of a fully automated scanner and an interactive review station. The scanner searches the conventionally prepared Pap test and selects by a neural network computer 128 epithelial cells. The scanned images are digitized and transferred to a compact disc (CD). The review station consists of a monitor, computer, keyboard, and mouse interfacing with a digital drive to replay the scanned images. The cytotechnologist interprets these video based images according to standard criteria of cellular atypia and decides if the Pap test should be categorized as review or negative. The glass slides of all tests triaged as review have a full, manual screen by a cytotechnologist.

The Papnet system may function as well as and in some instances better than conventional, manual screening in the primary evaluation of the Pap test [5-11]. In this application, the Pap tests are first scanned, the images are triaged, and only the review tests undergo full, manual screening. The increased sensitivity appears to be mostly at the ASCUS (atypical squamous cells of undetermined significance) level raising concerns about the clinical significance and specificity of automated screening [12]. The conclusions of published studies however, must be carefully interpreted and are not directly comparable because of limitations and inconsistencies in study design, case selection criteria, reporting terminology, and data analysis. Other more critical limitations and inconsistencies are definitional in regard to error classification and the gold standard against which the results of automated and manual screening were measured and compa-

Revised manuscript accepted for publication October 18, 1999

red. Consequently, the effectiveness of the Papnet system as a primary screener is in need of more formal and controlled investigation.

Material and Methods

A retrospective, systematic search of the files of the Cytopathology Division of Calgary Laboratory Services at the Foothills Hospital was conducted to select 2,200 Pap tests. The slide set was to include 2,000 negative slides (WNL, within normal limits or BBC, benign cellular change) and 200 abnormals. The abnormals were to consist of approximately 50 slides in each of the categories: 1) ASCUS/AGUS (atypical glandular cells of undetermined significance), 2) LSIL (low grade squamous intraepithelial lesion), 3) HSIL (high grade squamous intraepithelial lesion), and 4) carcinoma (including squamous cell and adenocarcinoma).

All slides had been screened by one of five cytotechnologists, those targeted for quality control were rescreened by one of two cytotechnologists, and all abnormals were interpreted by one of four pathologists. The starting date for the slide selection was December 1995. The Bethesda system (TBS) reporting terminology was adopted by the laboratory in April of 1994 [13]. Before that time, cervical abnormalities were reported in a classification system that included the CIN (cervical intraepithelial neoplasia) terminology [1, 14]. Results in the CIN terminology were translated to the TBS by a cytotechnologist not otherwise involved in the study. Every abnormal slide in each of the four abnormal categories was selected until approximately 50 in each category were retrieved. For each abnormal slide, the subsequent 10 consecutive negative slides were selected. Excluded from selection were slides that were: 1) cracked or damaged, 2) repaired, 3) had more than half of the surface covered with bubbles from improper cover slipping, 4) multiples of the same accession number, 5) cover-slipped with plastic, 6) ink marked with a water insoluble pen, and 7) cover-slipped so that it extended more than 50 mm from the slide edge opposite to the laboratories identification (ID) label.

Prior to the slides being cleaned in preparation for scanning, the original ink markings were photocopied to maintain a permanent record of the conventional, manual screening. The slides were bar-coded with a unique Papnet ID (identification) number and cross referenced with the laboratories accession number. Scanning with software version 2.10 was performed at the scanning centre in Suffern, New York according to standard procedures [15]. On completion of scanning, the slides and CDs were returned to the Foothills Hospital. The Papnet ID number was entered into the study information form which was designed to collect patient demographic data and pertinent clinical information. This information was transcribed from the original report by the cytotechnologist not otherwise involved in the study. The Papnet 2.00 review station was used by one of two Papnet certified cytotechnologists to study the images on the CD. All unsuccessfully scanned slides were assigned a technical code by the scanner and all had a full, manual screen. Image features triggering a full, manual screen were documented on the form and the slide triaged as review (Fig. 1). The same triage cytotechnologist did the full, manual screen and that interpretation was also captured on the study form. All slides with a result other than WNL or BCC were forwarded to the study pathologist (MAD) who interpreted the abnormal features and recorded the result. The triage cytotechnologists and study pathologist were masked to the results of the conventional, manual screening.

