Informatics / Web-Based Automated Gynecologic Screening

Internet-Based Gynecologic Teleytology With Remote Automated Image Selection

Results of a First-Phase Developmental Trial

John H. Eichhorn, MD,1,2 Leigh Buckner, CT(ASCP),3 Sally Beth Buckner, CT(ASCP),4 David P. Beech, CT(ASCP),1 Kimberly A. Harris, MM,5 Douglas J. McClure, MIM,5 Barbara A. Crothers, DO,3 and David C. Wilbur, MD1,2

Key Words: Cervical cytology; Automated screening; Internet; Web-based software; Telepathology; Telecytology

DOI: 10.1309/GRAV16QP8JR5XTPF

Abstract

A retrospective set of 191 gynecologic cytology slides with reference interpretations was run on an automated screening device that selects fields of view (FOVs) based on a hierarchical probability of abnormality being present. An interface was developed between the device and a remote server using customized image review software. FOVs were reviewed by 3 cytotechnologists and 3 cytopathologists, and binary triage (unsatisfactory for evaluation/negative for intraepithelial lesion or malignancy [NILM] vs “abnormal” [neither unsatisfactory nor NILM]) and specific interpretations were done. No morphologic training before FOV review was provided. Three or more reviewers agreed on the correct categorization of NILM/unsatisfactory in 89% (85/96) and abnormal in 83% (79/95). Three or more reviewers triaged cases to abnormal as follows: atypical squamous cells of uncertain significance, 83% (5/6); atypical squamous cells, cannot exclude high-grade lesion, 100% (3/3); low-grade squamous intraepithelial lesion (SIL), 83% (52/63); high-grade SIL, 94% (17/18); and atypical glandular cells, 40% (2/5).

This procedure may have comparable sensitivity and specificity and possibly could provide effective initial triage to further evaluation. A review of individual cases suggests that further accuracy can be achieved with additional training and experience.

We previously hypothesized that the melding of 2 technologies, the Internet and computerized automated scanning, could lead to the provision of effective cervical cancer screening programs for countries lacking a trained cytology workforce and to an efficient, economical, and centralized way of triaging specimen data. In a pilot study, we showed that satisfactory and reliable interpretations of device-selected, electronically transmitted images can be made on a computer monitor, without manual microscopy.1

The instrument used in that study, the FocalPoint screening system for cervical cytology (Becton Dickinson/TriPath Imaging, Burlington, NC), is used in many laboratories to triage cases to no further review or complete, manual review. Studies have shown that the use of this automated device allows cytologists to efficiently and reliably identify more abnormal cases than manual screening alone.2-4 This device also has 3 features that make possible the Web-based evaluation of device-selected images: (1) Up to 30 fields of view (FOVs) are selected using a proprietary algorithm based on a probability of that FOV containing cellular abnormality.4 (2) FOVs are arranged in a hierarchical rank order based on this probability. (3) The FOV images are black and white and low resolution, so that in JPEG compressed format, they are suitable for rapid digital transmission using Internet conduits of modest bandwidth. This image-capture capability had initially been engineered into the FocalPoint device to allow for the accurate localization of FOVs to observers during subsequent manual microscopic review, and, therefore, the images are linked to slide location Image 1.

In our previous study,1 the FOV images were transmitted as e-mail attachments, which made their display and manipulation cumbersome and inefficient. The image attachments for
Informatics / Original Article

each case had to be bundled and sent to the specific electronic address of the reviewer. In any clinical application, the resulting diagnoses would have had to be further delivered to other specified addresses for hierarchical review and/or consultation. The present study was designed to ameliorate those limitations by using a secure Web site and software that permits a variety of image display options.

The software application was custom designed for the efficient display and manipulation of FOV images, the retrieval of demographic and device-generated information, reporting capabilities, and data storage. Demographic data and device-captured images from individual cases were uploaded to a remote server, queued, and made available to assigned persons for review. A secure Web site displays the interpretations and images to additional individuals with access for further processing or reporting. Such a system allows for the following: (1) distribution of case images to people with diagnostic expertise, who may be in remote locations; (2) less transportation of perishable materials; (3) timely, encrypted transmittal of device-generated images and interpretations; (4) communication of results to patients, caregivers, or both, who may be in a remote location or in transit; and (5) efficient, economical triage of patients to additional clinical evaluation, testing, treatment, or any combination thereof.

