Automation and the Endangered Future of the Pap Test

Dorothy L. Rosenthal*

Deaths from uterine cervical cancer have been dramatically reduced since the development in the 1940s of the Papanicolaou smear (also called the Pap smear or test) for detection of precursor lesions. Along with the vaccines for smallpox, polio, and other viruses, this screening test has been hailed as one of the most successful public health measures and is the only cancer screening test proved to reduce mortality (1).

Since its first availability for patient care, the categorization of the Pap smear as a clinical laboratory test, with no allowance for difficult and time-consuming microscopy, contributed to the disparity between actual cost and reimbursement by insurance companies. Even now, the U.S. Government’s Health Care Financing Administration continues to price the Pap smear at the same low reimbursement level as highly automated tests, despite expert testimony against such equivalence. As a result of this pricing, many laboratories are being forced to discontinue performing Pap tests; therefore, more and more Pap smears are being sent to larger reference laboratories for analysis.

A number of deaths from cervical cancer following misreadings of smears (2,3) has led to a near epidemic of legal actions based on perceived errors in Pap test diagnoses (4–7). Publicity given to these tragedies presents an impression that errors in diagnoses should be a rare event, which has reinforced a public misconception that the cervical cancer screening test is error-free. Thus, decision-makers at pathology laboratories face unrealistic expectations from a litigation-prone public, along with other critical economic questions, but they have no certain solution to their dilemma (8). Understandably, many cytopathologists are questioning the wisdom of continuing to interpret Pap smears.

Automation is one frequently proposed answer to the question of how to save this endangered screening test. To date, several automated systems have been designed to prevent the most common causes of errors encountered during processing and interpretation of cellular samples.

This commentary will describe the current status of the new technologies and will examine their potential as cost-effective solutions. Readers should be aware that much of the available performance data cited here are results from manufacturer-sponsored trials.

Potential for Improvement

Before exploring the potential of automated systems for improving the traditional Pap smear, it is useful to look at each stage of the testing process. In the traditional sequence of steps from preparation of the patient to acquisition of the cellular sample to microscopic decision, opportunities for error occur along the way. Viewed in this context, the scientific advantages and cost-effectiveness of the new automated systems or devices can be better appreciated.

Preparing the Patient

Patients need to be educated in two areas: 1) what they can do to optimize the quality of the cellular sample that is obtained from them by their clinician and 2) what they can expect in terms of the reliability of the screening test. Often, patients lack or choose to ignore advice regarding their personal physical preparation, specifically, no douching or intercourse during the 24 hours preceding the Pap test, and preferable sampling at mid-menstrual cycle. Some popular magazines have helped to educate the public. For example, a particularly thorough article on the topic appeared in a recent issue of the magazine Glamour (9). Programs aimed at dispelling misconceptions about the reliability of the test and its rate of error are being offered by many members of the pathology and oncology professions. Public education is also the responsibility of U.S. federal agencies such as the Food and Drug Administration (FDA) and the Federal Trade Commission (FTC), which protect the public from unfounded claims and misleading advertising by device manufacturers or their agents.

Obtaining the Sample

Cytologic diagnosis of the Pap smear is only as good as the quality of the sample. The cytologic sampling devices that were widely used in the early years included a cotton-tipped applicator that could be inserted into the endocervical canal of the patient and a contoured spatula named after its inventor, J. E. Ayre, to fit the squamo-columnar junction (transformation zone) for the collection of cellular samples. Several sampling devices, such as Cytobrush (MedScand, Hollywood, FL), Cervex...
Brush™ (Rovers, Boonton, NJ), and Papette (Wallach Surgical Devices, Orange, CT), that have been developed during the past 10 years and are now widely used have greatly improved the amount of the endocervical sample obtained from the patient. The use of these improved devices has facilitated the detection of lesions at the transformation zone (origin of 90% of cervical cancers) and in the canal. However, these devices do not eliminate inflammatory cells or excessive blood, both of which impede fixation, staining, light transmission, and interpretation of the test. In fact, if improperly used, these devices may cause the patient to bleed. In addition, the cells obtained may be distorted if the sample is applied to the slide with too much pressure.

To minimize these potential problems of cytologic sampling, instructions for the new liquid-based (thin-layer) preparation methods specify the use of collection devices other than cotton swabs or Ayre spatulas. The amount of material and the pressure with which it is applied to the slide are critical to proper fixation, staining, and optical transparency. Although a liquid-based sample avoids pressure distortion, it does not eliminate the possibility that the sample could contain artifacts.

Fixing and Staining

Traditionally, immediate fixation and staining of the cellular sample on the slide with 70% ethyl alcohol and Papanicolaou stain have been the professional standard. This fixation and staining combination results in a cellular sample that not only has well-defined and tinted morphologic features, but also has transparency to allow for microscopic visualization of nuclear and cytoplasmic boundaries through multiple layers of epithelial cells. Although the experienced human eye can be forgiving, the performance of an imaging system is considerably compromised if the slide quality is not optimal. A major advantage of the new automated screening devices is the stringent demands of each system for high-quality staining and coverslipping, including optical uniformity of the cellular target.

Locating Cellular Abnormalities

Accurate location of abnormal cells or patterns on a routine Pap smear requires competent training, experience, and constant vigilance on the part of the human scanner. Because of the relatively low prevalence of disease (less than 5%) in most populations of women tested, routine Pap smear screening consists of long, tedious intervals between interesting cases. Important abnormal cells may be missed because of inattention caused by monotony of the visual scene. The new automated stage controllers and scanning devices have been designed to eliminate errors that are borne of human distraction and fatigue. Thin-layer preparations remove obscuring blood and inflammatory cells and, unlike a tired human, automated scanning devices theoretically do not miss rare small cells (10).

Interpreting Cellular Abnormalities

The conventional Pap smear has been estimated to contain between 50,000 and 300,000 epithelial cells (11), which makes detection of rare abnormal cells one feat and their interpretation yet another. Cellular morphology is the basis of clinical cytology, and the range of cytomorphologic changes in both wellness and disease is wider in cervical–vaginal specimens than in samples from any other body site. Perhaps the greatest challenge facing cytologists and developers of artificial intelligence systems is defining permutations of normalcy.