Slides triaged as negative had an abbreviated, manual screen by the triage cytotechnologist. All four slide/cover-slip edges were screened using a 10x objective followed by a vertical screening sweep of one end of the slide from top to bottom. When the bottom was reached, the slide was screened horizontally for 10 mm, screening was stopped and a vertical screening sweep made. At the top of the slide, 10 mm of horizontal screening occurred and continued to the bottom of the slide. This pattern continued to the end of the slide. Cell abnormalities detected in this manner were viewed at 40x. Slides with results other than WNL or BCC were referred to the study pathologist.

Any discrepancy in diagnosis between the conventional, manual screening and the Papnet-assisted, primary screening was resolved by a consensus panel review of those slides. The panel consisted of the study pathologist and one of two additional experienced pathologists, and took place approximately four months after the study pathologist's interpretations of the Papnet screenings. Consensus was defined as full, mutual agreement. Tests diagnosed as negative by Papnet-assisted, primary screening had the original ink markings restored to the slides by one of the triage cytotechnologists prior to the consensus review. The review panel was masked to the results of both screening methods.

The effectiveness of detecting abnormal slides by conventional, manual screening and by Papnet-assisted, primary screening was compared by calculating the sensitivity and specificity rates and 95% confidence intervals of each method [16]. The panel's interpretation of all discrepancies between the two methods was used as the correct result and combined with the result of the concordant cases to form a true result. Results were ranked in ascending order from ASCUS/AGUS to carcinoma. Rates for each screening method were calculated using the true

Figure 1. — *Triage protocol list of features indicating a review decision and a full, manual screen.*

1. Pontentially abnormal cells are displayed
   Perinuclear Halo
   Binucleation
   Elongated (Fennel-shaped) Nucleus
   Overlapping Nuclei
   Overlapping Cells in Group
   Koilocytotic Cell
   Increased N/C Ratio and/or Nuclear size ≥ 2x size of normal intermediate cell nucleus
   Molding Nuclei
   Eosinophilic, Opaque Cytoplasm
   Irregular Nuclear Membranes
   Prominent Nucleoli
   Hyperchromatic Nucleus
2. Tumor diathesis or suspicious background is present.
3. Microorganisms are present.
4. Excessive inflammation, cytolysis, blood or artifacts are present in 12 or more tiles.
5. Hormonal pattern is inconsistent with age or clinical presentation.
6. Endocervical component is absent.
7. Endometrial cells present are morphologically aberrant and/or inconsistent with menstrual cycle.
8. Technical code is displayed.

result as the reference standard. Slides in which a diagnosis was not possible because the specimen was inadequate were excluded from the calculations. Rates were separately calculated for all abnormal results (>/= ASCUS/AGUS), results of LSIL and higher (>/= LSIL), and HSIL and higher (>/=HSIL). Tests with a negative result by conventional manual or Papnet assisted, primary screening, but with an abnormal standard result were labeled false negatives. The converse were labeled false positives. Undercalled and overcalled results were included as false negative and positive results respectively, when calculating the rates at the SIL thresholds.

False negative errors in conventional, manual screening were attributed to screening if the standard result was based on new abnormal cells not previously detected and to interpretation if no new cells were detected, but the same cells were reclassified. All false positives were attributed to interpretative errors as the change in result was due to a reclassification of the identified cells. Errors in Papnet-assisted, primary screening were similarly classified and the false negative rate determined from the sensitivity results. False negative errors by this screening method were divided into negative triage and review triage errors. Possible error sources in these two activities were investigated and attributed to 1) misclassification by the study pathologist if a result was overturned by the consensus review, 2) post triage, full manual screening errors if the cytotechnologist failed to detect or correctly interpret the abnormal cells previously detected by conventional, manual screening, 3) review station, triage failure if on the second-look of the images by the author and one of the triage cytotechnologists, criteria for review were present and not detected by the original, triage cytotechnologist, and 4) scanning failure if the images did not display an abnormal feature or technical code. The attributable, false negative rate for each component of the error source was then calculated.

