

Location-Guided Screening of Liquid-Based Cervical Cytology Specimens

A Potential Improvement in Accuracy and Productivity Is Demonstrated in a Preclinical Feasibility Trial

David C. Wilbur, MD,* Elizabeth M. Parker, CT(ASCP), and Joseph A. Foti, SCT(ASCP)

Key Words: Cervical cytology; Papanicolaou smear; Automation; Computerization; AutoPap

Abstract

A 2-armed, masked study was performed on 1,275 AutoCyte PREP (TriPath, Burlington, NC) slides. Manual screening (current practice [CP]) was compared with automated screening with location-guided screening (LGS) using the AutoPap system with the SlideWizard 2 automated review microscopy station (TriPath). Cytologic adjudication determined “truth” for each slide. LGS identified more abnormal cases at all interpretive levels and classified abnormalities more specifically. For high-grade squamous intraepithelial lesions and above, the overall sensitivity of appropriate triage to pathologist review was 98.4% for LGS and 91.1% for CP. The appropriate triage for all abnormal cases was 92.1% for LGS and 87.9% for CP. The potential exists for more accurate and productive cytologic interpretation using this system, which requires no manual screening of a portion of slides and only limited review of another portion of “predotted” slides.

Numerous factors have come together in recent years to threaten the viability of the cervical cytology screening system for early detection and, hence, prevention of cervical cancer. These factors include inadequate reimbursement for this complex medical test, a declining cytotechnologist workforce in the face of potential increased utilization, and the continuing threat of litigation for “mistakes” in what is inherently an imperfect and highly subjective analysis. Some laboratories have ceased to provide cervical cytology services, and many more have contemplated this course but ultimately have drawn back owing to a sense of duty to patients and to continue to provide a comprehensive approach to overall patient care.

Interestingly, this “negative milieu” for the practice of the most successful screening program ever developed has been manifest in an era of unprecedented improvement in technique. These improvements have been in the areas of specimen preparation, namely the introduction of liquid-based methods, and computerized analysis of specimens, more specifically the use of computerized image analysis scanning devices. Taken individually, these methods each have been shown to improve overall accuracy of disease detection (sensitivity and specificity),¹⁻³ and each has the theoretical promise of substantial improvements in cytology laboratory productivity (John Bishop, unpublished data, 1996).^{3,4} Furthermore, combinations of these technologies might be expected to provide further improvements in both areas compared with the use of each method individually. A recent approval by the US Food and Drug Administration (FDA) for the combination of liquid-based cervical cytology (AutoCyte PREP, TriPath Imaging, Burlington, NC) with computerized automated screening (AutoPap Primary

Screening System, TriPath Imaging) showed that the 2 methods working together were statistically equivalent to the manual screening of PREP specimens alone. However, the data showed numeric superiority with the combination of techniques in terms of sensitivity and specificity of cervical disease detection.⁵ Although definitive time and motion studies have yet to be documented for liquid-based preparation methods, it generally is considered by practitioners expert in their use that liquid-based specimens may be screened up to 30% faster than routine conventionally prepared specimens. Preliminary studies support these observations (John Bishop, unpublished data, 1996).⁴ In addition, use of the AutoPap system permits immediate archiving of up to 25% of smears (either liquid-based or conventional) as “within normal limits” (WNL) without further manual review.³ Hence, both methods can be expected to provide some increase in laboratory productivity when used individually, but to an even greater extent when used together.

The FDA-approved application for the AutoPap Primary Screening System permits use of the device to classify entire conventional or liquid-based (AutoCyte PREP, TriPath Imaging) cervical cytology slides as to the likelihood that abnormality is (or is not) present. However, inherent in the software of the AutoPap system there always has been the ability to designate areas on the slide (fields of view [FOV]) that have the highest probability of containing abnormal cells (AutoPap location-guided screening [LGS] system). The advantages of an extension of the approval of the AutoPap to permit use of this capability are obvious. First, in terms of screening sensitivity, cytologists would have the advantage of being directed to areas on the slide at greatest risk of containing abnormal cells. This should reduce omission errors in the otherwise “needle in a haystack” hunt that occurs in the routine unguided examination of a slide. Second, if FOV showing the greatest level of abnormality are highlighted, improved specificity of interpretation could be achieved. Third, if cytologists are directed immediately to abnormalities, then the process of screening a slide reliably and, hence, appropriate triage to the pathologist could be hastened substantially, leading to greater productivity in the laboratory.

Use of this capability of the AutoPap system has been shown previously using the so-called PAPMAP concept. In this early process, a paper printout of the FOV of highest likelihood of abnormality provided a means to trace these areas onto the slide for subsequent manual examination. Use of this system in a variety of differing settings and protocols has shown improvements in accuracy and, where studied, in productivity as well.⁶⁻⁹ To further improve the process, the present trial expanded on these studies to include the addition of an automated microscopy station (SlideWizard 2, TriPath Imaging). This station adds automatic motorized

location of the FOV of interest, as downloaded directly from the AutoPap screening of the slide, for subsequent cytotechnologist review.

