The Becton Dickinson FocalPoint GS Imaging System

Clinical Trials Demonstrate Significantly Improved Sensitivity for the Detection of Important Cervical Lesions

David C. Wilbur, MD,1 W. Stephen Black-Schaffer, MD,1 Ronald D. Luff, MD,2 Kurian P. Abraham, MD,3 Cynthia Kemper, CT(ASCP),4 James T. Molina, MD, PhD,5 and William D. Tench, MD6

Key Words: Cervical cytology; Computer-assisted screening; Cytology imaging system

Abstract

Location-guided screening in cervical cytology offers a potentially significant advance over routine manual screening. A prospective, 2-armed, masked clinical trial of the BD FocalPoint GS Imaging System using SurePath slides (BD Diagnostics-TriPath, Burlington, NC) compared routine manual screening and quality control rescreening with computer-assisted, field-of-view screening and device-directed quality control rescreening. The results obtained in the 2 arms were compared with adjudicated reference diagnoses for each slide. Sensitivity, specificity, and negative predictive value were calculated for the detection of atypical squamous cells of undetermined significance and greater (ASC-US+), low-grade squamous intraepithelial lesion and greater (LSIL+), and high-grade squamous intraepithelial lesion and greater (HSIL+) groups. We evaluated 12,313 slides. The detection sensitivities for HSIL+ were increased by 19.6% (P < .0001) and for LSIL+ were increased by 9.8% (P < .0001) in the computer-assisted arm, with small statistically significant decreases in specificity. For ASC-US+ sensitivity and specificity, the study arms were not statistically different. Use of this system might be expected to improve accuracy for clinically important entities without increasing equivocal case detection.

The era of clinical cervical cytology computer-assisted screening began with approval from the US Food and Drug Administration (FDA) of the AutoPap 300QC device (NeoPath, Redmond, WA) in 1995,1,2 This instrument was the first of the series of devices that followed this initial approval and are currently marketed and hereafter referred to in this article as the BD FocalPoint Slide Profiler (or Slide Profiler; BD Diagnostics-TriPath, Burlington, NC). The FDA evaluated the AutoPap 300QC (the initial quality control [QC] device) with assistance from the cytology community through the recommendations of the Intersociety Working Group for Cytology Technologies,3 which helped to establish acceptable operational parameters and to define the methods that should be used to test and validate present and future automated screening devices.

The Slide Profiler device was a conceptual breakthrough in computer-assisted screening. Prior experimental devices were specifically designed to identify, locate, and, therefore, present individual abnormal cells to cytologists.4,5 The Slide Profiler operational concept was different. It prioritized slides based on the probability and degree of abnormality present on entire slides, giving abnormal slides, particularly slides with the highest level of abnormality, scores placing them highest in the ranking. This concept allowed QC rescreening to selectively target the highest scoring slides, thus focusing the search for false-negative cases on the highest probability category. Implementation of this concept resulted in a significant and practical improvement over the random rescreening process. Data from the premarket approval (PMA) clinical trial showed that, used as a QC device, the Slide Profiler identified 4 times more false-negative slides of any abnormal type and 7 times more false-negative slides at the level of high-grade
squamous intraepithelial lesion (SIL) and above, as compared with routine Clinical Laboratory Improvement Amendments of 1988 (CLIA)-mandated rescreening procedures. Early adopters of the device into clinical use verified the sponsored trial results, reporting actual improvements in false-negative slide detection.

Of course, primary screening of cervical cytology slides is the ultimate goal for automated devices, and the Slide Profiler concept of hierarchical slide stratification could be used for this task as well. Primary screening by this method includes the use of slide scoring to identify low-probability populations of slides falling below a fixed score threshold. Slides below this “primary” threshold could be reliably reported as negative without any human manual screening (so-called no further review [NFR] slides), while the slides above the primary threshold would be manually screened in a routine manner (so-called review slides). In addition, the directed QC rescreening (referred to as Directed QC Technology, BD, Franklin Lakes, NJ) of high-scoring slides was retained, allowing for the unique advantage of rescreening only the population of cases most at risk of containing false-negative cases.