Results

By conventional, manual screening, there were 2,000 (90.9%) WNL slides and 200 (9.1%) abnormal slides. The abnormals were composed of 53 (2.4%) ASCUS/AGUS, 50 (2.3%) LSIL, 50 (2.3%) HSIL, and 47 (2.1%) carcinomas. There were 27 (1.2%) ASCUS and 26 (1.2%) AGUS. The carcinomas consisted of 27 (1.2%) squamous and 20 (0.9%) adenocarcinomas. The review station triage identified 1,409 (64%) slides for a full, manual screen and 791 (36%) for an abbreviated screen. Slides triaged for review included 36 (2.6%) with a technical code which were mostly due to a cornflake artefact (n=30, 83.3%). There were 196 (13.9%) and 28 (3.5%) abnormal results respectively, following the full and abbreviated, manual screenings. Therefore, Papnet-assisted, primary screening yielded 224 (10.2%) abnormal slides, 1,971 (89.6%) WNL slides, and five (0.2%) where a diagnosis was not possible as the specimen was unsatisfactory. The abnormals consisted of 70 (3.2%) ASCUS, 19 (0.9%) AGUS, 42 (1.9%) LSIL, 57 (2.6%) HSIL, and 36 (1.6%) carcinomas. The carcinomas were classified as 19 (0.9%) squamous cell carcinomas, 14 (0.6%) adenocarcinomas, and three (0.1%) undifferentiated carcinomas.

In a comparison of Papnet-assisted, primary screening with conventional, manual screening, there were 277 (12.6%) abnormal results by at least one screening method, and five (0.2%) unsatisfactory for evaluation results which were excluded from further analysis. The abnormal results were concordant in 101 (36.5%) and discordant in 176 (63.5%) tests (Table 1). The discordant results included 53 (30.1%) false negative which had the original, ink markings restored to the slides prior to the consensus review. Following the review, a standard result was generated for the 2,195 slides and was composed of 1,953 (88.9%) WNL, 73 (3.3%) ASCUS, 16 (0.7%), AGUS, 52 (2.4%) LSIL, 62 (2.8%) HSIL, and 39 (1.8%) carcinomas. The carcinomas consisted of 22 (1.0%) squamous cell carcinomas, 15 (0.7%) adenocarcinomas, and two (0.09%) undifferented carcinomas.

Conventional, manual screening compared to the true result had a lower rate of abnormal results due mainly to decreased numbers of ASCUS/AGUS, but also to reduc-

Table 1. — *Papnet-assisted, primary screening compared to conventional, manual screening.*

	CONVENTIONAL, MANUAL SCREENING					
Papnet screening	Negative	ASCUS/AGUS	LSIL	HSIL	CARCINOMA	TOTAL
NEGATIVE	1,918 (97.3%)	36 (1.8%)	15 (0.8%)	1 (0.05%)	1 (0.05%)	1,971 (89.8%)
ASCUS/AGUS	63 (70.8%)	11 (12.4%)	9 (10.1%)	2 (2.2%)	4 (4.5%)	89 (4.1%)
LSIL	10 (23.8%)	3 (7.1%)	20 (47.6%)	8 (19.1%)	1 (2.4%)	42 (1.9%)
HSIL	3 (5.3%)	1 (1.7%)	6 (10.5%)	38 (66.7%)	9 (15.8%)	57 (2.6%)
CARCINOMA	1 (2.8%)	2 (5.5%)	0	1 (2.8%)	32 (88.9%)	36 (1.6%)
TOTAL	1,995 (90.9%)	53 (2.4%)	50 (2.3%)	50 (2.3%)	47 (2.1%)	2,195

tions in the LSIL and HSIL categories (Table 2). There were 175 (1,928) slides with a true positive (negative) result of >/= ASCUS/AGUS, 126 (2,021) with a >/= LSIL result, and 91(2,088) with a >/= HSIL result. The respective false negative (positive) results for each of the three diagnostic thresholds were 67 (25), 27 (21), and 10 (6). Papnet-assisted, primary screening compared to the true result detected fewer abnormal results and the decrease was mainly in the LSIL category (Table 3). There were 212 (1,941) tests with a true positive (negative) result of >/= ASCUS/AGUS, 131 (2,308) with a result of >/= LSIL, and 90 (2,091) with an HSIL or greater result. The respective false negative (positive) results at each threshold were 30 (12), 22 (4), and 11 (3).