By using this combination of technologies, the routine manual practice of screening liquid-based cervical cytology slides was compared with AutoPap scanning, followed by examination of slides using field-of-view localization. Results were compared for the following: (1) appropriate case triage, both to full manual review and to pathologist review; (2) sensitivity and specificity of abnormality detection; and (3) evaluation of specimen adequacy in each arm. No attempts to directly assess or quantify screening productivity were made in this study, although potential productivity improvements were modeled.

Materials and Methods

The 2-armed trial used a retrospectively obtained population of previously reported cervical cytology slides derived from the routine cytology practice seeded with known abnormal slides from a variety of laboratory sources. All slides were prepared by the AutoCyte PREP liquid-based method according to the instructions for use¹⁰ and were stained by a routine Papanicolaou method. Slides were directed to reviewers in such a way that no slide was examined by the same cytotechnologist in both study arms (or by the same cytotechnologist who performed the original clinical assessment of the slide). Seeding of known abnormal slides was done by a random process, by personnel not otherwise involved in the study. All participants in the study were unaware of all previous results or interpretations on the study slides. All previous markings were removed from the slides before study initiation. Institutional review board approval was obtained for the study protocol before study initiation (Western Institutional Review Board, Olympia, WA). All study interpretations used the nomenclature of the 1991 Bethesda System,¹¹ as the study was initiated before the 2001 Bethesda System revisions.¹²

In the current practice (CP) arm, slides were screened by cytotechnologists in the routine manner, with 10% random quality control (QC) rescreening performed. Interpretations were made on each slide for cytologic abnormality and specimen adequacy, and these became the final CP arm results following completion of the entire process. All markings were recorded by using the electronic slide dotting capabilities of the SlideWizard 2 and removed from the slides before entrance into the second study arm.

In the LGS arm, slides were processed on the AutoPap Primary Screening System with LGS capabilities. All slides were bar-coded to ensure proper identification at each step in the screening process. Up to 25% of slides with the lowest

device scores were immediately designated as “no further review,” and final arm interpretations of WNL were recorded along with device-generated statements of adequacy. The remaining approximately 75% of slides were designated as “review,” requiring further analysis. AutoPap quintile rankings were generated on each slide indicating the device-determined likelihood of the presence of abnormality. In addition, 10 FOV designated by the AutoPap as having highest likelihood of containing abnormality were downloaded to SlideWizard 2 screening stations by a copy service application interface. A field of view is represented by a $\times 200$ magnification microscopic field that is centered on a point coordinate selected by the AutoPap. The reviewer sees a larger area in the microscopic field than actually is imaged and designated by the AutoPap device.

In the review population of slides, cytotechnologists examined only the FOV via a mouse- or foot pedal-activated motorized microscopy stage. Assessment of the accuracy of the translation of the locations of FOV was performed via a general stage calibration process (done before each run of slides) followed by visual comparison of cells in the first field of view from each slide with the actual image of that field downloaded from the AutoPap. For each slide, following review of all FOV, if no abnormality (cellular, pattern, background, adequacy, or no FOV generated owing to scanty cellular specimen) was identified, an LGS final arm interpretation of WNL was recorded for the slide along with a statement of adequacy. If a potential abnormality was identified in the FOV review, the slide was referred for a full manual screening. For this group of slides, the LGS arm interpretations for abnormality and adequacy were the result of this process. In slides designated as WNL following these reviews, the AutoPap selected the 15% highest scoring for “QC review,” and these slides received both FOV review and full manual rescreening. Following QC rescreening, LGS arm interpretations were finalized.

After completion of the 2 arms of the study, the final arm interpretation results were collated and compared. Further adjudication was performed by 1 cytopathologist (D.C.W.) on cases in which the cytologic interpretation differed between the 2 arms and for all cases in which a matching, but abnormal, interpretation had been entered in each of the study arms. In addition, all slides in which there was a discrepancy in terms of specimen adequacy were adjudicated by 1 senior cytotechnologist (J.A.F. or E.M.P.). For the purposes of this study, *adequacy* was defined as either satisfactory or unsatisfactory. For these abnormal and adequacy adjudication processes, slide markings were retained from the AP arm, and markings (if any) were added back to the slides based on the locations captured using the SlideWizard 2 following completion of the CP arm. All adjudicators were

unaware of all previous interpretations and did not know from which study arm the slide markings had been derived. The results of the adjudication process became the study “truth” interpretation for each slide. All nonadjudicated slides (WNL in both study arms) had a truth final designation of WNL.