Clinical trials of the Slide Profiler, using a primary threshold cutoff maximizing the NFR population at 25%, showed improvements in detection of abnormal slides at all abnormality levels when compared with manual screening. These data led to the Slide Profiler’s approval as the first computerized primary screening device in 1998. The Slide Profiler has been in wide use since FDA approval, and numerous postapproval studies have confirmed its clinical usefulness. Extended clinical studies also led to FDA approval of the device to screen liquid-based BD SurePath slides (BD-TriPath) in 2001.

In addition to the ability to prioritize slides based on the probability of abnormality being present somewhere in a slide, always inherent in the Slide Profiler device was the ability to perform the same exercise on the individual microscopic fields of view (FOVs). Each FOV represents a microscopic field viewed at a magnification equivalent to that obtained with a 10× objective (approximately ×100 final magnification). The device prioritizes all FOVs across an entire slide and provides the slide coordinates (and, hence, the locations) of the FOVs having the highest probability of containing abnormal cells.

This concept was originally demonstrated using a manual system of FOV location transfer using printed pages referred to as PAPMAPs, over which slides were placed and FOVs traced onto the slide, essentially providing the screening cytologist with a predotted slide showing the highest probability of abnormality locations. Conceptually, if no abnormal cells were identified in the FOVs, the slides could be reliably labeled as “negative” and did not require further manual review. If abnormality (or potential abnormality) was identified during the initial FOV screen, the entire slide would be triaged to a full manual screening. This process became known as “location-guided screening” (LGS). With this process, the combination of slide and FOV prioritization led to a workflow in which slides scoring below the primary threshold (the NFR population) could be reliably called negative without manual screening. With FOV screening performed on all review slides scoring above the primary threshold, full manual screening was necessary only when potential abnormality was identified in the initial FOV review.

Initial clinical studies of the Slide Profiler using PAPMAP tracings showed encouraging results. One such study showed improvement of abnormal slide detection sensitivity when compared with the sensitivities obtained using the Slide Profiler with slide ranking alone. But key lessons were learned from other studies, including the importance of the triage to full manual screening when any abnormality was detected. Using the Slide Profiler to make final interpretations from FOV review alone, without triage to full manual screening, showed that such a practice was more sensitive than manual screening for the detection of abnormal cases. However, the FOV-only screen did not always label the slide with the correct level of abnormality; full manual screening when any abnormality was identified in FOVs was necessary to ensure maximal accuracy for the final interpretation. In later versions of the Slide Profiler system with LGS, the PAPMAP tracing was automated, using an automated stage that would guide users to the most suspicious FOVs. Preliminary studies using this method of interface have also shown improved performance compared with manual screening.

Slide ranking and FOV ranking by the Slide Profiler device have been shown to be very robust. By using BD SurePath liquid-based slides on the Slide Profiler with LGS, 1 study showed very reliable slide scoring, with 98% of all high-grade squamous intraepithelial lesion (HSIL) and cancer cases sorting in the top 30% of all slides and with 79% in the top 15% of all slides. Another study, also using BD SurePath slides with the Slide Profiler with LGS, showed 82% of all HSIL and cancer slides sorting in the top 30% of all slides and 58% in the top 15%. In both cases, not a single HSIL+ slide fell into the NFR population. In addition, a study using the Slide Profiler with LGS in a specialized investigational telecytology application showed that abnormal cells were identified in the first ranked FOV in 50% of cases.

The present report details the next phase of BD FocalPoint Slide Profiler development. Guided screening of liquid-based BD SurePath slides with FOV review of all cases is more adaptable to risk-averse practitioners desiring to have a manual review performed in every case. The BD FocalPoint GS Imaging System provides FOVs for all slides scanned (in adequate cases). There is no NFR population of low-scoring slides that is not manually screened. FOVs are localized for cytologist review using an automated stage retrofitted...
on commonly used microscopes (GS Review Station, BD Diagnostics-TriPath) that automatically presents FOVs in order based on decreasing FOV score, meaning that the highest probability FOV is presented first. The cytologist controls the microscope FOV advance using a mouse or foot pedal interface. This article describes the results of a large, 2-armed, prospective, masked clinical trial comparing the BD FocalPoint GS Imaging System with Directed QC Technology for the primary screening and directed QC rescreening of slides compared with the standard practice of manual screening and CLIA-mandated QC rescreening.