The sensitivity and specificity of both methods as well as the 95% confidence intervals were calculated (Table 4). The narrow range of 95% confidence intervals for each method at each of the three diagnostic thresholds indicated sample size was not a confounding variable. In a comparison of the degree of overlap of the 95% confidence intervals, Papnet-assisted, primary screening was significantly more sensitive than conventional, manual screening at the ASCUS/AGUS threshold. Differences in sensitivity at the LSIL and HSIL thresholds and differences in specificity at each threshold were not significant.

The false negative results by conventional, manual screening were all due to a screening error, whereas interpretative errors by the professionals who formulated the original results accounted for the 25 false positive tests. The false negative rate of Papnet-assisted, primary screening at the ASCUS/AGUS threshold was 12.4%.

Table 2. — *Conventional, manual screening compared to the reference standard result.*

	STANDARD RESULT					
Manual Screening	Negative	ASCUS/AGUS	LSIL	HSIL	CARCINOMA	TOTAL
NEGATIVE	1,928 (96.6%)	44 (2.2%)	21 (1.1%)	2 (0.1%)	0	1,995 (90.9%)
ASCUS/AGUS	18 (33.9%)	31 (58.5%)	2 (3.8%)	1 (1.9%)	1 (1.9%)	53 (2.3%)
LSIL	7 (14.0%)	9 (18.0%)	28 (56.0%)	6 (12.0%)	0	50 (2.4%)
HSIL	0	2 (4.0%)	0	48 (96.0%)	0	50 (2.3%)
CARCINOMA	0	3 (6.4%)	1 (2.1%)	5 (10.8%)	38 (80.9%)	47 (2.1%)
TOTAL	1,953 (88.9%)	89 (4.1%)	52 (2.4%)	62 (2.8%)	39 (1.8%)	2,195

Table 3. — *Papnet-assisted, primary screening compared to the reference standard result.*

	STANDARD RESULT					
Manual Screening	Negative	ASCUS/AGUS	LSIL	HSIL	CARCINOMA	TOTAL
NEGATIVE	1,941 (98.5%)	24 (1.2%)	5 (0.25%)	1 (0.05%)	0	1,971 (89.8%)
ASCUS/AGUS	12 (13.5%)	61 (68.5%)	14 (15.7%)	0	2 (2.3%)	89 (4.1%)
LSIL	0	1 (2.4%)	33 (78.6%)	8 (19.0%)	0	42 (1.9%)
HSIL	0	1 (1.8%)	0	52 (91.2%)	4 (7.0%)	57 (2.6%)
CARCINOMA	0	2 (5.6%)	0	1 (2.8%)	33 (91.6%)	36 (1.6%)
TOTAL	1,953 (88.9%)	89 (4.1%)	52 (2.4%)	62 (2.8%)	39 (1.8%)	2,195

Potential sources of error were investigated and 17 (56.7%) were tests that had been triaged as review and had a full, manual screen and 13, (43,3%) as negative and had an abbreviated screen (Table 5). Half (50%) of the false negative rate was due to a post triage, full manual screening error, 30% to a triage error, 13.3% to a scanning error, and 6.5% to a classification error. All four error sources accounted for misses in the ASCUS category and all except scanning error, resulted in AGUS misses. LSIL misses were attributed to post triage, full manual screening and triage errors and the one HSIL miss to a post triage, full manual screening error. Classification error by the study pathologist accounted for the false positive results by Papnet-assisted, primary screening.

Discussion

Automated screening must be at least as effective and preferably more effective than conventional, manual screening before changes to the laboratory evaluation of the Pap test can be considered. Fully automated, primary screening of Pap tests is desiderable, but not yet possible. Meanwhile semi-automated devices in use as postscreeners are being adapted to a primary screening role. The Autopap system (Neopath, Inc., Redmond, Washington, USA) is currently available as a primary screening device and demonstrates a superior sensitivity to manual screening in the detection of all levels of cervical anormality [17]. The Papnet system shows some promise as a primary screener, but efficacy is unproven because of design limitations and inconsistencies in the published

Table 4. — *Sensitivity and specificity of conventional, manual screening and Papnet-assisted, primary screening.*