The present study augments the scope of our prior study with a much larger population of cases, use of SurePath (Becton Dickinson/TriPath) and ThinPrep (Cytyc, Marlborough, MA) liquid-based samples, more diagnosticians, and more reading locations. Our goal was to test the reliability of Web-based image review for clinically meaningful interpretative groups. As in the pilot study, the goal was to distinguish, by means of primary triage, cases that require prompt intervention or further clinical evaluation from those that do not.

Materials and Methods

The study protocol was reviewed and approved by the institutional review boards at all investigational sites. The study population consisted of a set of liquid-based SurePath and ThinPrep cervical cytology slides from reference cases with known diagnoses from the files of 2 participating institutions. Interpretations were made using the criteria of the 2001 Bethesda System.5

An interface was established between a FocalPoint GS gynecologic screening device (Becton Dickinson/TriPath) located in Redmond, WA, and a computer server located in Cambridge, MA, using custom-designed, Web-based image and data acquisition and review software (Consult, Wellogic, Cambridge, MA) that automatically queries the FocalPoint database at regular intervals during the screening process. The software was developed in consultation with the investigators to address the specific needs of the project. The software application is based on an existing one that has been certified...
as compliant with the Health Insurance Portability and Accountability Act of 1996 for use in patient care settings.

Web site–based image review and data management software were developed for the present study to overcome the limitations inherent in the use of e-mail transmission and case review. The custom-designed software used in this study incorporated the following capabilities: (1) entry of clinical and demographic data of individual cases and linkage to captured images; (2) upload of specimen-derived images that are device selected from an automated instrument at a remote site; (3) encryption of transmitted data and images; (4) secure storage of case-specific data and images; (5) orderly presentation of clinical data and images on a computer monitor; (6) access at the Web site to individual cases or to lists of cases, by registered cytotechnologists, cytopathologists, clinicians, and patients; (7) presentation of thumbnail FOV images in rank order Image 2 and Image 3; (8) full-screen presentation of any 1 image, panels of 4 or 9 enlarged sequential images (with

**Image 2** Cases were accessed on a Web site with retrieval of images that had been uploaded to a remote server from the automated screening device; 30 JPEG images with the highest probability of abnormality are displayed in rank order (left), with an adjacent scroll tab. A panel of 4 selected highlighted images (fields of view 9-12) is enlarged (right). LSIL. ThinPrep slide.

**Image 3** Shown are 30 JPEG images in rank order according to probability of abnormality (left), enlarged view of JPEG image 15 (right), and drop-down view of specimen details (top), including distribution quintile and adequacy. High-grade squamous intraepithelial lesion involving glands. ThinPrep slide.
selection of the “starting” image) (Image 2), or of 4 enlarged nonsequential images (with point-and-click selection from thumbnail images); (9) presentation of clinical (age, reproductive, and menstrual status) and device-generated (distribution quintile, computer-assessment of adequacy) data from a dropdown folder tab on the monitor (Image 3); (10) recording of each interpretation on every case by a cytotechnologist, cytopathologist, and (if needed) consultant; (11) electronic communication between any or all people responsible for any individual case; and (12) presentation of the finalized evaluations of selected cases to clinicians, patients, or both.