The Bethesda System (12) introduced the term “atypical squamous cells of undetermined significance” (ASCUS) to categorize those cases not considered normal, but whose morphology falls short of neoplastic changes. Although many epithelial changes diagnosed as ASCUS behave in a benign fashion, a substantial number of cell samples classified as ASCUS precede diagnoses of more severe lesions, resulting in heightened surveillance of the patient.

The use of washed, liquid-based samples has eliminated a substantial proportion of ASCUS diagnoses, allowing the observer to place the case in either a benign or a potentially premalignant category (13–17). The scanning devices that display all cells of concern together present an “enriched” sample to the viewer, whereas the routine Pap smear may contain only rare and scattered abnormal cells of no apparent importance.

New Technologies

New technological applications for Pap test improvement have been designed to ensure adequate sampling of abnormal cells; proper fixation and staining; location, identification, and interpretation of abnormal cells; and final classification (diagnosis) of the entire sample. They can be grouped into two general categories: 1) specimen preparation devices and 2) computerized alarm-detection devices (Table 1).

Specimen Preparation Devices

ThinPrep® 2000 (Cytyc Corporation, Boxborough, MA)

The concept of a suspension of disaggregated cells captured on a membrane filter and transferred to a glass slide has gradually gained acceptance as an improvement over the traditional Pap smear. Approved by the FDA in 1995, Cytyc’s ThinPrep 2000 Pap test was originally conceived to accommodate the exacting demands imposed when traditional imaging devices are used for analysis of cervical epithelial cells.

How it works. In brief, the cellular sample is scraped from the cervix with one of the recommended collection devices (see above in the section entitled “Obtaining the Sample”) and then swished vigorously in the provided vial containing the proprietary preservative/fixative which is called PreservCyt®. The resulting cell suspension is stable at room temperature for days to weeks, thus allowing reasonable time for transport to the laboratory and storage. The single-sample processor utilizes a rotory device inserted into the collection vial to disaggregate any cell groups; this action is followed by suction of the suspension but not the cells through a polycarbonate filter. Suction spontaneously ceases once all the filter holes are blocked by cells or debris. Trapped cells are automatically touch-transferred from the filter to an area 20 mm in diameter on a glass slide (16).

The ThinPrep test has several advantages over previous methods (18,19) that have been used to produce microscope slides from cells in suspension. Cellular samples containing separate epithelial cells in a monolayer, and with no inflammatory cells, were necessary to enable computerized microscopes to digitize information regarding the cellular features to be pro-

Journal of the National Cancer Institute, Vol. 90, No. 10, May 20, 1998

COMMENTARY 739
cessed by image analysis algorithms for the purpose of cell identification and specimen classification. However, the Thin-Prep system is designed to retain clusters of cells as well as single, small epithelial cells from high-grade lesions. The benefit of flushing away the visually obscuring neutrophils is obvious. In addition to the generation of a preparation without obscuring inflammatory cells, the cervical sample is immediately transferred from the collection device into PreservCyt, which partially fixes the cells. Fixation is completed during the processing of the cervical sample, which results in minimal deterioration of cell features.

Performance data. Performance data submitted to the FDA prior to approval of Cytyc’s ThinPrep were compiled from a six-center clinical trial involving 7360 women who were enrolled in a split-sample, matched-pair, double-blinded study (15). The sites included three screening centers and three hospitals, which resulted in a study population that was generally balanced in geographic and socioeconomic representation. A single cervical sample collected from each woman was first smeared on a glass slide and processed in the traditional manner. The specimen that remained on the sampling device was then transferred from the collection device into PreservCyt, which partially fixes the cells. Fixation is completed during the processing of the cervical sample, which results in minimal deterioration of cell features.

Table 1. Steps involved in Pap smear acquisition, processing, and interpretation, chances for errors, and possible remedies by new technologies

<table>
<thead>
<tr>
<th>Task</th>
<th>Error</th>
<th>Remedy</th>
<th>New device</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>Douching/menses</td>
<td>Patient education</td>
<td></td>
</tr>
<tr>
<td>Sampling</td>
<td>Lack of endocervix</td>
<td>Clinician education</td>
<td>Cytyc ThinPrep®</td>
</tr>
<tr>
<td></td>
<td>Inflammation/blood</td>
<td>New samplers</td>
<td>AutoCyte PREP®</td>
</tr>
<tr>
<td>Fixation/staining</td>
<td>Air-drying, poor stain</td>
<td>Liquid-based preparations</td>
<td>Cytyc ThinPrep®, AutoCyte PREP®</td>
</tr>
<tr>
<td>Locating cells</td>
<td>Incomplete scanning</td>
<td>Stage controllers</td>
<td>AutoCyte PREP®, Cytyc ThinPrep®</td>
</tr>
<tr>
<td></td>
<td>Cells obscured</td>
<td>Liquid-based preparations</td>
<td>AutoPap®, PapNet®</td>
</tr>
<tr>
<td></td>
<td>Too few cells</td>
<td>Optimal scan</td>
<td>AutoCyte SCREEN®</td>
</tr>
<tr>
<td>Identify cells</td>
<td>Significance not apparent</td>
<td>Display all alarms together</td>
<td>PapNet®, AutoCyte SCREEN®</td>
</tr>
</tbody>
</table>

Over a population with lower prevalence of disease and that were prepared by the conventional method. Lee et al. acknowledged that the lower grade lesions, which are more common in the population served by the screening centers, were easier to detect in smear specimens prepared by the ThinPrep system than by conventional methods. These authors also speculated that cytologists at the hospital centers may have been more likely than those at screening centers to have diagnosed on conventional Pap smears a borderline abnormality in the presence of inflammation in order to ensure follow-up for high-risk patients.

Additional results of the clinical trial (15) by Lee et al. included fewer cellular specimens that were obscured by inflammation or blood when processed by the ThinPrep method (P < .001), but the conventional smears had a lower percentage of samples that did not contain endocervical cells (9.4% for conventional smears versus 15.8% for ThinPrep). This difference can be explained by an ethical constraint. Because ThinPrep 2000 had not yet received FDA approval for clinical use, the single cervical sample had to be first applied to a slide for the conventional Pap smear, with only the residual specimen rinsed into the PreservCyt vial for processing with use of the ThinPrep 2000 system. Presumably, this procedure resulted in less diagnostic material, including endocervical cells, for the ThinPrep slide. Similar data on the presence of endocervical samples in specimens prepared by use of both the ThinPrep and conventional methods were reported by Laverty et al. (20) in a study involving 2125 women in Australia.