After completion of both arms and the adjudication process, the results of each arm were compared with one another for each truth-determined Bethesda System category. Statistical analysis to determine equivalence and superiority between the procedures was done by the conditional binomial method.¹³ In addition, data were generated relating to the rate that slides were designated as one of the following: (1) process review, indicating that slides could not be scanned successfully by the AutoPap system; (2) no further review, indicating slides classified as WNL by the AutoPap Primary Screening System alone; or (3) WNL by FOV examination alone in the LGS arm. An assessment also was made of discrepancies between FOV review alone vs LGS arm interpretation and against study truth. In addition, data were generated regarding the false-negative and false-positive rates in each arm, and, hence, the sensitivity and specificity of the procedure in each study arm could be derived.

Results

Initially, 1,300 slides were entered into the protocol: 1,049 were slides derived in consecutive order from the routine clinical laboratory, and 251 were randomly seeded known abnormal slides. Twenty-five (25) cases were excluded from the analysis owing to processing failures on the AutoPap instrument (22 cases; 1.73% process review rate), because no AutoPap Review was performed (2 cases), and owing to an inappropriate reading (1 case; slide reviewed by the same cytotechnologist in more than one assessment). Therefore, the total number of slides entered into the analysis was 1,275. The final truth-determined slide designations are given in **Table 1**.

Table 1
Distribution of 1,275 Slides

Truth Interpretation	No. (%) of Slides
Unsatisfactory	12 (0.94)
Within normal limits	1,049 (82.27)
Atypical squamous cells of undetermined significance	52 (4.08)
Atypical glandular cells of undetermined significance	32 (2.51)
Low-grade squamous intraepithelial lesion	6 (0.47)
High-grade squamous intraepithelial lesion	32 (2.51)
Adenocarcinoma in situ	5 (0.39)
Carcinoma	87 (6.82)

Table 2
Location-Guided Screening Arm Review Rates for 1,275 Slides

Type of Review	Rate (%)
No further review (n = 218)	17.10
Fields of view only (n = 619)	48.55
Full manual review (n = 438)	34.35
With fields of view (n = 418)	32.78
Without fields of view (n = 20)	1.57

Table 3
Comparison of Location-Guided Screening and Current Practice Arms for Cases With Truth Determination of ASCUS+*

Location-Guided Screening	Current Practice (Manual Screening)		
	ASCUS+	Not ASCUS+	Total
ASCUS+	166	31	197
Not ASCUS+	17	0	17
Total	183	31	214

ASCUS, atypical squamous cells of undetermined significance.
* P = .0003 for equivalence; P = .0297 for superiority.

Table 4
Comparison of Location-Guided Screening and Current Practice Arms for Cases With Truth Determination of LSIL+*

Location-Guided Screening	Current Practice (Manual Screening)		
	LSIL+	Not LSIL+	Total
LSIL+	88	21	109
Not LSIL+	12	9	21
Total	100	30	130

LSIL, low-grade squamous intraepithelial lesion.
* P = .006 for equivalence; P = .0814 for superiority.

Table 5
Comparison of Location-Guided Screening and Current Practice Arms for Cases With Truth Determination of HSIL+*

Location-Guided Screening	Current Practice (Manual Screening)		
	HSIL+	Not HSIL+	Total
HSIL+	67	36	103
Not HSIL+	6	15	21
Total	73	51	124

HSIL, high-grade squamous intraepithelial lesion.
* P < .00001 for equivalence; P < .0001 for superiority.

In the LGS arm following the AutoPap initial screening, 218 slides (17.10%) were included in the no further review category and, therefore, were designated immediately as satisfactory and WNL. After review of the FOV only, 619 slides (48.55%) were designated as satisfactory and WNL

and received no full manual screening. The remaining 438 slides (34.35%) were triaged to a full manual screening. Because of scant cellularity as determined by the AutoPap, 20 slides did not generate FOV review. These slides were automatically sent for a full manual review (Table 2).

For cases with a final study truth of atypical squamous (or glandular) cells of undetermined significance and above (ASCUS+), the LGS arm identified 197 of 214 cases (sensitivity, 92.1%) and the CP arm identified 183 of 214 cases (sensitivity, 85.5%). For cases with final study truth of low-grade squamous intraepithelial lesion and above (LSIL+), the LGS arm identified 109 of 130 cases (sensitivity, 83.8%) and the CP arm identified 100 of 130 cases (sensitivity, 76.9%). For cases with a final study truth of high-grade squamous intraepithelial lesion and above (HSIL+; including endocervical adenocarcinoma in situ [AIS]), the LGS arm identified 103 of 124 cases (sensitivity, 83.1%) and the CP arm identified 73 of 124 cases (sensitivity, 58.9%). The results are shown in Table 3, Table 4, and Table 5, and graphically in Figure 1.