Demographic information for each slide was entered manually into the slide review database. The slides were tagged with bar codes, stripped of other identifiers, and scanned at the remote site on the FocalPoint GS system. A set of up to 30 of the highest scoring FOV images were captured in a JPEG compressed format (image size, 12-16 KB each). The FocalPoint GS system captures only black-and-white images, each of which corresponds closely to an approximately ×200 microscopic magnification when viewed on a monitor. Of note, the image captured by the device represents only the central area of a ×200 magnification FOV that, in the commercial application of the FocalPoint GS device, is intended to be viewed as a full circular microscopic FOV. Hence, each image used in the study does not contain the entire FOV intended to be viewed in the normal operation of the device. Sets of JPEG images for each case were transmitted over the Internet and uploaded to the remote server. For each case, 30 JPEG images were device generated and transmitted unless fewer FOVs were selected by the scanner, as occurred in 22 cases of low cellularity (for which 5-29 FOVs [mean, 22 FOVs] were selected). To capture maximal information from the abbreviated FOVs, 30 images were selected for review; in the intended use of the device, only the top 10 scoring FOVs are used. The maximum file size per case was less than 500 KB. Known demographic information (eg, age and menstrual status) and standard FocalPoint device-generated data (distribution quintile, computer assessment of adequacy) were entered into the database for each case.

The review stations were the standard personal computers (Microsoft Windows 2000 Professional operating system, Microsoft Explorer browser, Microsoft, Redmond, WA) used by the reviewers in their secure institutional networks, augmented by high-resolution 19-inch LCD monitors to enhance the clarity of images. The monitor settings and image display options were standardized across institutions. The study did not seem to be affected by network delays at either institution. The software application displayed up to 30 thumbnail images for each case in a panel on the left side of the monitor (Images 2 and 3). Each individual image could be displayed as a single large image or as a panel of 4 or 9 consecutive images at less than maximal enlargement, depending on the preference of the observer. In addition, for image comparison purposes, any 4 selected nonconsecutive FOVs could be displayed together as a panel of enlarged images. Demographic and FocalPoint data for each case were accessed by using a drop-down tab. Interpretations for each case could be entered in a results field after image review and forwarded to another individual for final review (“escalate” tab), or submitted as a final diagnosis (“complete” tab).

Glass slides were not manually reviewed by any participant during the study, and the reviewers received no training in the evaluation of the black-and-white FOV images. Each of the study cases was interpreted independently by 3 ASCP-registered cytotecnologists (D.P.B., L.B., and S.B.B.) and 3 board-certified cytopathologists (B.A.C., J.H.E., and D.C.W.) who were masked to the reference diagnoses. During the study, an off-line worksheet was used that allowed the recording of participant observations in a format intended as an aid to future data analysis and further technique refinement. Reviewers annotated the reasons that a given interpretation was favored and features of each case that made evaluation difficult. For every case, they also recorded the first FOV for which an abnormality was suspected, the first FOV at which the final interpretation became fixed, and the specific FOVs that contained abnormalities or in which other observations were made. Assessments regarding adequacy and any other comments about each case were recorded.

Cases were interpreted as specifically as possible by the reviewers and then assigned to 1 of 2 categories: (1) negative for intraepithelial lesion or malignancy (NILM; normal findings, reactive changes, organisms) or unsatisfactory for evaluation; or (2) any cellular abnormality diagnostic of an intraepithelial lesion (or worse) or worrisome for its presence (high-grade squamous intraepithelial lesion [HSIL]; low-grade SIL [LSIL]; atypical squamous cells, cannot exclude high-grade lesion [ASC-H]; atypical squamous cells of uncertain significance [ASC-US]; or atypical glandular cells [AGC]). It was intended that this binary categorization would determine if the combination of automated screening and Web-based transmittal and presentation of images could be used to triage patients appropriately, with the cutoff point representing clinically significant lesions requiring further evaluation.

After reviewers had recorded their initial binary assessments for each case, their “trial” interpretations were compared with the reference diagnoses. All 6 reviewers met at 1 location to review the images of selected cases for which there were 1 or more discrepant interpretations, with simultaneous access to the images of these cases (from the Web site) and written observations that they had made at the time of masked review recorded in a computerized database. The potential sources of error were discussed at this time. The slides of cases for which all, or nearly all, reviewers incorrectly categorized the binary triage compared with the reference diagnosis were screened.
manually by the investigators to confirm or change a reference diagnosis for data analysis purposes.