THRESHOLD	SENSITIVITY		SPECIFICITY	
	Manual	Papnet	Manual	Papnet
>/= ASCUS/AGUS	72.3%	87.6%	98.7%	99.4%
95% C.I.	66.7-77.9	83.4-91.8	98.2-99.2	99.1-99.7
>/= LSIL	82.4%	85.6%	99.0%	99.8%
95% C.I.	76.4-88.4	80.0-91.2	98.6-99.4	99.6-100.0
>/= HSIL	90.1%	89.1%	99.7%	99.9%
95% C.I.	84.3-95.9	83.0-95.2	99.5-99.9	99.8-100.0

Table 5. — *Papnet-assisted, primary screening of 30 false negative results: error sources.*

Error source	Triage = review	Triage = negative	Total	FNR
Scanner		4 (30.8%)	4 (13.3%)	1.6%
Result: ASCUS		2 (15.4%)		
LSIL		2 (15.4%)		
Triage		9 (69.2%)	9 (30.0%)	3.7%
Result: ASCUS		7 (53.8%)		
AGUS		1 (7.7%)		
LSIL		1 (7.7%)		
Manual screening	15 (88.2%)		15 (50%)	6.2%
Result: ASCUS	11 (64.7%)			
AGUS	1 (5.9%)			
LSIL	2 (11.8%)			
HSIL	1 (5.9%)			
Classification	2 (11.8%)		2 (6.7%)	0.9%
Result: ASCUS	1 (5.9%)			
AGUS	1 (5.9%)			
Total	17 (56.7%)	13 (43.3%)	30	12.4%

FNR = False negative rate.

studies [5-10]. A rewiew of some of these studies [6, 7, 9] and coded so as to achieve comparability of the studies designs suggested the system may achieve an 18-40% increase in the number of abnormal results relative to manual screning [18]. The PRISMATIC trial complied with the proposed guidelines for the evaluation of primary screening instruments and reported a sensitivity of Papnet-assisted, primary screening equal to manual screening and a significantly higher specificity [11, 19].

The accurancy of cytopathology results are traditionally measured against the biopsy result [20]. In Pap test screening however, such correlation is not always possible as biopsy follow-up may never occur or there may be a long interval between the abnormal test result and the biopsy. To accommodate such limitations, consensus review of the Pap test result by a panel of experienced pathologists is the accepted method of generating a reference standard, in particular when primary screening instruments are being compared [19, 20]. The goal of the panel review is to achieve a consensus in terms of the number and types of all abnormal results possible by each method and to use this result as the reference standard against which the performance of both screening methods can be measured and compared. With this approach, specificity is defined by the panel's opinion.

A retrospective design was chosen for the current study to control the number and types of abnormal test results seeded into the study set. Other advantages included smaller study numbers and faster accrual. Observer bias associated with a testing scenario that did not mirror usual laboratory practice was a disadvantage of the design, however. All components of the Papnet assisted, primary screening with the possible exception of scanning were subject to some observer bias. These components were carried out by an experienced group of two cytotechnologists and one pathologist which would result in more consistency in the review station triage, manual screening, and classification of abnormalities. Conventional, manual screening in contrast reflected the practice of the study period as performed by a larger number of laboratory personnel with varying levels of experience. The study group's prior positive experience with the system as a pontential primary screener may also have introduced some bias [8].

A triage review rate of 15-30% is usual when the Papnet system is used to rescreen negative Pap tests [8]. The rate should be higher for the device in an assisted, primary screening mode since the frequency of abnormal tests is higher. The scanning and triage components were successful in sorting the tests into those that were negative (36%) and those that were a review (64%), negating concerns that all tests would be a review status because the triage was governed by a very specific protocol. This protocol has both qualitative and quantitative criteria regarding cell and smear features on a video image that the cytotechnologists evaluate, and was developed to correspond with a manual screening abnormal threshold of >/=BCC [8]. Thus this triage review rate was due not only to a higher rate of abnormal tests, but also to the comprehensiveness and customization of the protocol items. This rate is a benchmark, triage review rate for a disease prevalence of 9% and cannot be compared with other Papnet-assisted, primary screening studies either because the rate was not included, details of the protocol were not provided, or the prevalence of abnormal tests was different [5-7, 9-11].