In the LGS arm, 1 slide (0.5%) represented a no-further-review population false-negative interpretation. It was interpreted as atypical glandular cells of undetermined significance (AGUS) by the adjudication process. All other no further review slides were truth determined to be WNL.

In the LGS arm, 8 cases with truth interpretations of ASCUS+ were triaged after FOV-only review to WNL. These cases were considered FOV false-negative cases and included 4 cases of ASCUS, 2 cases of AGUS, and 2 cases of LSIL. No case of HSIL+ was triaged inappropriately to WNL on the basis of the FOV-only review. However, in 2 cases of HSIL+, the full manual review downgraded slide interpretations to false-negative WNL after an appropriate FOV triage to full manual review. One case was truth determined to be cancer and had FOV triage interpretation of benign cellular changes (BCCs), and 1 truth-determined HSIL had FOV triage interpretation of ASCUS. In addition, 4 cases with a truth determination of AGUS and 3 cases of truth-determined ASCUS were downgraded to WNL after appropriate FOV triage to full manual review. In the truth-determined AGUS cases, FOV triage interpretations were all BCCs (4), and in the truth-determined ASCUS cases, FOV triage interpretations were 2 BCCs and 1 ASCUS. One additional case with truth determination of ASCUS was triaged to full manual review for adequacy reasons but received an interpretation of WNL following completion of that review. Two initially false-negative cases (ASCUS, 1; LSIL, 1) subsequently were identified as abnormal during the QC rescreening phase of the protocol, leaving 17 total false-negative ASCUS+ cases in the final LGS arm interpretations. Therefore, of 214 truth-determined abnormal (ASCUS+) cases in the

study, 205 (95.8%) were triaged appropriately by the FOV review. In addition, 128 (98.5%) of 130 LSIL+ cases, and all 124 (100%) HSIL+ cases were triaged to full manual review appropriately on the basis of the FOV review process (Table 6).

Of the 21 cases of HSIL+ that failed to be identified as such in the LGS arm, 19 (90%) were captured as abnormal at the level of ASCUS+, while 2 (10%) were false-negative cases. In the CP arm, of the 51 truth-determined HSIL+ cases that failed to be identified as such, 38 (75%) were captured at the level of ASCUS+, 11 (22%) received false-negative WNL interpretations, and 2 (4%) were called unsatisfactory.

Therefore, the overall sensitivity of identifying an HSIL+ case as abnormal (at any level) was 98.4% (122/124) in the LGS arm and 91.1% (113/124) in the CP arm (including 2 unsatisfactory cases as clinically “abnormal” results requiring pathologist review). In the LGS arm overall, there were 17 false-negative cases at the level of ASCUS+ (AGUS, 7; ASCUS, 7; LSIL, 1; HSIL, 1; cancer, 1). In the CP arm there were 26 total false-negative cases at the level of ASCUS+ (AGUS, 7; ASCUS, 8; HSIL, 2; AIS, 2; cancer, 7). An additional 3 unsatisfactory cases in this group were considered clinically abnormal, requiring pathologist review. Therefore, the sensitivity of detection of truth-determined ASCUS+ as the final-arm interpretation were 92.1% for the LGS arm and 87.9% for the CP arm. These figures represent the rates at which cases would have been triaged appropriately for pathologist review (Table 6).

Forty false-positive cases were identified in the LGS arm. By interpretive category, the false-positive cases were as follows: ASCUS, 23; AGUS, 5; LSIL, 9; HSIL, 1; and cancer, 2. Forty-six false-positive cases were identified in the CP arm. By interpretive category, they were as follows: ASCUS, 26; AGUS, 6; LSIL, 11; HSIL, 2; and cancer, 1. Calculations were made determining the false-positive and specificity rates in each arm compared with truth. In the LGS arm, the false-positive rate was 3.8% with a specificity of 96.1%. In the CP arm, the false-positive rate was 4.4% with a specificity of 95.1% (Table 7).

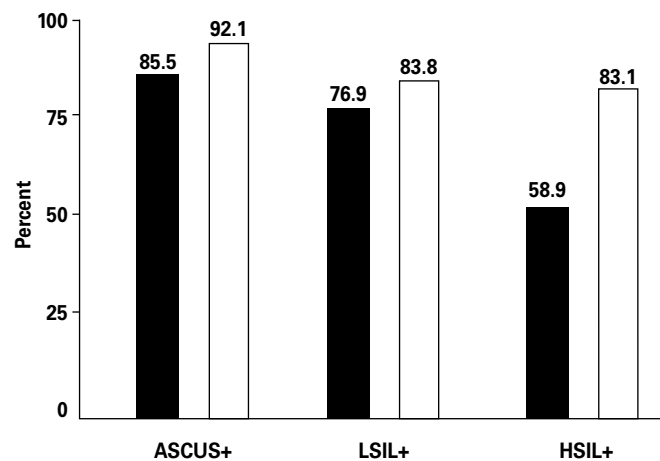


Figure 1 Relative detection sensitivities for abnormal cases of the current practice (manual screening) (black bars) and location-guided screening (white bars) arms. ASCUS, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