A comparison was done of reviewer categorization (binary triage) vs the reference diagnosis for each case and tabulated based on how many reviewers achieved the correct result. The numbers of reviewers correctly triaging the cases into these 2 categories were also tabulated for each Bethesda diagnostic group for the reference cases. In addition, rates of sensitivity [True Positive/(True Positive + False Negative)] and specificity [True Negative/(True Negative + False Positive)] were calculated at the level of agreement by 3 or more reviewers because it was postulated that this degree of concordance indicated that a correct interpretation could have been made in each case.

Results

The study set included 101 SurePath and 90 ThinPrep slides. We excluded 11 other ThinPrep slides from the study because they could not be adequately scanned by the FocalPoint device. In these cases, it seemed that scanning was impeded by technical problems such as coverslip placement, artifacts of preparation, and staining irregularities. Reference diagnoses included the following: HSIL, 18; LSIL, 63; ASC-H, 3; ASC-US, 6; AGC, 5; NILM, 93; and unsatisfactory for evaluation, 3 Table 1.

Images 2 and 3 and Image 4 include representative device-captured FOV images from ThinPrep slides, and Image 5, Image 6, and Image 7 show images from SurePath slides. The images derived from ThinPrep slides tended to have fewer cells for evaluation in each FOV than those produced from SurePath slides, and the cells in the images from ThinPrep slides were less evenly spread across the FOV. These findings directly relate to the lower cell concentration and increased “open” areas routinely present on each ThinPrep slide compared with the higher and more evenly distributed cellularity of a routine SurePath slide. SurePath slides, on the other hand, showed fewer well-focused cells and poorer overall image resolution, especially in hypercellular

Table 1

<table>
<thead>
<tr>
<th>Bethesda Category</th>
<th>SurePath (n = 101)*</th>
<th>ThinPrep (n = 90)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSIL</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>LSIL</td>
<td>45</td>
<td>18</td>
</tr>
<tr>
<td>ASC-H</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>ASC-US</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>AGC</td>
<td>0</td>
<td>5†</td>
</tr>
<tr>
<td>NILM</td>
<td>48</td>
<td>45</td>
</tr>
<tr>
<td>Unsatisfactory for evaluation</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

AGC, atypical glandular cells; ASC-H, atypical squamous cells, cannot exclude high-grade lesion; ASC-US, atypical squamous cells of undetermined significance; HSIL, high-grade intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignancy.

* One slide per case.
† The slides from these cases were rescreened after the trial; 3 of 5 were diagnosed as NILM by majorities of reviewers, and these patients had subsequent surgical procedures that were negative.

Glandular cell atypia was difficult to assess on low-resolution JPEG images. Adenocarcinoma in situ. ThinPrep slide.

Case judged negative for intraepithelial lesion or malignancy with abundant normal endocervical cells. The 3 reviewers not accustomed to screening SurePath slides called this case atypical.
samples. This finding could be related to the characteristic 3-dimensional depth of routine SurePath slides, which might have led to more variability in the device-determined plane of focus.

The reviewers’ ability to correctly assign the cases to either of the 2 categories, NILM/unsatisfactory or abnormal, is shown in Table 2. Three or more investigators agreed on the correct general categorization of NILM/unsatisfactory in 89% of cases (85/96) and abnormal in 83% of cases (79/95). This corresponds to a sensitivity of 88% and a specificity of 93% for binary categorization at the level of agreement by at least 3 reviewers. The ability of the reviewers to triage individual cases to the abnormal group and target a specific Bethesda category of reference diagnosis is shown in Table 3. For specific Bethesda interpretations, the frequencies at which 3 or more reviewers triaged the cases to abnormal were as follows: ASC-US, 83% (5/6); ASC-H, 100% (3/3); LSIL, 83% (52/63); HSIL, 94% (17/18); and AGC, 40% (2/5). That is, the reviewers’ ability to triage to the abnormal group the cases with possible high-grade squamous lesions (HSIL or ASC-H) was better than their ability to assign low-grade squamous lesions (LSIL) to that category. A majority of the reviewers incorrectly categorized 7 (16%) of 45 SurePath LSILs and 4 (22%) of 18 ThinPrep LSILs, whereas a majority incorrectly assigned only 1 (6%) of 18 HSILs. These observations are consistent with the known sensitivity data of the FocalPoint screening device.4,6

The data in Table 3 suggest that the system did not perform as well for glandular as for squamous cell abnormalities, but this apparent disparity was all but eliminated when the individual cases were examined more closely. When the slides (all ThinPrep) of the 5 cases of AGC were reexamined by the reviewers as a group, 3 cases were judged to have an incorrect reference diagnosis by majorities of the participants.