When the entire cervical scrape was placed directly into the PreservCyt vial, Corkhill et al. (21) found that the endocervical component was present approximately as often on the ThinPrep slide as on the conventional Pap smear slide. However, study results that would include archival data are invalid because the ThinPrep procedure restricts the sampling device to exclude the use of Ayre spatulas and cotton-tipped applicators, the standard devices used for many years to make Pap smears. This difference in methodology introduces a variable that has not usually been considered or for which there have not been controls because the type of collection device used was not recorded when the sample was obtained.

These results of Corkhill et al. (21) have been reinforced by the results of an evaluation of the use of the ThinPrep method in a large clinical practice, reported by Papillo et al. (22). Over a
7-month period, 16314 ThinPreps were diagnosed, and a subset of 8574 tests was compared with historic data from patients of the same practitioners who had been screened using conventional Pap smears. When the ThinPrep method was used, the percentage of test results that were normal improved by 1.71%, the percentage of ambiguous or borderline diagnoses decreased by 26.59%, and the percentage of cases diagnosed as LSIL or higher increased by 52.15% (P<.001 for all three differences; proportional differences were analyzed for significance with the use of a Z statistic test). Since the samples were not split, only a small percentage of ThinPreps (4.38%) lacked endocervical cells, which compared favorably with the Pap smears (4.8%). However, the collection device may play an important role in obtaining endocervical cells, and the Pap smears were obtained by a variety of unspecified samplers, whereas the ThinPreps were collected with several of the newer devices (16).

Since all of the PreservCyt cell suspension is not initially used to prepare the first slide, the remainder is available to prepare additional slides if only a few abnormal cells are present in the first aliquot. Ancillary studies, e.g., human papillomavirus (HPV) subtyping (23), can be performed on remaining cells. Finally, multiple slides from a single sample of a known lesion can be used for teaching, proficiency testing, and routine laboratory quality-control checks.

**Training.** The manufacturer, Cytex Corporation, offers a 3-day training program for new users (Table 2). The bases for the cytologic criteria taught in this course are well-established cell features encountered in benign and neoplastic processes of the uterus. Because of differences in fixation and staining (chromatin tends to stain more palely with the ThinPrep method, which uses a modified Papanicolaou stain), cells do not appear as abnormal as they do with standard fixation. One of the benefits of the ThinPrep method—a clean background—can also be a detriment when detecting invasive lesions. Tumor necrosis, a classic cytologic criterion of epithelial basement membrane invasion, is washed away in the filtration process. Its absence can lead to a misinterpretation of the sample, classifying it into a diagnostic category that is lower than appropriate. The result could be a potentially disastrous false-negative diagnosis. The learning curve for laboratory personnel to reliably classify the variety of lesions under the ThinPrep system is directly related to the size and complexity of the laboratory’s case load and may require several months of training for each person. Also, because invasive carcinoma is seen infrequently, it takes time and experience before the observer can master the microscopic characteristics of malignant cells prepared by this new technique.

**Cost-effectiveness.** Currently, the ThinPrep 2000 is a single-sample processor, requiring time for laboratory personnel to prepare a slide from each specimen (there is no processing time for the Pap smear), with more costly disposable accessories (approximately $10) compared with the routine Pap smear (less than $2). These factors have made the system unattractive financially, despite the clarity of the microscopic images. A version of the machine that can simultaneously process multiple samples is now under development and may improve the cost-effectiveness of the system.

**AutoCyte PREP® (AutoCyte, Inc., Elon College, NC)**

AutoCyte, Inc. (formerly Roche Image Analysis), has submitted to the FDA an application for pre-market approval of AutoCyte PREP, a system for processing multiple samples (48 per hour).

**How it works.** Instead of the rotor method of cell disaggregation used by ThinPrep, AutoCyte PREP utilizes the fluid friction of a syringe device to disaggregate clusters of cells. Cells processed by this method do not completely disaggregate (24): Cells remain in architectural relationships to one another, even though the average size of the clusters (and therefore, the thickness on the slide) is decreased. Density-gradient centrifugation is used to separate epithelial cells from inflammatory cells and debris. Epithelial cells are then transferred to the slide by gravity, which helps to prevent shape distortion by any additional physical forces such as suction. After depositing a sample of 40 000–70 000 washed cells on a circular area (13 mm in diameter) of the slide, the system automatically fixes and stains slides individually before their removal from the processor, thereby eliminating the chance of cross-contamination of samples. Since these critical steps are controlled before automated scanning or visual review by a cytotechnologist or cytopathologist, the sample contains cells with optimal morphologic criteria and, thus, greatly facilitates alarm location and interpretation.

**Performance data.** Results from a study reported by Wilbur et al. (17), involving six clinical centers and 286 patients, indicated that there was exact agreement between the CytoRich (since renamed AutoCyte PREP) thin-layer system and conventional Pap smears for 78% of all samples and for 95% of samples classified as being separated by a single Bethesda System diagnostic category. A larger and more recent multicenter trial reported by Bishop (25) compared diagnostic performance of cytologists using liquid-based samples prepared by the AutoCyte

### Table 2. Training requirements for new technologies*

<table>
<thead>
<tr>
<th>Technologies</th>
<th>ThinPrep®</th>
<th>AutoCyte PREP®</th>
<th>AcCell</th>
<th>Pathfinder®</th>
<th>AutoPap®</th>
<th>PapNet®</th>
<th>AutoCyte SCREEN®</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample</td>
<td>Thin layer</td>
<td>Thin layer</td>
<td>Pap smear</td>
<td>Pap smear</td>
<td>Pap smear</td>
<td>Pap smear</td>
<td>Thin layer</td>
</tr>
<tr>
<td>Training†</td>
<td>3 days</td>
<td>TBE</td>
<td>&lt;1 wk</td>
<td>&lt;1 wk</td>
<td>None</td>
<td>3 days</td>
<td>TBE</td>
</tr>
<tr>
<td>Ancillary tests</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>DNA estimate possible</td>
</tr>
</tbody>
</table>

*TBE = to be established; N/A = not applicable.
†Exclusive of operating device.
PREP system and conventional cervical smears. In a masked, split-sample protocol, 8983 patients were enrolled from eight clinical centers. With the use of AutoCyte PREP slides, a 31% overall increase in detection of important lesions was realized—specifically, 46% more LSIL, 6% more high-grade squamous intra-epithelial lesions (HSIL), and 25% more cancers (P<.001 with use of the McNemar’s test for all three differences) than the conventional Pap smears. Concordance between the two diagnostic methods was achieved in 86.9% of the 8983 cases and in 97.4% of those classified within a single diagnostic category.