Of the 12 truth-determined unsatisfactory specimens identified in the study population, 8 were identified in the LGS arm (67%) and 10 were identified in the CP arm (83%) (Table 8). Three of the 4 slides not interpreted as unsatisfactory in the LGS arm were designated as “satisfactory but limited by,” and the 1 remaining slide was designated as “satisfactory for interpretation.” Both of the slides not interpreted as unsatisfactory in the CP arm were designated “satisfactory but limited by.”

The correlations between FOV review interpretation (n = 1,037) and final LGS arm and truth interpretations are shown in Table 9 and Table 10. In the discrepant cases (WNL vs ASCUS+) between FOV review and final LGS interpretation (n = 39), there were 21 cases (54%) upgraded to a higher level of interpretation, and 18 cases (46%) were downgraded to a lower level of interpretation. In the discrepant cases (WNL vs ASCUS+) between FOV review and study truth (n = 80), there were 30 cases (38%) upgraded

Table 6 False-Negative Cases and Appropriate Triage to Full Manual Review (Location-Guided Screening Only) and Pathologist Review

Interpretive Category	Current Practice (Manual Screening)		Location-Guided Screening	
	No. of Slides*	Triage to Pathologist (%)	No. of Slides*	Full Manual Review (%) / Triage to Pathologist (%)
ASCUS+	26	87.9	17	95.8/92.1
LSIL+	11	91.5	3	98.5/97.7
HSIL+	11	91.1	2	100.0/98.4

ASCUS, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.
* False-negative slides.

Table 7
False-Positive Cases and Specificity

Interpretive Category	Current Practice (Manual Screening)	Location-Guided Screening
ASCUS+	46	40
LSIL+	14	12
HSIL+	3	3
False-positive rate (%)	4.4	3.8
Specificity (%)	95.1	96.1

ASCUS, atypical squamous cells of undetermined significance; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

Table 8
Comparison of Location-Guided Screening and Current Practice Arms for Cases With Truth Determination of Unsatisfactory*

Location-Guided Screening	Current Practice (Manual Screening)		
	Unsatisfactory	Other	Total
Unsatisfactory	6	2	8
Other	4	0	4
Total	10	2	12

* Differences were not significant.

Table 9
Comparison of Agreement Between Field of View Review Only vs Final Location-Guided Screening Arm Interpretation

Interpretive Category	No. of Slides	Agreement*
Atypical squamous cells of undetermined significance+	228	207 (90.8)
Low-grade squamous intraepithelial lesion+	143	90 (62.9)
High-grade squamous intraepithelial lesion+	114	71 (62.3)

* Data are given as number (percentage).

Table 10
Comparison of Agreement Between Field of View Review Only vs Truth Interpretation

Interpretive Category	No. of Slides	Agreement*
Atypical squamous cells of undetermined significance+	205	175 (85.4)
Low-grade squamous intraepithelial lesion+	126	76 (60.3)
High-grade squamous intraepithelial lesion+	120	69 (57.5)

* Data are given as number (percentage).

to a higher level of interpretation, and 50 cases (62%) were downgraded to a lower level of interpretation.

QC rescreening (10% random) was performed on 106 cases in the CP arm. This rescreening found no additional abnormal cases. AutoPap-directed QC rescreening was performed on 304 cases in the LGS arm. This rescreening

process identified 3 additional abnormal cases (ASCUS, 1; LSIL, 2). These cases are included in the preceding totals, as they represent the final interpretations in the LGS arm **Table 11**.

Discussion

The only current FDA-approved applications of a computerized cervical cytology scanning device (the AutoPap system) are as follows: (1) as a triage instrument to identify slides having a low probability of abnormality that can be interpreted reliably as WNL without the need for human manual screening (no-further-review population), (2) the ranking of slides within the remaining review population to stratify slide risk of abnormality, and (3) the identification of slides with the highest probability of containing cytologic abnormality for use in directed manual QC rescreening. Because the AutoPap system scores slides by analysis of FOV containing cellular material, it is logical to expect that the device also would be able to identify FOV on each slide having high probabilities of containing abnormal cells. This ability to “predot” a slide before human review has substantial potential to improve the sensitivity of the cervical cytology slide examination process by helping to prevent “omission” errors that frequently occur in the manual screening process. In addition, the possibility of immediate identification of the “most abnormal” cells on a slide could lead to more specific interpretation. Finally, screening productivity also might be expected to improve because cytology screening personnel would not need to examine entire slides to arrive at reliable WNL interpretations on a substantial number of cases. Various levels of triage to WNL or potentially abnormal could take place: (1) instrument identification of lowest scoring population; (2) FOV review and triage to WNL when no potential abnormality is identified; (3) FOV review followed by routine manual screening with triage to WNL or a potential abnormal slide needing further pathologist review; and (4) directed FOV and full manual QC rescreening of the highest AutoPap-scoring cases remaining as WNL after triages in numbers 1 through 3.