Materials and Methods

A prospective, multicenter, 2-armed, masked clinical study was designed to evaluate the effectiveness of manual screening of BD SurePath Papanicolaou (Pap) tests compared with LGS using the BD FocalPoint GS Imaging System. This study was conducted at 4 geographically diverse CLIA-certified clinical laboratory sites in the United States. All clinical trial sites were experienced in the processing of the BD SurePath Pap tests and had annual volumes ranging from 35,000 to 122,000. Four cytotechnologists per site, ranging in experience from 2 to 36 years, participated in each arm of the trial. All cytotechnologists participated in both study arms but did not screen the same slides in each of the 2 study arms. An independent site, also experienced in BD SurePath slide interpretation, was designated as the adjudication center, as is described subsequently. Institutional human subjects review board approval was obtained at all participating sites before initiation of the protocol.

BD SurePath Pap tests prepared using the BD PrepStain system (BD) from women 18 to 75 years of age were obtained consecutively from the recent files of the participating institutions. Slide exclusion criteria included broken or cracked slides or coverslips, lack of essential clinical information for diagnosis, slides that were part of a multiple-slide case, slides that were not able to be successfully processed on the GS Imaging System, and BD SurePath Pap tests with markings that could not be removed because of standard laboratory policies and procedures.

The manual initial screening (control arm) of the study consisted of 100% manual primary screening of slides processed per routine laboratory practice. This represented the original clinical processing of each specimen. The results obtained in the control arm were, therefore, the results originally reported by the laboratory. Laboratory procedures were all performed in accordance with the regulations of CLIA and any additional policies and procedures of the clinical site.

To ensure the analysis of adequate numbers of abnormal slides to provide statistical power to the trial, additional slides were added to each site’s routine archival slides. These “seeded” slides were prepared at the same time as the other clinical samples from BD SurePath residual pellets that were from cytology samples previously interpreted as abnormal at the same site or from BD SurePath collection vials supplied by BD-TriPath. These seeded samples were processed with fabricated plausible clinical information and entered into the laboratory information system at each site so as to “mask” the cytotechnologists and pathologists from identifying the seeded slides as part of the trial. The seeded samples were processed in a true prospective manner such that the interpretations of the original slides from these specimens were not known. Therefore, the control arm diagnosis for the seeded cases was that resulting from the screening and interpretation of each newly processed slide, not from the diagnosis of the original slide made from that sample.

All slides included in the control arm were then entered into the experimental GS arm of the study. In the experimental GS arm, 100% of slides were processed using the BD FocalPoint GS Imaging System in its intended use mode of operation. The GS Imaging System consists of the BD FocalPoint Slide Profiler and the BD FocalPoint GS Review Station. The GS Imaging System classified each slide as “review,” “process review,” or “rerun.” Any slide receiving a process review or rerun classification (each category is an indication of a technical processing failure) was excluded from the study if it could not be successfully run in a total of 3 attempts. For each review slide, the cytotechnologist performed the initial screen, reviewing all device-selected FOVs, up to 10 FOV locations and 1 visual confirmation field, using the GS Review Station and recorded an adequacy determination and a diagnosis. The first FOV reviewed was always a “location verification” field that contained a unique morphologic feature to ensure that the slide and microscope were properly aligned. The following FOVs were the ranked locations, beginning with the field with the highest probability of containing abnormality. If the FOVs showed no potential abnormality and were deemed adequate for interpretation, the cytotechnologist screened only the FOV locations and reported the case as “satisfactory for interpretation” and “negative for intraepithelial lesion or malignancy (NILM).” At this point, the slide became eligible to be selected for directed QC rescreening using the Directed QC Technology.