Table 2

<table>
<thead>
<tr>
<th>General Category of Reference Cases</th>
<th>No. of Reviewers With Correct General Categorization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>NILM/unsatisfactory for evaluation</td>
<td>35</td>
</tr>
<tr>
<td>Abnormal</td>
<td>32</td>
</tr>
</tbody>
</table>

NILM, negative for intraepithelial lesion or malignancy.

* Data are given as number of cases.

3 Image 6 | High-grade squamous intraepithelial lesion cell (center right) and benign cells for comparison. Hypercellular fields of view sometimes had poorer resolution of the nuclei of atypical cells. SurePath slide.

4 Image 7 | Abnormal cells and cell clusters are not always positioned in the center of the selected field of view (FOV) owing to the full-field interpretation algorithmic method used in the FocalPoint system. Abnormal cells can be localized to any portion of the selected FOV. Low-grade squamous intraepithelial lesion. SurePath slide.
and to have shown no more than reactive glandular changes. In each of these 3 cases, no glandular lesion was identified after subsequent surgery (cone biopsy and endocervical and endometrial curettage, 2 cases; hysterectomy, 1 case). On the initial FOV review, 2 of these cases had been assigned to the NILM/unsatisfactory group by 5 reviewers and the other case by 3 reviewers.

Of the remaining 2 AGC cases, 1 patient had endometrial adenocarcinoma found by subsequent hysterectomy. When the device-selected FOVs of this case were reexamined, however, no convincing atypical glandular cells were found. This case had been called NILM by 4 participants and ASC-US by 2. Finally, 1 patient with a reference diagnosis of AGC had endocervical adenocarcinoma in situ in a subsequent hysterectomy specimen. This case (Image 4) had been called abnormal by 5 of 6 participants after FOV review, but there was little agreement on specific categorization: 2 called it cancer and 1 each called it LSIL, ASC-US, or AGC. Thus, of only 2 true AGC cases in the study, one lacked abnormal cells in device-selected FOV images and was missed on this basis, and the other had abnormal cells in the FOVs and correct binary triage by 5 of 6 reviewers. Differences in the specific categories to which this latter case was assigned may be related to difficulties in resolving fine details in small glandular cells.

In 4 cases with a squamous abnormality (SurePath LSIL, 2; ThinPrep HSIL, 1; ThinPrep LSIL, 1), none of the 6 reviewers gave an interpretation of abnormal. Reexamination of the 30 device-selected FOV images of each of the 4 cases revealed no convincing squamous abnormalities. The cases screened from SurePath slides (both containing LSIL) showed rare (1 or 2) dense blotches on the FOVs; these may be koilocytotic nuclei, but there was insufficient detail to be sure. In the cases from ThinPrep slides, subsequent biopsies confirmed the presence of squamous dysplasia that was severe in one case and mild in the other. The ThinPrep HSIL case appeared to contain many small cells of parabasal or metaplastic type, and it is possible that similarly sized dysplastic cells may have been missed at the level of resolution of the FOV images. These 4 cases indicate that the abnormal cells in squamous lesions (particularly LSIL) may not be captured in the rectangular images captured by the automated device from the selected FOV.