Training. The time necessary for cytotechnologists and cytopathologists to become proficient at making diagnoses based on samples prepared by this product has not yet been assessed. However, even if the liquid-based samples are used strictly for human review, the smaller area of the slide that is to be scanned, as well as the improved cell preservation, should help reduce fatigue and prevent misinterpretation or oversight of partially or completely obscured cells. A smaller screening area should also help to increase the analyst’s productivity. On the other hand, there may be fewer abnormal cells in each sample, and background (contextual) clues and cell associations are diminished; therefore, heightened vigilance may be required, leading to increased screening time and stress for the cytologists.

Cost-effectiveness. Bishop (25) has attempted to determine the comparative costs of analysis between the conventional Pap test and a primary automated screening method that uses a liquid-based sample. Some of his data regarding costs were based on general experience. Other costs were computed on the basis of economic assumptions made from available information, since no system had as yet been approved for primary screening. Using the AutoCyte PREP System, Bishop processed and evaluated 2106 cervical samples. His calculations included wages for the technologist and pathologist, consumable and processing costs, overhead, and amortization of capital equipment. He concluded that the cost of production of the conventional Pap smear would be $9.74 (in 1994), whereas a primary screening system using liquid-based samples would cost an estimated $12.07 per slide. In 1994, the local Medicare reimbursement for a uterine cervical screening test was $7.15. When economy of scale is considered for larger volumes, the difference in cost of production of an automated primary screening system can be reduced from 24% higher to 14% higher than the traditional screening method, based on an annual volume of 40000 specimens.

Bishop’s calculations, however, were based on the premise that a cytotechnologist can screen the maximum number of samples allowed by federal law, i.e., 12.5 conventional slides per hour, or 25 slides per hour when one half or less of the slide area is covered by cells. In another study, Laverty et al. (20) estimated that 3 minutes would be the average time necessary to screen a slide prepared by the CytoRich system. Anyone who has “pushed glass” knows how difficult it would be to place 25 slides in an hour on a mechanical stage (required for optimal coverage of the area), consider the patients’ histories, contemplate the complex visual scenes on each slide, mark abnormal cells, and classify the slides for release or further review.

**Computerized Alarm-Detection Devices**

Two types of computerized systems—1) stage control devices and 2) automated scanners—have been designed to improve location and identification of abnormal cells and to assist in the interpretation of abnormalities.

**Stage Control Devices**

Two commercially available products, both of which are attachments to standard light microscopes, are designed to enhance the cell location portion of screening by producing slide maps of areas visited. CompuCyte’s Pathfinder® system automatically records the mechanics of slide review. AccuMed’s AcCell™ system controls and records the functions of screening, especially those related to slide travel, such as field-of-view overlap and focus. A new component of the AcCell system, called TracCell™, automatically eliminates acellular areas of the slide before it is subjected to human screening, thereby reducing the amount of review time needed per slide. FDA approval for use of the basic system in diagnostic applications was not required. However, because TracCell excludes portions of the slide from visual inspection, pre-market approval by the FDA was required and was obtained in 1997.

**Pathfinder® (CompuCyte, Cambridge, MA)**

**How it works.** The Pathfinder system was designed to prevent errors of nondetection during screening (26). This system does not actually control the process; instead, it records the mechanics of screening. The system consists of a microcomputer interfaced to a standard light microscope stage by means of position sensors. As the cytotechnologist traverses the slide, the \(x\)-\(y\) coordinates can be tracked and displayed on a 5-inch video monitor. The degree of overlap is automatically recorded for each field, and the average overlap is calculated for the entire slide. Dwell time is also captured each time the cytotechnologist stops to inspect an object, usually a cell or cell cluster. Objects can be electronically marked according to their degree of abnormality so that only the worst objects are recalled for final diagnosis. Quality-assurance data, such as time per slide and number of slides reviewed per 8-hour day, are automatically captured. Pathfinder constructs a permanent record of the physical events that preceded the final slide classification by the cytotechnologist.

**Performance data.** Early performance data support the manufacturer’s claim that use of Pathfinder enhances screening accuracy and decreases rescreening time for quality-assurance evaluation (27). Berger (28) used Pathfinder during routine 10% rescreening to increase the detection of abnormal cases that were initially missed. Over a 9-month period the false-negative rate for ASCUS cases fell from 1.6% (37 of 2336) to 0.7% (12 of 1772), a 56% decrease. Berger used no tests to calculate statistical significance. Furthermore, the statistical significance of differences for higher grade abnormalities could not have been determined with accuracy because of the relatively low numbers of such cases in the study.

**Cost-effectiveness.** Berger (28) estimated the direct incremental cost of the device to be approximately 10 cents per slide, calculated with a 5-year amortization of the equipment and training. The increased cost is justified by the resultant increased
efficiency of the staff when diagnosing the slides and the decreased time collecting quality-assurance data. Computerized documentation of standardized screening can be used to show laboratory performance.

NeoPath is currently developing an interface with the Pathfinder microscope system (29) that will allow the cytologist to relocate abnormal cells already detected by AutoPap®.

AcCell™ Series 2000 (AccuMed International, Inc., Chicago, IL)

How it works. AcCell Series 2000 is an automated support system for screening, marking, and classifying specimens (30). Through links to the laboratory information system, patient demographics are automatically accessed via a bar-code reader and displayed on the computer screen for the cytotechnologist before screening begins. Personal preferences of the individual screener can be preset, such as mode of screening (horizontal versus vertical), scanning speed and direction, dwell time, field overlap, and area to be reviewed—the last two of which are especially critical with the variation in area of cellular deposition by the liquid-based processors. Focus is controlled by the screener. As the slide is traversed, the viewer can stop the automated process and explore an area of interest. When screening resumes, the automated stage returns to the exact location where the diversion occurred and proceeds in proper direction, so that no area is missed. Cells or areas of interest are electronically recorded and the slides marked automatically.

During the screening process, the cytotechnologist captures diagnostic information on the computer screen in a customized format. At the end, if the case is normal, a case report is generated; if the case requires further review, the screening information can be transferred to the pathologist’s AcCell workstation. Quality-assurance data, such as the number of slides screened per 8-hour period and the total time spent screening, are automatically captured.