The present study attempted to answer the question of the clinical feasibility and operating characteristics of such a process using liquid-based preparation slides in conjunction with the AutoPap Primary Screening System with LGS functionality. This device has been described conceptually with preliminary functionality described in previous preliminary studies.⁶⁻⁹ The present study differs from previous work in that coordinates of FOV (predots) were downloaded directly from the AutoPap into a computerized cytology workstation (SlideWizard 2). Slides were matched positively

Table 11
Results of Quality Control Rescreening

Study Arm	No. of Slides Rescreened	Abnormal Slides Identified
Current practice (manual screening)	106	0
Location-guided screening	304	3*

* Atypical squamous cells of undetermined significance, 1 slide; low-grade squamous intraepithelial lesion, 2 slides.

to computer records by barcode identification, and review of FOV was streamlined through a motorized microscope stage that automatically moved between FOV with the click of a mouse or the use of a foot pedal.

The study was performed using a 2-armed masked protocol directly comparing the results of the routine manual screening process with the AutoPap LGS system process. The study was fully adjudicated in a masked manner such that a cytologic “truth” was obtained for each slide that then could be compared with the result obtained for that slide in each study arm. As the LGS cytology screening process represents a multilayered approach to the overall final assessment of the individual slide, the protocol measured multiple outcomes in the process. First, the process of screening is a triage function that at its most basic level decides which slides are sent to the pathologist for further review. Hence, the first triage is at the level of WNL vs potential abnormality/refer to pathologist. The second level of assessment is actual interpretation of the level of the potential abnormality. The third level, pathologist review and, hence, final clinical interpretation, was not addressed in this study with regard to the final interpretations in each study arm. However, the final truth adjudication, with which the interpretations in each arm were compared, was achieved by pathologist review.

The results show the effectiveness of the first 2 triage levels of the AutoPap LGS system. The no further review fraction of cases contained 1 false-negative case, a slide interpreted as AGUS. The FOV triage of cases was perfect in the HSIL+ category, sending all truth-determined HSIL+ cases for a full manual review. In 2 HSIL+ cases (1.6%), subsequent full manual review erroneously downgraded the final interpretation to a false-negative interpretation of WNL. At lower levels of interpretive category (ASCUS+), 4 AGUS (1.9%) and 3 ASCUS (1.4%) cases also were triaged appropriately to full manual review but were downgraded to false-negative final interpretations. Of note is that the majority of these false-negative cases were sent for full manual review with FOV interpretations of BCCs, indicating that substantial subjectivity may have been involved in the discrepancy between final-arm interpretation and study truth.

By comparison, in the CP arm, in which triage is a singular process of full manual review and designation as

WNL or referral to pathologist, 11 false-negative HSIL+ cases (8.9%) were identified, with an additional 40 cases of HSIL+ being captured at the lower abnormal interpretive level of ASCUS+ (38 [30.6%]) or unsatisfactory (2 [1.6%]).

Therefore, using truth-determined HSIL+ as an end point, the sensitivity of detection of these cases at any level of abnormal or unsatisfactory interpretation was 98.4% for the LGS arm and 91.1% for the CP arm. These detection rates are important, because in actual clinical practice, these are the rates at which HSIL+ slides would be sent for pathologist review. In addition, appropriate triage rates for the LSIL+ and ASCUS+ categories are shown in Table 6.

Comparison with previous studies using earlier versions of AutoPap LGS (PAPMAP System) showed similar results. Huang et al⁷ reported 98.1% sensitivity for HSIL+ using this method. Also using the earlier system, a study sponsored by the National Health Service in the United Kingdom reported a 92.1% sensitivity for moderate to severe dyskaryosis+ (near equivalent to HSIL+ in the Bethesda System).⁹ In the present study, the sensitivity for an exact match within the category of HSIL+ in the LGS arm was 83.1% in comparison with the CP arm sensitivity of 58.9%. Despite the relative similarity between the 2 arms in proper triage of HSIL+ cases as abnormal, this substantial difference in proper final classification potentially may be explained by the LGS arm FOV review directing the cytologist to cells exhibiting greater degrees of abnormality than the cells found in the untargeted reviews of the CP arm. Further study is required to test this hypothesis.