The success of binary triage by the reviewers is separately shown by preparatory technique in Table 4 and Table 5 for the abnormal cases. There were differences in the composition of abnormal cases in these 2 study populations, how and where the reference sets were compiled, and the availability of follow-up data from subsequent procedures. Of 53 abnormal SurePath slides, 45 (85%) showed LSIL and none showed AGC; follow-up data were not readily available for some of these cases. There were fewer cases with ThinPrep slides (11 cases were initially eliminated from the study because their slides could not be adequately scanned by the FocalPoint device), but among the 42 abnormal cases in this set of slides, there was a broader distribution of Bethesda groups, including 18 LSIL (43%), 12 HSIL (29%), 5 AGC (12%), 4 ASC-US (10%), and 3 ASC-H (7%) cases. In this population, follow-up data from subsequent surgical procedures were available for all cases with abnormal cytology. Moreover, the reviewers differed in their experience with each of the 2 liquid-based methods: reviewers at one institution had experience with both types of slides but had been routinely using SurePath slides in clinical practice, whereas reviewers at the other institution had little or no recent experience with SurePath slides, having used the ThinPrep method exclusively in their practice.

As in our previous study,1 FOV images displayed on a computer monitor showed poorer resolution for a given degree of magnification than would have been achieved by standard light microscopy. Their plane of focus is set by the automated instrument and cannot be manipulated. These features and the lack of a full-color spectrum made the evaluation of fine or subtle nuclear and cytoplasmic detail more challenging than in light microscopy, particularly for the smaller cells within a
given image. Each cytologist reported spending approximately 1 to 2 minutes per case on average to review the images and make an interpretation, although no attempt was made to do a formal time analysis in this first-phase study.

Discussion

In the United States, the incidence and mortality rates of cervical cancer have decreased 50% and 70%, respectively, in the 50 years since widespread screening programs were instituted. Other countries have also benefited greatly from population-based screening programs. Worldwide, however, cervical cancer continues to be a leading cause of death for young women who reside in countries without systematized screening programs. These regions lack an adequate pool of trained specialists to screen and interpret specimens. They typically lack funding for evolving new technologies such as liquid-based preparations, automated screening devices, and human papillomavirus testing. Even in the United States, laboratories experience fluctuating workloads out of synchrony with their available workforce, resulting in specimen backlogs.

Remote interpretation of automated device-selected digitized images transmitted over the Internet has the potential to provide effective screening and clinical triage to individuals in these populations. “Proof of concept” was provided by a recent feasibility study in which it was shown that low-resolution images selected by an automated screening device and transmitted over the Internet as e-mail attachments could be adequately interpreted at a distant site without manual microscopy, using primary triage methods for clinically meaningful diagnostic categories. An algorithm of remote review of a laboratory’s total cervical cytology workload with triage of the most potentially abnormal cases to on-site review by the originating laboratory has the potential to select the “needles out of the haystack” for expedited review. In addition, laboratories could more effectively share personnel resources across distances without resorting to the shipment of glass slides, thereby easily redistributing the workload among available resources.

The use of customized Web site–based data management software facilitated the accommodation of more reviewers and reviewing stations in the present study than there had been in the previous one, thereby permitting the evaluation of a much larger population of cases. Multiple investigators could evaluate each of the cases at their own particular location, in any order, at their own pace, and with as many “visits” to the

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Reviewer Triage to Abnormal vs Bethesda Category (SurePath Cases)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bethesda Category of Reference Cases</td>
<td>No. of Reviewers Triaging Cases to “Abnormal” Group</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>ASC-US</td>
<td>0</td>
</tr>
<tr>
<td>ASC-H</td>
<td>0</td>
</tr>
<tr>
<td>LSIL</td>
<td>0</td>
</tr>
<tr>
<td>HSIL</td>
<td>0</td>
</tr>
<tr>
<td>AGC</td>
<td>0</td>
</tr>
</tbody>
</table>

AGC, atypical glandular cells; ASC-H, atypical squamous cells, cannot exclude high-grade lesion; ASC-US, atypical squamous cells of undetermined significance; FOV, field of view; HSIL, high-grade intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

* Data are given as number of cases.

† Reexamination of FOVs from these cases revealed no convincing abnormality, indicating device failure of abnormality localization to selected FOVs.