A related product from AccuMed International, Inc., the TracCell 2000 specimen mapping system, is a prescreening device (31). TracCell, which selectively maps slides for automatic presentation to the viewer, is designed to operate unattended through use of cassettes that have been preloaded with bar-coded slides. The same cassettes can then be inserted into the AcCell carrier for robotic transfer of slides to the automated stage. During screening with AcCell, instead of scanning the entire slide, the TracCell mapping routes the motor-driven stage to those areas that contain cellular material and presents them in approximate focus, thus automatically decreasing the amount of time needed to review the slide.

Performance data and cost-effectiveness. Performance data on this system are not yet available, and cost-effectiveness analysis has not been possible.

AUTOMATED SCANNERS

Automated scanners are designed for use as rescreening devices, but they may also be used for prescreening. Some have potential use for primary screening. Performance standards differ, depending on whether the device is intended for rescreening or primary screening (32). A rescreening device scans after the cytotechnologist has reviewed the slide, either verifying those samples previously designated as normal or automatically selecting as potentially abnormal those slides worthy of human review for quality control. Systems that display digitized images select for human reconsideration the most troublesome areas of a slide after a human screener has designated the slide as normal. At that point, the cytotechnologist decides whether those already screened slides are definitely normal or in need of further review. A primary screening device is intended to scan a cellular sample before a human sees it and then indicate whether it is normal or abnormal; normal slides would never be seen by a human unless randomly chosen for quality control of the system.

Two automated devices have received FDA approval for rescreening Pap smears. Differing system designs and degrees of human interaction distinguish Neopath’s AutoPap® 300 QC (FDA approval received in 1995) and Neuromedical System’s PapNet® (FDA approval received in 1996) (Fig. 1). Both use light microscopes, Papanicolaou stain, and traditional smears. AutoPap automatically scans the slide with the use of traditional image analysis techniques. PapNet adds neural networks as a second tier of decision-making (33). A third device currently undergoing clinical trials prior to submission to the FDA for pre-market approval is AutoCyte SCREEN® (AutoCyte, Inc.), which combines features of AutoPap and PapNet and utilizes a sample prepared from a cell suspension.

The various systems employ different decision-making processes (34). Standard image processing algorithms (as used in the AutoPap 300 QC system) depend on classic morphometrics, e.g., nuclear and cytoplasmic area, nuclear–cytoplasmic ratio, nuclear optical density, and nuclear shape. Neural networks (such as those used in the PapNet system) involve more global pattern recognition tasks, including correct identification of cell groups and noncellular artifacts (33). Both software designs are tested and refined until decision errors are minimized and the desired false-negative rate is achieved.

For every device, there is a trade-off between sensitivity (measure that the sample is definitely abnormal) and specificity (measure that the sample is truly free of disease) (35). Cytologic slide classification is based on the number of abnormal cells and the severity of the abnormality. A human will detect and interpret one abnormal cell and classify the slide as abnormal. If an automated device is designed to classify on the basis of a preset threshold that allows even a few abnormal cells to be included in the normal category, then it is conceivable that one cancer cell could be missed by the machine (36). Most clinicians agree that, even with these new scanning devices, perfection is desirable but not practical and are satisfied that a standard of performance “equal to or better than the human” is reasonable. While a 5% false-negative rate is considered by many cytopathologists as the lowest attainable, a standard and acceptable human false-negative rate has yet to be defined by the medical profession for the routine Pap test (37).

AutoPap® 300 QC System (NeoPath, Inc., Redmond, WA)

How it works. Conventionally prepared Pap smears are loaded into the scanning device on slide trays. A robotic mechanism presents the slide to the optical pathway. Accumulated information acquired during digitization and morphometric cell feature analysis is used to establish an evaluation score and to rank the likelihood of the object’s being abnormal. Based on the single object rankings, a slide classification score is established.
Slides are normalized for stain and fixation qualities to ensure robust classification and highest signal-to-noise ratio. The final classifier score is, thereby, theoretically independent of preparation variability. All of these functions are performed without human attendance or intervention. The scanner can process eight slides per hour virtually round the clock, except for intermittent maintenance and slide reloading.

Performance data. Performance data from clinical trials, sponsored by Neopath, Inc., were included in the submission to the FDA for pre-market approval of AutoPap 300 QC as a rescreening device for quality assurance. Another study, performed independently of the manufacturer, has also evaluated the performance of the device. Stevens et al. recently assessed the ability of the AutoPap 300 QC system to identify false-negative smears that were previously screened as normal. The prevalence of disease in the population was very low: Seven cases were identified among 1840 subjects screened (<1%). With the device threshold between normal and abnormal set at 20% and 30% (i.e., 80% and 70% of the smears, respectively, would most likely be normal), three of seven (20% threshold) and four of seven (30% threshold) of the smears originally diagnosed as normal were reclassified as ASCUS. By contrast, randomly selecting 30% of the previous normal smears for human review, only one of seven diagnoses was changed to ASCUS. Neither method detected abnormalities at 10% threshold levels. While this difference may seem impressive, the cases were not selected in the same way. AutoPap 300 QC scanned all the slides before selecting the cases most likely to contain abnormal cells, a selection based on computer-detected abnormal cell features. The human rescreen was a random selection of only a percentage of the cases. To make a nearly equivalent comparison, 100% of the slides should have been rescreened by the cytopathologist. But the comparison would still not be equivalent, since there would not have been any computer-based assistance to detect subtle morphometric feature differences. Furthermore, there are basic differences in human and machine performance. Specifically, humans tend to miss the isolated, small HSIL cells that are so accurately identified by the imaging devices, whereas computer scanners still have difficulties with the overlapping groups of cells that human observers can easily distinguish.

As noted above, the point at which the threshold between normal and abnormal is set is especially critical when an automated system performs as a primary screener. If the threshold in the population to be screened does not coincide with that preset on the device, some cases with cytologic evidence of disease will be included in the group not subjected to human view. Lee et al. described the system’s performance on low prevalence and small cell abnormalities. Dependent upon where the device threshold was set, a percentage of HSIL slides remained in the normal category. However, in a normal population of women to be screened, the percentage of abnormal smears is predictably low, allowing threshold levels to be set safely below the level of disease prevalence.