The category of AGUS, while classified broadly with the ASCUS+ cases, seems to be distinct from ASCUS in terms of ultimate prognosis and management; hence, a breakout analysis is warranted. In this study, the AGUS false-negative rates were equivalent between the 2 study arms. Seven false-negative AGUS cases (of 32) were present in each arm (22% false-negative rate). Of note is that 2 additional cases of AIS (of 5) were missed in the CP arm (40% false-negative rate), while no case of AIS was missed in the LGS arm. The latter finding again suggests that the immediate attention drawn to potentially abnormal cells or cell groupings in the FOV review process ultimately may lead initial screeners to overall higher detection rates, particularly in this area of well-known screening difficulty.¹⁴

Comparison of specificities between the 2 study arms also shows a slight improvement when using the AutoPap LGS method. The specificity of the CP arm was 95.1%, while the specificity of the LGS arm was 96.1%. The types of false-positive interpretations (by interpretive category) were similar between the 2 arms. More than one half of all false-positive interpretations in each arm were associated with truth BCCs interpretations compared with the ASCUS+ truth, again pointing out the issue of potential subjectivity of

interpretation. This improvement in specificity closely mirrors that of the AutoPap Primary screening trials (without LGS),³ again suggesting that human review of the very-low-risk, AutoPap-defined, no-further-review population of cases may generate more false-positive results than warranted, when weighed against the very small increments of sensitivity improvement that might accrue from this practice. In this study, the sensitivity increment achieved would have been 0.47% with the only additional case generated being an AGUS.

A study such as this begs the question of generating final interpretations of all cases (not just those with interpretations of WNL) by the use of FOV review alone. The use of FOV triage, as the data show, is very efficient, with 95.8% of truth-determined abnormal cases (ASCUS+) being appropriately sent on for full manual screening. Furthermore, 100% of truth-determined HSIL+ cases were triaged for a full manual review. However, data generated in this study show that final interpretation is not performed as accurately with this method alone. Abnormal interpretations (ASCUS+) after FOV review alone agreed with the final LGS arm and truth interpretations of ASCUS+ in 90.8% and 85.4% of cases, respectively. As the interpretive category increases in severity, the correlation between FOV review and final LGS arm and truth interpretation diminishes. At the level of LSIL+, the correlations are 62.9% and 60.3%, respectively, and for HSIL+ the correlations are 62.3% and 57.5%, respectively. As would be expected, the correlation between FOV review and LGS arm interpretation is higher than between FOV review and truth because the former 2 interpretations are made by the same person (unmasked) at the same evaluation, whereas the latter are 2 independent, masked assessments. These data indicate that FOV review may not, in some cases, show the highest level of abnormality present on the slide. This confirms work by Chang et al,⁸ who showed that by using LGS, higher levels of “abnormality capture” were achieved compared with manual screening, although the highest interpretive level of abnormality was not always attained.

As noted in previous clinical trials,¹⁵ the value of directed QC rescreening was documented in the present study. In the CP arm, the 10% random QC rescreening process detected no false-negative cases. In the LGS arm, the directed QC rescreening process, aimed specifically at the highest scoring cases remaining as WNL after initial evaluation, detected 3 additional abnormal cases (ASCUS, 1; LSIL, 2) that had been missed by initial FOV and full manual reviews. Therefore, this type of directed QC process further enhances the already more sensitive process of AutoPap LGS initial evaluation. In addition, as was shown in previous clinical trials,¹⁵ the robust nature of the no-further-review population of slides was documented. In this study, only 1

false-negative case of AGUS was identified in this non-manually screened fraction of cases. No cases of squamous intraepithelial lesion or malignancy were misinterpreted as WNL by the computerized system alone, and by the very nature of the process, no false-positive cases can ever be present in this group.

Based on the results of this study, screening of liquid-based cervical cytology slides using the AutoPap LGS system has the potential to be an effective method of prescreening slides, completely eliminating some slides from human review, and subjecting yet another substantial percentage to FOV review triage only. Both the sensitivity and specificity of the procedure seem to be improved compared with routine manual screening practice.

Regarding potential improvements in productivity by using this procedure, we made no attempt in this trial to directly study the time taken in the review process. However, results from a previous time and motion study performed in conjunction with the National Health Service study using an earlier LGS method in a similar manner showed a 24% improvement in productivity over the manual screening process.⁹ To speculate about the potential for productivity improvement with this new LGS method, data generated from this study can be modeled to show what might be expected from the use of this procedure. The College of American Pathologists Workload Recording data generated from time studies on the screening of conventional cervical cytology smears shows that screening a negative smear takes, on average, 5.0 minutes, while screening a positive smear takes, on average, 6.5 minutes.¹⁶ If one makes the assumption that screening liquid-based smears is 25% faster⁴ (John Bishop, unpublished data, 1996, and our anecdotal observations), a negative PREP specimen might be expected to take, on average, 3.8 minutes to screen, while a positive PREP specimen might take, on average, 4.9 minutes to screen. A further assumption is an average time of 1 minute for FOV review. Based on these assumptions, the total screening workload in the CP arm would have been as follows:

$$(4.9 \times 230 \text{ abnormal slides}) + (3.8 \times 1,045 \text{ WNL slides}) \\ = 5,098 \text{ minutes}$$