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Reviewer Triage to Abnormal vs Bethesda Category (ThinPrep Cases)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bethesda Category of Reference Cases</td>
<td>No. of Reviewers Triaging Cases to “Abnormal” Group</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>ASC-US</td>
<td>0</td>
</tr>
<tr>
<td>ASC-H</td>
<td>0</td>
</tr>
<tr>
<td>LSIL</td>
<td>0</td>
</tr>
<tr>
<td>HSIL</td>
<td>0</td>
</tr>
<tr>
<td>AGC</td>
<td>0</td>
</tr>
</tbody>
</table>

AGC, atypical glandular cells; ASC-H, atypical squamous cells, cannot exclude high-grade lesion; ASC-US, atypical squamous cells of undetermined significance; FOV, field of view; HSIL, high-grade intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

* Data are given as number of cases.

† Reexamination of FOVs from these cases revealed no convincing abnormality, indicating device failure of abnormality localization to selected FOVs.

‡ The slides from these 3 cases were rescreened after the trial and were interpreted as negative for intraepithelial lesion or malignancy by majorities of the participants; subsequent surgery failed to detect a glandular abnormality.
cases as needed.

Although the FocalPoint system is approved by the US Food and Drug Administration (FDA) for automated scanning of SurePath slides, it can also be reconfigured for the assessment of ThinPrep slides, for which it has shown screening efficacy. It is FDA approved for the screening of conventional smears as well, but the present study was designed to evaluate slides made from liquid-based preparations. We included both liquid-based slide types in the present study, not to compare the desirability of one method over the other, but to show that both can be used in such a system and to judge the special features of each type of preparation when evaluated on the monitor-based FOV reviewing platform. This was necessary because we anticipated that one or the other type of liquid-based preparation may be favored or in use in the target population of our subsequent studies or their intended applications. Differences in the images derived from the 2 types of samples were seen. ThinPrep slide images tended to have fewer cells per FOV than those from SurePath slides, and the cells were distributed less uniformly across each FOV. The images produced from SurePath slides, however, had poorer image resolution, especially in cellular samples, generally because a thicker layering of cells in these preparations produced more variability in the device-determined plane of focus.

Even though image quality and size were augmented by the use of larger, flat-screen monitors with better resolution than the monitors used in our prior study, the resolution of the FOV images still was not as optimal as that of standard light microscopy, and the fixed plane did not allow manipulation of cells that were not optimally focused when captured. In the present study, these limitations may have hindered the appreciation of fine cellular details in glandular cells, which tended to be small and to present in closely packed groups (Image 4), which led to difficulty in classifying a case of adenocarcinoma in situ (although it was recognized as abnormal by 5 of 6 reviewers). Conversely, 3 investigators working at hospitals that screen only ThinPrep slides tended to overcall glandular atypia from images derived from SurePath slides that had abundant normal endocervical cells (Image 5).

Despite a poorer focus and resolution of images derived from some SurePath slides that were hypercellular (because of a thicker layer of cells per FOV), HSIL was correctly recognized despite the small size from cellular features unaffected by these limitations, such as nuclear/cytoplasmic ratio, nuclear shape, and degree of hyperchromasia, compared with neighboring cells in the FOV image (Image 6). The reviewers noted that LSIL cells were easier to recognize than HSIL cells because of their larger nuclear dimensions and the occasional presence of human papillomavirus–associated cytopathic perinuclear halos.

The most abnormal cells, however, were not always positioned near the center of the FOV (Image 7), nor were they always present in the FOVs that had been assigned the highest probability scores for having cellular abnormality. These observations relate to FocalPoint system programming because the device selects FOVs with the highest probabilities of cellular abnormality based on its assessment of the field “as a whole” and not of individual cells within the field or their position within it. Moreover, the computer algorithm used to assign these probabilities has an intended bias, based on training and prioritization, toward the detection of high-grade squamous lesions over the identification of other types of abnormality.

In follow-up group discussions, it seemed that some reviewers had been more bothered than others by the lack of a full-color spectrum, particularly in the evaluation of subtle differences in cytoplasmic characteristics of squamous metaplastic and high-grade squamous dysplastic cells, and that some investigators were more adept than others in making cellular and nuclear size comparisons from FOV images. Different reviewers developed their own preferred strategies of approach to individual cases using the display software, which probably evolved between review of early and later cases.