During its submission to the FDA (January 28, 1998) for pre-market approval of the AutoPap system for use in primary screening, NeoPath, Inc., presented data from a clinical trial that compared the device with conventional screening of Pap smears by cytopathologists in a routine practice. Wilbur et al. conducted a two-arm prospective study at five commercial cytology laboratories. Using strict inclusion and exclusion criteria, they processed 25124 slides first in the conventional manner, including a randomly selected 10% of the normal slides for quality-control rescreening. They then screened the same slides.
PapNet® (Neuromedical Systems, Inc., Suffern, NY) screening system is not yet available. 75% of the slides will still require conventional screening with Pap system for primary screening, stipulating that 25% of the diagnosis. The FDA panel recommended approval of the AutoPap system for primary screening, stipulating that 25% of the slides will still require conventional screening with the assistance and added cost of AutoPap. Final FDA approval is still pending. Cost–benefit analysis of the AutoPap primary screening system is not yet available.

**How it works.** PapNet differs from AutoPap 300 QC in both its intended use and the timing and extent of human intervention. By locating cells that the human screener tends to miss because of their small size, infrequency, or location under cellular and noncellular debris, the PapNet system presents to the cytotecnologist a video display of an enriched sample. Human review of a nonranked collage of 128 digitized cell images (64 images on each of two video screen “pages”) determines whether the slide will ever be seen again. The device does not make the decision as to whether or not the slide contains only normal cells.

Another major difference between AutoPap 300 QC and PapNet is the location of the scanning station. Whereas NeoPath, Inc., sells or leases its instruments to individual laboratories, Neuromedical Systems, Inc., requires that the cytology slides be shipped to one of their scanning stations (Fig. 1). Once there, conventional Pap smear slides are loaded into 100-slide cassettes adjacent to the computerized microscope. A robot arm transfers a slide to the microscope stage where it is scanned at three magnifications. Cellular material is located at ×50 magnification, and potentially abnormal cells or groups are then mapped at ×200 magnification. Next, two independent neural networks process and numerically score the cells on the basis of the degree of abnormality. The neural networks that are an integral part of the software design are able to recognize abnormal cells regardless of shape, staining quality, or cellular overlap. A neural network makes decisions much like the human brain, by “experience” gained from integrating information gathered from many previous decisions which were made on slides and were then classified as either right or wrong by the human designing the system. The neural network’s interconnections which resulted in “correct” answers are retained in the network while those which led to “incorrect” answers are discarded. Based on the ranking made at a ×200 magnification, the 128 areas on the slide automatically determined to be the most abnormal are then scanned at a ×400 magnification. This selection process results in two video display pages consisting of 64 single-cell images on the first page and 64 cell clusters on the second. These two pages containing a total of 128 digitized images are then captured on a digital tape or CD-ROM and constitute the “enriched” specimen, which is sent back to the referring laboratory with the original glass slide.

At the referral laboratory, the tape or disc is loaded into the appropriate player and the images are projected onto a video screen. Both screen and player are supplied by Neuromedical Systems, Inc., as part of the contractual agreement between the company and the laboratory, which includes training for those who will be responsible for triaging the slides on the basis of the images. The person reviewing the images, usually a cytotecnologist, can display the image “tiles” in groups of 64, 16, or 4, each at successively higher magnification with some progressive degradation in image fidelity. Depending on what is seen in these two-dimensional images, the human viewer decides whether the slide is normal or in need of human review. The glass slides are subjected to conventional microscopic review, utilizing the computer-captured coordinates to relocate the cells selected by the cytotechnologist at the viewing station. Denaro et al. (43) have written an in-depth description of the system’s specifications and capabilities.

**Performance data.** Data from a clinical trial involving 10 laboratories and a total of 10,000 smears were included in the submission that Neuromedical Systems, Inc., presented to the FDA for pre-market approval. The study consisted of an index group of patients who had been diagnosed with invasive cervical cancer and whose screening Pap smears in the preceding 2 years were originally diagnosed as negative by the conventional method. As controls, 20 consecutive negative smears that came after each index patient’s negative smear were also scanned by the PapNet system. Koss et al. (44) found that diagnoses for 37 (16%) of the 228 smears that had been scanned and triaged by the PapNet system were revised (15 LSIL, 20 HSIL, and two invasive cancers). Of the 9666 negative controls, 127 diagnoses were revised to 98 LSIL and 29 HSIL, a false-negative rate of 1.3%. The results are convincing evidence of the system’s greater efficacy in detecting potentially invasive cancers at earlier stages than had been accomplished by the traditional method.

Three of the university laboratories involved in the above clinical trial had previously documented their false-negative detection yield, defined as the percentage of rescreened negative slides classified as abnormal. Mango and Valiente (45) compared a subset of controls from the trial (2293 negative smears) with the results of conventional screening (13761 negative smears). The controls that had been scanned by the PapNet system had a false-negative yield of 6.2% (142 of 2293) versus 0.6% (82 of 13761) (P<.001 with use of the chi-squared test) for the routine Pap smear.

In another study, Jenny et al. (46) subjected a set of 516 abnormal smears mixed with an equal number of normal smears to double scanning with the PapNet system and to independent review by two cytotecnologists in a double-blind trial. Results...
were impressive: The original laboratory false-negative rate was 5.6% compared with 0.4% and 0.8% for the two cytotechnologists involved in the study. Sensitivity increased from 94.4% for malignant and premalignant lesions to greater than 99%. The throughput of slides was increased, with an estimated reduction in screening time from 5 to 6 minutes per negative slide to about 1 to 2 minutes. However, although there was always sufficient evidence to classify a sample as abnormal, the most abnormal cells were not always displayed. This finding suggests that, regardless of the grade of the abnormality or the quantity of the abnormal cells seen on the video monitor, the final classification must be based on human review of the original slide.

The PapNet system was designed to enrich the sample presented to the human eye–brain axis and not as a primary screening device. Early clinical trials (44,46–48) provide convincing evidence that PapNet performs as expected. However, one of the difficulties in assessing the relevance of these clinical trials is the necessity to increase the percentage of abnormal cases required to determine the sensitivity of the system in a relatively short period. The prevalence of serious cervical disease in a normal population is much lower and thus more problematic to detect. Therefore, more clinical data are needed to assess the effectiveness of the system in an ordinary laboratory setting.