In the LGS arm, 218 slides received no FOV or manual screening, and 619 slides received FOV screening only (619 minutes). This leaves 438 slides (234 abnormal and 204 WNL) receiving FOV (but 20 slides in this group with scant cellularity received only full manual review [438 – 20 = 418 minutes]) and/or full manual screenings (4.9×234) + $(3.8 \times 204) = 1,921.8$ minutes, for a total of 2,958.8 minutes. The screening workload in the LGS arm, based on the aforementioned assumptions, is therefore 58.04% of that in the CP arm. One can change the assumptions to change the overall

workload percentages; however, the combination of the AutoPap no-further-review population of cases and the FOV-only review triage function is highly likely to improve laboratory screening productivity.

Further study of this method with full-scale clinical trials will be necessary before LGS capability can be used clinically in the United States. The results of the present study, carried out using a trial-type protocol, suggest that this system not only will be effective as a method to improve the accurate practice of cervical cytology but also will be an important productivity tool in the potentially understaffed cytology laboratory.

From the Cytopathology Unit, ViaHealth, Rochester, NY.

Supported by TriPath Imaging, Burlington, NC (to ViaHealth).

Presented in part at the 91st Annual Meeting of the United States and Canadian Academy of Pathology, Chicago, IL, February 25, 2002.

Address reprint requests to Dr Wilbur: Dept of Pathology, Massachusetts General Hospital, 55 Fruit St, Boston, MA 02114.

** Dr Wilbur is currently with the Cytopathology Unit, Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston.*

Dr Wilbur is a member of the speakers' bureau of TriPath Imaging. Ms Parker and Mr Foti have no financial relationships with manufacturers of any products discussed in this article.

NOTE: During manuscript preparation, TriPath Imaging changed the names of 2 products referred to in this study: AutoCyte PREP is now SurePath; AutoPap is now the FocalPoint System.

References

1. Lee KR, Ashfaq R, Birdsong GG, et al. Comparison of conventional Papanicolaou smears and a fluid-based, thin-layer system for cervical cancer screening. *Obstet Gynecol.* 1997;90:278-284.
2. Bishop J. Comparison of the CytoRich System with conventional cervical cytology: preliminary data on 2,032 cases from a clinical trial site. *Acta Cytol.* 1997;41:24-28.
3. Wilbur DC, Prey MU, Miller WM, et al. 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;42:214-220.
4. Bur M, Knowles K, Pekow P, et al. Comparison of ThinPrep preparations with conventional cervicovaginal smears: practical considerations. *Acta Cytol.* 1995;39:631-642.
5. PREPStain [product insert]. Burlington, NC: TriPath Imaging; 2002.
6. Lee JSJ, Kuan L, Oh S, et al. A feasibility study of the AutoPap system location-guided screening. *Acta Cytol.* 1998;42:221-226.
7. Huang TW, Lin TS-M, Lee JS-J. Sensitivity studies of AutoPap System Location-Guided Screening of cervical-vaginal cytologic smears. *Acta Cytol.* 1999;43:363-368.
8. Chang AR, Lin WF, Chang A. Can technology expedite the cervical cancer screening process? a Hong Kong experience using the AutoPap Primary Screening System with location-guided screening capability. *Am J Clin Pathol.* 2002;117:437-443.
9. FocalPoint, Export Only Version [product insert]. Burlington, NC: TriPath Imaging; 2002.
10. PrepStain System Operator's Manual. Burlington, NC: TriPath Imaging; 2002.
11. Kurman RJ, Solomon D. *The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses.* New York, NY: Springer-Verlag; 1994.
12. Solomon D, Davey D, Kurman R, et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA.* 2002;287:2114-2119.
13. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York, NY: John Wiley; 1981.
14. Mitchell H, Medley G, Gordon I, et al. Cervical cytology reported as negative and risk of adenocarcinoma of the cervix: no strong evidence of benefit. *Br J Cancer.* 1995;71:894-897.
15. Patten SF, Lee JSJ, Wilbur DC, et al. The AutoPap 300 QC System multicenter clinical trials for use in quality control rescreening of cervical smears, I: prospective intended use study. *Cancer.* 1997;81:343-347.
16. College of American Pathologists. *College of American Pathologists Workload Recording Method and Personnel Management Manual.* Northfield, IL: College of American Pathologists; 1992:94.