The reviewers’ success in binary triage of the cases was analyzed at the level of agreement of 3 or more individuals (half the total number of viewers). This level of agreement was selected for this first-phase trial because most of the reviewers had no experience in the evaluation of black-and-white FOV images, none had used the Web-based image display software, no prior training in FOV assessment had been given, and some of the reviewers were familiar with one liquid-based preparation but not sufficiently with the other. It was hypothesized that if 3 or more of 6 observers were able to classify a case correctly, there was most likely good reason, and, hence, training would improve performance of the remaining 3 or fewer observers. At this level of agreement, sensitivity of 88% and specificity of 93% were obtained for the categorization of the cases into 1 of the 2 groups. When various threshold levels of abnormality were separately examined for cases with squamous abnormalities, the sensitivity of detection of HSIL or worse (94%) was better than the sensitivities at lower threshold levels (85% for LSIL or worse and 86% for ASC-US or worse). These figures compare favorably with reports of the sensitivity of manual screening reported in clinical trial settings. In the original FDA clinical trials for the FocalPoint screening system, overall sensitivity for binary triage of all abnormal cases (ASC-US or worse) was 78%, for LSIL or worse was 83%, and for HSIL or worse was 93%, figures that are very close to the results obtained in this study.

One limitation of the present study is the fact that
Table 6

<table>
<thead>
<tr>
<th>Abnormality Threshold</th>
<th>Sensitivity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC-US or worse</td>
<td>86</td>
</tr>
<tr>
<td>LSIL or worse</td>
<td>85</td>
</tr>
<tr>
<td>HSIL or worse</td>
<td>94</td>
</tr>
</tbody>
</table>

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

abnormal cases were overrepresented in the current study population compared with the aforementioned prospective study. This difference could enhance the apparent sensitivity of detection owing to increased disease prevalence and, hence, heightened vigilance of observers. A prospective study to more thoroughly test these operating characteristics is underway at the time of this writing. When specific types of cellular abnormality were examined, it seemed that reviewers were better at recognizing as abnormal the cases that contained potential high-grade squamous lesions than cases with low-grade squamous abnormalities. Identification of glandular lesions remains a challenge to the system, as in the standard manual review process; however, these lesions generally represent a small minority of cases in actual practice.

We anticipate that the interpretation of FOVs can be improved with training and the development of strategies to compensate for the difficulties of FOV assessment in particular types of cases. The identification of common and recurring pitfalls in evaluation and of problematic types of cases will inform us in the assembling of study sets for the purpose of training. Can the quintile data provided by the instrument be used as an ancillary clue in individual cases? In separate studies, we are investigating the optimal number of FOVs for effective specimen triage and whether determinations of specimen adequacy can be made reliably. It also needs to be established if, and how often, diagnostic cells are present in the device-selected FOV (which is circular) but not in the captured JPEG image (which is rectangular and a centrally located subset of the larger circular field). For example, in the 4 squamous lesions (3 LSIL and 1 HSIL) that were missed by all reviewers, the abnormal cells were present on the slides but not in the rectangular JPEG images. Were they present at the periphery of the device-selected FOVs but outside the boundaries of the JPEG image? In the future, additional avenues of investigation might include the addition of color image capture, wider FOVs, and the making of virtual slide images for whole-slide manipulation and/or focusing. Possibly, virtual microscopy or whole-slide imaging might be used to identify solutions for the shortcomings of static digital image transmission.

We are encouraged that Web-based interpretation of automated device–selected images transmitted electronically can permit optimization of interpretative expertise and potentially provide cytology screening services to populations that currently have none, enhance resource sharing between laboratories, and centralize cytology interpretation services. Reviewers will be further trained from study sets acquired from the present exercise in an attempt to further enhance the capability of interpreting FOV images in this system. In addition, a larger prospective trial with specimens from current patients is underway.

References