The cytotechnologist reviewed the entire slide whenever one of the following conditions was met: evidence of abnormality in any of the FOVs, specimen adequacy could not be determined on FOVs, endocervical component was not identified in the FOV locations, squamous component was not identified, or the device did not obtain specified FOV locations for the slide (which most commonly occurred with low-cellularity specimens). Full slide review occurred immediately following the FOV review and was performed by the same cytotechnologist who performed the FOV review.
For each review slide, the GS Imaging System provided the following information regarding slide adequacy to the screener: (1) presence or absence of a squamous component and (2) presence or absence of an endocervical component. By using this information in conjunction with the review of FOVs, the cytotechnologist determined whether a slide was satisfactory or unsatisfactory, in accordance with the criteria defined by the 2001 Bethesda System.

The GS Imaging System also selected at least 15% of the highest scoring NILM slides for QC rescreen. All QC designated slides received a full slide review and, if needed, were subject to the laboratory’s normal protocol for hierarchical review. All QC policies that applied to the control arm were also applied to the GS arm. For example, any patient history that would have resulted in targeted rescreening in the control arm was provided in the GS arm. The laboratory was instructed to use standard practice for targeted QC rescreening. Any QC task was assigned to a cytotechnologist who had not previously reviewed the slide.

In this study, performance in each study arm was compared with an adjudicated “truth” standard for each slide. A reference diagnosis or “cytologic truth” for each slide was determined by 1 of 2 methods: (1) For slides interpreted in both arms of the study as NILM/satisfactory, this result was taken as the reference diagnosis. (2) For slides interpreted as abnormal by either or both study arms, slides interpreted as unsatisfactory by either or both study arms, and slides with differing interpretations or adequacy determinations between the two study arms, the slides were forwarded to a separate site serving as the cytology adjudication center (CAC). A subsample of slides interpreted as NILM by both study arms was reviewed by the CAC to seed the referred abnormal slide population because of the potential bias that would be created if only potentially abnormal slides were reviewed in the CAC. In addition, adjudication of some NILM slides was necessary to allow for specificity determinations in the 2 arms of the study.

In the adjudication process, 2 cytopathologists initially examined each slide after the slides had been screened by cytotechnologists. Each CAC cytotechnologist and cytopathologist was masked from all site diagnoses and the GS Imaging System data information for each slide. The reviewers were given the subject’s available case history. If the 2 cytopathologists were in agreement regarding the slide diagnosis, that result became cytologic truth. If no consensus was achieved, the slide received a third cytopathologist review, and if all 3 interpretations differed, the slide received a 3-cytopathologist multthead microscope review until a consensus cytologic truth determination was achieved.

Data analysis compared the diagnoses made on each slide in the 2 study arms with the cytologic truth determination for that slide derived from the adjudication process. Follow-up biopsy results were not used as the “gold standard” reference in this trial, as per FDA protocol. The goals of the study as reported herein were determinations of the sensitivity and specificity for the detection of abnormal cases at the levels of HSIL and greater (HSIL+), low-grade squamous intraepithelial lesion and greater (LSIL+), and atypical squamous cells–undetermined significance and greater (ASC-US+). The specificity calculations were performed using the subset of NILM slides that had been sent for adjudication. An additional “bootstrap” analysis was performed to model the specificity calculation across the entire population of NILM cases. Briefly, this method extrapolates to the entire NILM population, the rate of reclassification of NILM cases sent to the adjudication process, and calculates the specificities based on the entire data set following this extrapolation. Further analysis of the detection sensitivity for slides adjudicated as invasive carcinoma was also performed. Calculations of the negative predictive value (NPV) for HSIL+, ASC-US/SIL ratios, false-negative rates, and adequacy determinations were performed for each study arm. In addition, slide ranking determinations based on the diagnostic category were tabulated to assess the robustness of the GS Imaging System classification algorithms.

Statistical analyses to compare the sensitivity and specificity results of the control and GS arms of the study used the McNemar exact test, which was calculated using StatXact (version 7) software (Cytel, Cambridge, MA).