Scanning was unsuccessful in 36 tests due to technical difficulties. Most were due to a cornflake artefact that interfered with the scanner's ability to focus the cells. While technical difficulties increased the number of slides requiring a full, manual screen, the number was inconsenquential relative to the 1,373 requiring the same because of an abnormarl image or images. The rate of 1.6% is nearly similar to that previously reported from the same laboratory and suggests this is the technical failure, benchmark rate for the laboratory [8]. In routine practice however, the rate may be higher as the exclusion criteria attempted to eliminate some slides would result in a technical difficulty for the scanner.

Papnet-assisted, primary screening had a higher frequency of abnormal results when directly compared to conventional, manual screening (10.2% vs 9.1%) and an insignificant number of unsatisfactory results (n=5). Lack of complete agreement between the results of both screening methods was due to screening errors and/or errors in classification. The number and types of abnormal tests in the study set were standardized based on the 101 (36.4%) inter-method concordant and the consensual review of the 176 (64.0%) discordant results. Importantly, the reference standard included examples of all seeded abnormalities including AGUS as well as three types of carcinoma, ie, squamous cell, adenocarcinoma, and undifferentiated carcinoma. This contrasts with the prospective study of Ashfaq et al. which did not have any examples of AGUS in the study period [5]. Some observer bias may have influenced the reference standard result because the study pathologist who would have memory of the Papnet-assisted, primary screening results was a panel member. This possibility was diminished by the study control measures. They included masking the panel members to the results of both methods and delaying the review for several months so that the memory of the Papnet results would have faded. Moreover, memory would not have been an issue for the 53 false negatives by Papnet-assisted screening as they would not have been referred to the study pathologist for interpretation.

Papnet-assisted, primary screening and conventional, manual screening compared to the reference standard detected all seeded types of cellular abnormalities. Both methods were very sensitive, but Papnet-assisted, primary screening was significantly more sensitive than conventional, manual screening in the detection of any abnormality (87.6% vs 72.3%), slightly more sensitive in detecting abnormalities at the LSIL threshold (85.6% vs 82.4%), and marginally less sensitive in the detection of abnormalities at the HSIL threshold 89.1% vs 90.1%). The specificity results of both screening methods were also very high. Papnet-assisted, primary screening was more specific at each threshold. Undercalled and overcalled tests at the SIL thresholds were included with the

absolute false positive and negative results in the calculation of their effectiveness rates. Justification for this inclusive definition of a false positive and negative result was based on the clinical impact incorrect classification would have on the management of the woman's disease [14]. Overcalled tests may be incorrectly referred for colposcopy and undercalled tests may be inappropriately followed with repeat Pap tests.

Abbreviated, manual screening was used to confirm the negative triage and 13 false negative results were discovered. All were either an ASCUS/AGUS or LSIL and importantly did not include an HSIL or carcinoma. This technique is used in some laboratories as a quality control check of negative tests and although controversial, may be as effective as the standard practice of a full, manual rescreening of a proportion of negative tests in detecting false negatives [21]. Abbreviated screening is not routinely used in this laboratory so that inexperience with the technique may have underestimated the number of false negatives. Consequently, the effectiveness of Papnet-assisted, primary screening may be overstated. Alternatively, since the 5.3% attributable, false negative rate is close to the 5% false negative fraction of standard practice, the abbreviated screening may have been performed successfully [22]. More importantly, since the proposed guidelines on automated instruments endorse a 5% false negative rate for the negative triage, it would appear that the Papnet system functions appropriately in this regard [19].

The 15.3% sensitivity advantage of Papnet-assisted, primary screening approximated the lower end of the 18-40% range determined from a review of other studies [18]. At this time, the increased detection of LSIL lesions is considered advantageous, but whether the same applies to the increased detection of ASCUS/AGUS tests is debatable. Tests classified as ASCUS/AGUS are by definition of uncertain clinical significance and in follow-up studies show a high spontaneous resolution rate [13, 23, 24]. These properties may undermine the clinical relevance of the sensitivity advantage of automated screening. Alternatively, it is prudent to consider that a proportion of women with an ASCUS/AGUS result will have a biopsy confirmed SIL at the subsequent colposcopy exam [23, 24]. Thus the increased sensitivity at the ASCUS/AGUS threshold should not be discounted.