In a recent study conducted by the Victorian Cytology Service in Australia without sponsorship from Neuromedical Systems, Inc., 195 abnormal slides were included in a case load of 20,000 slides previously determined to be normal by two manual screens (49). This prevalence of 1% disease is virtually equivalent to that of real life. The 195 slides had originally been diagnosed as normal; however, upon manual rescreening, they were found to contain abnormal cells. The abnormal slides were subdivided into three groups: Group 1 included 67 slides from patients whose lesions were found on histologic biopsy to be carcinoma in situ (CIS) and, on review, contained cytologic evidence of HSIL; group 2 included slides that, upon quality-control rescreening, were found to have cells reflective of HSIL and subsequently confirmed by cervical biopsy as HSIL; and group 3 included 95 slides that, on quality-control review, were diagnosed as containing cells from LSIL and later histologically or cytologically were confirmed to be LSIL. The PapNet-assisted detection rate for these three groups of abnormal smears were 46% (CIS), 61% (HSIL), and 37% (LSIL). Repeated assessments of PapNet tiles increased the detection rate. The authors concluded that the estimated sensitivity of 44% that they found for PapNet-assisted detection of abnormal slides that had initially been missed by traditional screening performed by human contrasts unfavorably with detection rate estimates of 83% (50) to 97% (47) for human rescreening.

Ashfaq et al. (50,51) have examined the effectiveness of PapNet, as both a prescreening and a rescreening device. In one study (50), they prospectively scanned and triaged 5,170 consecutive cervicovaginal smears by the PapNet system and then manually screened all of the slides without knowledge of the 128 areas selected for the PapNet tiles. Abnormal cases identified by either method were reviewed by a cytopathologist who was blinded with regard to the content of the specimens. Diagnostic concordance was achieved in 84% of cases (3,167 normal, 1,038 abnormal, and 135 unsatisfactory). Diagnostic sensitivity of the PapNet system was 82% (compared with 77% for smears analyzed by humans without PapNet assistance), specificity was 85%, predictive value of a positive test was 66%, and predictive value of a negative test was 93.4% (P = .0001, with the use of McNemar’s test on paired comparison of false-negatives). The authors estimated that a single case can be reviewed at the PapNet viewing station in 2–5 minutes, depending on the complexity of the case. One cytotechnologist can triage a batch of 50 cases in 1.5–3 hours.

Cost-effectiveness. Cost efficiency will depend on the number of cases sent forward for routine slide review. Ashfaq et al. (51) also analyzed the usefulness of the PapNet system for rescreening slides in a large prospective study; they concluded that the system afforded no benefit over the manual method (five of 2238 missed cases detected by PapNet; six of 2000 missed cases detected manually). In another study (52), they focused on 861 high-risk patients and found that PapNet rescreening of negative smears detected only one LSIL at an estimated cost of $34,500. On the basis of their own experience with PapNet, Ashfaq et al. (52) concluded that its use as a prescreening tool could effectively reduce the false-negative rate, reduce the number of manual rescreens, and reduce the workload. However, they pointed out that both human and machine errors are additive, and quality-assurance procedures must include both components of the screening process.

AutoCyte SCREEN® (AutoCyte, Inc., Elon College, NC)

How it works. AutoCyte, Inc. (formerly Roche Image Analysis Systems, Inc.), has recently presented to the FDA their submission for pre-market approval of the company’s AutoCyte PREP machine, which is capable of simultaneously processing 48 cell suspension samples. Once that approval has been granted, the company will next seek approval for its interactive screening system, AutoCyte SCREEN, which combines many of the design features of the AutoPap 300 QC and PapNet systems.

Like AutoPap 300 QC, it is based on algorithmic classifiers, but it also presents screens of cell images to the human for selection of the most abnormal cells on the slide, as does PapNet. The system differs from PapNet by displaying six very large low-power overview fields in addition to 120 of the most abnormal cell images. In addition to the digitized cell images, AutoCyte SCREEN derives a cell population histogram for the entire slide based on computer evaluation of certain cellular features. Ten cassettes, each containing 40 thin-layer slides, can be loaded onto the system’s carousel; the slides are then transferred by robot arm to the scanning stage. AutoCyte SCREEN can process up to 300 specimens in a 24-hour period, a rate of less than 5 minutes per slide (25).

Performance data. Preliminary performance data, resulting from a study by Takahashi et al. (53) of 583 cervical cellular samples processed by the AutoCyte PREP method, indicated a false-positive rate (using a threshold of LSIL or higher) of 20.8% and a false-negative rate of 1.8%. Among those cases incorrectly diagnosed with the use of the system were five LSILs, one HSIL, two squamous carcinomas, and one adenocarcinoma.

Measuring Outcomes

The currently FDA-approved systems would add cost to the processing and analysis of any sample, adversely affecting ac-
cess by patients to a widely utilized screening test. Unless use of the devices can be shown to increase throughput, reduce costs, or significantly increase accuracy, their commercial survival is questionable. Hutchinson (54) has analyzed various quality-control rescreening options, including none, 10% of negatives, targeted (high-risk patients), 100% rapid review, 10% selected by AutoPap, and 100% scan by PapNet followed with triage by a qualified cytotechnologist. Using economic assumptions based on her own laboratory’s experience and data from the literature, she cost-compared the automated systems with the various human approaches. On the basis of appropriate statistical analyses, Hutchinson concluded that the most expensive rescreening method is the PapNet method and that rapid human rescreening of all negative slides is the most economical.

Kaminsky et al. (55) have designed a mathematical model for laboratory decision-makers that describes the overall sensitivity, specificity, and cost of the screening process. They conclude that, if either automated or human screening processes improve the detection of the higher grade abnormalities, then the impact of false-negatives on the overall costs of health care will be reduced.

Schechter (56) determined the relative cost per life-year saved (in 1995 U.S. dollars) using PapNet and compared this cost to the costs associated with the use of other common screening and diagnostic interventions, such as mammography ($67,918) for breast cancer and serum prostate-specific antigen ($113,000) for prostate cancer screening. The costs of biennial PapNet testing and biennial routine Pap smear screening were calculated to be $48,474 and $98,880, respectively. Schechter argued that the increased detection of abnormalities by PapNet would allow for increasing the intervals between screening that would be an alteration in the current standard of care and might thereby decrease overall costs.