Results

Study identification numbers were assigned to the 12,732 slides originally enrolled. Of these, 345 slides (2.7%) were excluded from the trial based on failure to meet the inclusion and exclusion criteria or failure to be successfully processed by the GS Imaging System. The exclusions were well distributed among the 4 trial sites. An additional 71 slides (0.6%) were never processed through the GS Imaging System and were, therefore, excluded. After completion of the study, 3 additional slides were excluded owing to incomplete data being available. Based on the aforementioned exclusions, 12,313 slides (96.7% of slides initially enrolled) were fully evaluated in both study arms, the results of which were available for comparisons. Final numbers of slides fully evaluated at each of the sites ranged from 2,695 to 3,424. The number of seeded slides fully evaluated at all sites was 361 (299 derived from reprocessed site material and 62 from BD-TriPath–supplied vials). The number of seeded slides at each site ranged from 45 to 123.

Screening and interpretation of BD SurePath Pap Test slides using the GS Imaging System were found to be statistically superior to the manual screening and interpretation process for the detection of HSIL+ and LSIL+ HSIL+ sensitivity was 85.3% in the GS arm vs 65.7% in the
control arm ($P < .0001$), demonstrating a 19.6% increase in sensitivity with a small, but statistically significant decline in specificity ($-2.6%; P < .0001$). When interpretations of LSIL, atypical squamous cells—a high-grade lesion cannot be excluded (ASC-H), or atypical glandular cells (AGC) are included as “matches” for slides with adjudicated diagnoses of HSIL+ (therefore, a clinically relevant process having identical initial management was identified), the sensitivity of HSIL+ detection increased to 96.6% in the GS arm and 96.1% in the control arm. LSIL+ detection in the GS arm showed a sensitivity of 86.1% compared with 76.4% in the control arm ($P < .0001$), a 9.8% increase in disease detection with a 1.9% statistically significant ($P = .0032$) decline in specificity.

ASC-US+ sensitivity and specificity were not significantly different between the 2 arms of the study (sensitivities and specificities, respectively, of 81.1% and 84.5% in the GS arm vs 82.6% and 82.7% in the control arm). Specificities for the entire population of cases as determined by the bootstrap method showed only minor changes in the results (HSIL+ specificity decline of 0.7% in the GS arm [$P < .001$], LSIL+ specificity decline of 0.4% in the GS arm [$P = .037$], and ASC-US+ specificity increase of 0.3% in the GS arm [$P = .18$]). The NPV (for not-HSIL+) of a slide in the GS arm was 99.7% and in the control arm was 99.4%, showing that despite the minor decrease in specificity for HSIL+ detection, the GS arm is overall more accurate in the precise detection of HSIL+ and the reliable exclusion of HSIL+.

There were 1,443 ASC-US+ slides identified in the control arm of the study with a resulting ASC-US/SIL ratio of 0.69. This is significantly greater than the ASC-US/SIL ratio of 0.40 in the GS arm ($P < .0001$). The ASC-US/SIL ratio therefore decreased by 42% in the GS arm, representing a significant reduction in equivocal results. Excluding the seeded slides from ASC-US/SIL calculations resulted in an ASC-US/SIL ratio of 0.45 in the GS arm compared with 0.88 for the control arm, maintaining a significant decrease in equivocal results ($-49$%) in a more representative standard screening population.

The overall unsatisfactory for evaluation rates were 0.16% in the control arm and 0.22% in the GS arm. Compared with the component of slides that were adjudicated by the CAC, the GS arm correctly assessed the slide as being unsatisfactory 91.3% of the time, and the control arm correctly assessed the slide as being unsatisfactory 73.9% of the time. This resulted in a 17.4% increase in unsatisfactory slides correctly assessed in the GS arm compared with the control arm (difference not statistically significant).

For every slide that the GS Imaging System determines to have sufficient cellularity, a quintile ranking is provided that corresponds to the slide’s likelihood of containing abnormality. The quintile rank is expressed as a number from 1 to 5, where quintile 1 indicates the highest risk (top 20%) of abnormality, and so forth. Table 3 shows the number of abnormal slides of sufficient cellularity to generate FOVs for analysis, as determined

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Sensitivity and Specificity Calculations With Corresponding 95% CL and $P$ Values by Diagnostic Category$^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference Diagnostic Category/Measure</td>
<td>GS Arm (%)</td>
</tr>
<tr>
<td>ASC-US+</td>
<td>Sensitivity</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
</tr>
<tr>
<td>LSIL+</td>
<td>Sensitivity</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
</tr>
<tr>
<td>HSIL+</td>
<td>Sensitivity</td>
</tr>
<tr>
<td></td>
<td>Specificity</td>
</tr>
</tbody>
</table>