The sensitivity of Papnet-assisted, primary screening may be enhanced by improvements in cell classification, full manual screening, review station triage, and scanning. Further refinement of cytologic criteria to reduce or eliminate classification errors could reduce the number of false positives and reduce the false negatives by 6.7%. The false negative misclassifications related to one test with an ASCUS and one with an AGUS result, entities which are subject to considerable inter and intra-observer variation [25]. The sensitivity improvement therefore, would be marginal and difficult to achieve. More effective post triage, full manual screening may improve the sensitivity as 50% of the false negatives were due to errors in this activity, and importantly did include the single missed test result of HSIL. This may not be possible because the attributable, false negative rate (6.2%) of the post triage, full manual screening approximated what is described as the irreducible false negative fraction (5%) of manual screening [22]. While quality control of the post triage, full manual screening to detect manual screening misses may not be necessary, future studies should carefully address this need.

The sensitivity may be improved even further by reducing or eliminating the 30% attributable error due to review station triage failures. The nine triage misses were mostly examples of ASCUS and there was only one LSIL. A quality control check of the triage process may result in some improvement. Additionally, the manufacturer should consider alarming the triage process so that cytotechnologists are alerted to potential, false negative triages. The scanner accounted for 13.3% of the false negative errors and likely represents the limits of the system's detection capacity in it's current design. Additional enhancements may improve the scanner's sensitivity and wuold be advantageous as 50% of the lesions missed were examples of LSIL.

Some laboratory error in the conventional, manual screening practice was expected, however, the sensitivity and specificity rates were more than acceptable, particularly at the SIL threshold [2]. The relatively lower sensitivity rate at the ASCUS/AGUS threshold may not be solely due to screening and interpretative errors but rather to some inconsistent application of cytologic criteria by the laboratory personnel consequent to the introduction and lack of experience with the TBS reporting terminology [13]. Diagnosis of SIL lesions would be unaffected by a change in terminology since the cytologic criteria for diagnosis did not differ between reporting systems [1, 13, 14]. The almost equal effectiveness of both screening systems in the detection of SIL results supports this conclusion. Regardless, it is important to note that the effectiveness of Papnet-assisted, primary screening in this study was not due to a substandard level of conventional, manual screening. Thus Pap test screening in this laboratory could be improved at the ASCUS/AGUS and LSIL thresholds by the implementation of Papnet-assisted, primary screening. Such decisions are contingent on the results of future appropriately designed, prospective trials and whether the Papnet system will become commercially available once the companies economic issues are resolved [26].

Conclusions

Papnet-assisted, primary screening equaled conventional, manual screening in the detection of a wide range of cell abnormalities and was more effective in the detection of abnormalities at the lower end of the disease spectrum.

Acknowledgements

The following are thanked for their contribution to the study: Robert Gordon R. T. and Marianne Andriuk R. T. for the cytotechnology, and Paul Kneafsey M.B. F.R.C.P.C. and Jill Wanklyn M.D. F.R.C.P.C. for the panel review.

References

[1] Richart J. M.: "Cervical intraepighelial neoplasia: A review. In: Pathology Annual, Somer S., ed., East Norwalk, Connecticut, U.S.A., Appleton-Century-Crofts, 1973, 301.

[2] Koss L. G.: "The Papanicolaou test for cervical cancer detection. A triumph and a tragedy". *JAMA*, 1989, *261*, 737.

[3] Grohs H. K., Husain O. A. N.: "Automated cervical cancer screening". New York, U.S.A. Igaku-Shoin M. C., 1994.

[4] Mango L. J., Philip V. T.: "Neural Network - assisted analysis and microscopic rescreening in presumed negative cervical cytologic smears". *Acta Cytol.*, 1998, *42*, 227.

[5] Ashfaq R., Salinger F., Solares B., Thomas S., Liu G., Liang Y., Saboorian H. G.: "Evaluation of the PAPNET system for prescreening triage of cervicovaginal smears". *Acta Cytol.*, 1997, *41*, 1058.