Medley [reviewed in (57)] estimated the increased cost of including some of the new technologies in screening programs in the Australian state of Victoria. In 1994, 240 new cases of cervical cancer were reported in Victoria, and 80 women died. Only 10% of the 80 women had been screened by at least three Pap smears in the preceding 10 years. Of the women who died, 56% had never had a Pap smear. Therefore, only eight deaths (10%) could be considered screening failures. The number of Pap smears performed each year in Victoria is approximately 600,000. If costs for the new technologies are added to each routine test ($20 for the ThinPrep and $30 for PapNet), the annual cost of screening the women in Victoria would be increased by $30 million—possibly save eight lives. Translating these results for screening all sexually active women in the United States, assuming 50 million Pap smears per year, 4800 annual deaths from cervical cancer, and a 10% Pap smear failure rate, the total additional cost would be $2.5 billion—possibly detect 480 false-negative smears.

In an independent investigation, O’Leary et al. (58) re-screened 5478 Pap smears using PapNet to determine the effectiveness and cost of the device in detecting cellular abnormalities not identified in routine screening. Although 1614 (29%) of the slides were triaged for human review, only 448 (8% of the total) were found to have possibly abnormal cells. After review by a consensus panel, only six of 11 were reclassified as being equivocally abnormal (ASCUS category). No additional intraepithelial lesions were identified on the 11 smears, but one of the six patients whose smear was reclassified was found to have an LSIL on subsequent biopsy. A time savings of 2 minutes per case was realized for those slides not requiring further human review (71%). However, the calculated cost to detect one additional LSIL was $17,475 based on the contractual cost with Neuromedical Systems, Inc., of $7.50 per slide. If the advertised cost of $40 per slide is used, the cost to detect one additional LSIL is $101,345.

The issue of cost of false-positives rarely is factored into the health care equation. In an examination of the most popular cancer screening tests, Russell (59) warned that false-positive test results occur much more often than a relatively rare disease such as cervical cancer. She estimated that, for women who are screened regularly, increasing the frequency of testing from every 2 years to annually saves an additional year of life at the cost of more than $1 million and an uncounted number of false-positives. In addition to the cost of screening, false-positive tests require unnecessary follow-up studies and increase the patient’s anxiety and distrust of the test.

What is now needed are real life experiences with the automated systems in various clinical settings, with carefully structured studies so that comparisons are valid. In 1996, the Inter-society Working Group for Cytology Technologies was formed. The purpose of the group is to monitor the dissemination of information relating to novel approaches to cytology screening and diagnosis and to provide guidelines for federal agencies, especially the FDA, as well as for the manufacturers, regarding the proper design and conduct of clinical trials (60). The members’ greatest concern is that the profession continue to determine the standard of practice and not be unduly influenced by device manufacturers or by the threat of litigation. Recently, there has been great emphasis on public disclosure of any industrial affiliation by scientists involved in clinical trials being conducted to evaluate the performance of the screening devices irrespective of whether or not the trial itself has industry sponsors. Such disclosure should be encouraged so that objective appraisal of study results is possible.

Laboratory directors need to determine what improvements, if any, these technologies will make in their laboratories. In the case of the populations who are unscreened and for whom there are insufficient health care resources, such as in South Africa (48), the automatic scanners may provide substantial improvement over the current situation. By contrast, the new technology may offer little benefit to laboratories employing highly competent and experienced cytotechnologists who work under conditions of limited workloads and stringent adherence to quality control and assurance (41).

Conclusion

All of the innovative technologies described above are automated modifications of traditional cytologic methods. New technologies capitalizing on molecular biologic discoveries, such as HPV DNA testing (23), or reinventions, such as laser scanning microscopy (61) or spectral classification (62) applied to Pap smears, may solve the diagnostic problems that confront us every day. An ongoing clinical trial sponsored by the National Cancer Institute (contract No. 55159) is examining conventional and innovative testing of women with low-grade neoplasia of the...
utere cervix. The ASCUS/LSIL Triage Study will enroll 7200 women over the next 2 years to determine if HPV testing is an efficient and effective adjunct to cytologic analysis when deciding management of the patient with an abnormal Pap test. Liquid-based cytologic samples will be used in place of the standard Pap smear in this trial after comparison studies on the first 1000 women show at least equivalence, if not improvement, of the ThinPrep sample over the traditional Pap smear in detecting cytologic evidence of cervical neoplasia. One of the FDA-approved scanning devices, AutoPap 300 QC, has been included in the trial as an additional quality-control measure. Results will be analyzed during the year 2000 and may influence the practice of gynecologic cytopathology in the next millennium.

Those of us who have invested many years of our scientific careers in developing automated systems for cellular analysis are fearful that the public’s high expectations, fed by manufacturers’ claims for performance of these devices, will be met with disappointing results. The technology is exciting and, if given the time to develop in tandem with standard good laboratory practice, i.e., parallel studies in routine settings, then the most effective components of these systems will prevail. Cooperation among pathologists, clinicians, and manufacturers will ensure that the technology performs as expected and contributes to affordable and reliable patient care.

References

(8) Pellehach L. Appraising the Pap smear: will society, insurers, put their money where the value is? Lab Med 1997;28:440–9.


(57) Check W. Too early to solve Pap device puzzle. CAP Today 1997;11:1,4,46,48–9,52,54.


Notes

The author has been a practicing cytopathologist since 1970 and has worked in both private and academic sectors. She was actively involved in research conducted from 1978 through 1990 that contributed to the development of the automated scanning devices. She served as an unpaid consultant to Neuromedical Systems, Inc., from 1989 through 1991. Since 1988, she has promoted the refinement and acceptance of these evolving technologies by chairing and/or participating in conferences showcasing their progress and problems interfering with their success. She served as chair of the cytopathology subcommittee for the inaugural Clinical Laboratory Improvements Advisory Committee from 1992 through 1994 and currently represents the International Academy of Cytology on the Intersociety Working Group for Cytology Technologies. The analysis of the current state of the art is her own and has been influenced only by her experiences and not by any industrial affiliations.

The author wishes to thank Patricia Stephens, Ph.D., independent author’s editor, for her major help in focusing the paper. The secretarial skills of Shawnice M. Foster added greatly to the completion of this manuscript. Manuscript received August 27, 1997; revised March 9, 1998; accepted March 31, 1998.

Journal of the National Cancer Institute, Vol. 90, No. 10, May 20, 1998 COMMENTARY 749