[6] Boon M. E., Kok L. P., Beck S.: "Histologic validation of neural network-assisted cervical screening: Comparison with the conventional procedure". *Cell. Vision.*, 1995, *2*, 23.

[7] Doornewaard H., Van de Seijp H., Woudt J. M. C., van der Graaf Y., Van den Tweel J. G.: "Negative cervical smears before CIN 3/Carcinoma - Reevaluation with the PAPNET testing system". *Acta Cytol.*, 1997, *41*, 74.

[8] Duggan M. A., Brasher P.: "A paired comparison of manual and automated Pap test screening using the PAPNET system". *Diagn. Cytopathol.*, 1997, *17*, 248.

[9] Kok M. R., Boon M. E.: "Consequences of neural network technology for cervical screening". *Cancer*, 1996, *78*, 111.

[10] Michelow P. M., Hlongwane N. F., Leiman G.: "Simulation of primary cervical cancer screening by the Papnet system in an unscreened, high risk community". *Acta Cytol.*, 1997, *41*, 88.

[11] PRISMATIC project management team: "Assessment of automated primary screening on PAPNET of cervical smears in the PRISMATIC trial". *Lancet*, 1999, *353*, 1381.

[12] Krieger P., Bibbo M.: "Ediotrial; Our journey towards improved accuracy in cytology: The role of new technologies". *Acta Cytol.*, 1997, *41*, 11.

[13] Kurman R. J., Solomon D.: "The Bethesda system for reporting cervico/vaginal cytologic diagnoses. Definitions, criteria, and explanatory notes for terminology and specimen adequacy". Kurman R. J. and Solomon D. Eds, Springer-Verlag Inc., New York, 1994.

[14] Task Force on Cervical Cancer Screening Programs: "Cervical cancer screening programs". *Can. Med. Assoc. J.*, 1976, *114*, 1003.

[15] Mango L.: "Neuromedical Systems, Inc.". *Acta Cytol.*, 1996, *1*, 53.

[16] Fleiss J. L.: "Statistical methods for rates and proportions". Second edition, New York: John Wiley & Sons, 1981.

[17] Wilbur D. C., Prey M. U., Miller W. M., Pawlick G. F., Colgan T. J.: "The autopap system for primary screening in cervical cytology: Comparing the results of a prospective, intended use study with routine manual practice". *Acta Cytol.*, 1998, *42*, 214.

[18] Mango L. J., Radensky P. W.: "Interactive neural network assisted screening. A clinical assessment". *Acta Cytol.*, 1998, *42*, 233.

[19] Intersociety Working Group for Cytology Technologies: "Proposed guidelines for primary screening instruments for gynecologic cytology". *Acta Cytol.*, 1997, *41*, 924.

[20] 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.

[21] Dudding N.: "Rapid rescreening of cervical smears: An improved method of quality control". *Cytopath.*, 1995, *6*, 95.

[22] Krieger P., Naryshkin S.: "Random rescreening of cytologic smears: A practical and effective component of quality assurance programs in both large and small cytology laboratories". *Acta Cytol.*, 1994, *38*, 291.

[23] Sidaway M. K., Tabarra S. O.: "Reactive changes and atypical squamous cells of undetermined significance in papanicolaou smears: A cytohistological correlation". *Diagn. Cytopathol.*, 1993, *9*, 423.

[24] Zweizig S., Noller K., Reale F., Collis S., Resseguie L.: "Neoplasia associated with atypical glandular cells of undetermined significance on cervical cytology". *Gynecol. Oncol.*, 1997, *65*, 314.

[25] Young N. A., Naryshkin S., Atkinson B. F., Ehya H., Gupta P. K., Kline TS, Luff R. D.: "Interobserver variability of cervical smears with squamous - cell abnormalities: A Philadelphia study". *Diagn. Cytopathol.*, 1994, *11*, 352.

[26] Kreuger F. A., Van Ballegooizen M., Doornewaard H.: "Is PAPNET suitable for primary screening". *Lancet*, 1999, *353*, 1374.

Address reprint requests to:
MÁIRE A. DUGGAN, M.D., F.R.C.P.C.
Dept. of Pathology, Foothills Hospital,
1403, 29th Street, N. W.
Calgary, T2N 2T9
Alberta (Canada)