ASC-US, atypical squamous cells of undetermined significance; CL, confidence limits; GS, Becton Dickinson FocalPoint GS Imaging System; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

$^*$ Specificities as determined by the subset of cases evaluated as negative for intraepithelial lesion or malignancy sent to the adjudication process.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>ASC-US/SIL Ratios by Study Arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS Arm</td>
<td>Control Arm</td>
</tr>
<tr>
<td>382</td>
<td>952</td>
</tr>
</tbody>
</table>

ASC-US, atypical squamous cells of undetermined significance; GS, Becton Dickinson FocalPoint GS Imaging System; LSIL, low-grade SIL; SIL, squamous intraepithelial lesion.
by the truth adjudication process, with their associated quintile rank. These data demonstrate that a high proportion of these abnormal slides are ranked in quintile 1, and that progressively fewer slides containing abnormality are ranked in the lower likelihood quintiles. In the most important category of HSIL, 1, 161 (86.6%) of 186 cases were ranked in the top quintile, and 94.6% were ranked in the top 2 quintiles, indicating a very robust abnormal slide classification algorithmic process.

As summarized in Table 4, the detection of cancer was numerically higher in the GS arm. Of the 49 slides with adjudicated diagnoses of cancer, 34 were identified as such in the GS arm, and 22 were identified in the control arm. Of the 15 cancers undercalled in the GS arm but appropriately triaged to full manual review by the GS Imaging System-assisted practice, 12 were classified as HSIL, 2 as adenocarcinoma in situ, and 1 as NILM. This “negative” slide was indicated for a full slide review by the device owing to a designation of “scant cellular- ity.” It received a full manual screening but was subsequently classified as NILM by the cytotechnologist. A GS Imaging system “safety net” function ensures that all “scant cellularity” diagnoses must go to a manual full slide review. In the study, there were 12 samples of carcinoma that were found by the device to be limited in squamous cellularity that were ranked as quintile 5 with an accompanying designation that they were “low-cellularity” cases and were, by device operation protocol, triaged to a full manual screening. Of the 27 cancers undercalled in the control arm, 22 were classified as HSIL, 4 as AGC, and 1 as NILM. The overall clinical sensitivity for cancer in each of the trial arms was, therefore, 98% (48/49), but the GS arm correctly classified 12 more of the total of 49 cancer cases as such (a 25% increase in precise diagnosis).

Discussion

The goal of innovation in cervical cytology is to improve the overall accuracy of the final interpretation and to improve laboratory productivity through efficiency of operation, improved workflow processes, and enhancement of task interest and overall satisfaction of operating personnel. The FDA PMA clinical trial of the BD FocalPoint GS Imaging System was designed to test performance in both areas: accuracy and productivity. The current report details the accuracy performance of the device when integrated into a total system with cytotechnologist and pathologist screening and interpretation. Details of productivity performance will be presented in a separate article focused specifically on that topic.

The BD FocalPoint GS Imaging System trial was a large-scale, prospective, 2-armed, masked study, directly comparing cervical cytology slide final interpretation performance between the current standard of practice of manual screening with CLIA-mandated QC rescreening (the control arm) and the GS Imaging System. The GS Imaging System is composed of device prescreening with slide ranking and selection of high-risk FOVs for cytotechnologist review, followed by full manual screening when potential abnormality or adequacy issues are raised in the FOV review. It includes 15% directed QC rescreening of high ranking slides initially interpreted as NILM (the GS arm). A truth determination was made for each slide in the study to reconcile which study arm was actually correct in cases in which there was a difference in final interpretation between the 2 arms and as a measure of specificity in the case of negative slides. This consisted

---

**Table 3**

<table>
<thead>
<tr>
<th>Quintile</th>
<th>ASC-US</th>
<th>ASC-H</th>
<th>AGC</th>
<th>LSIL</th>
<th>HSIL</th>
<th>Squamous Cancer</th>
<th>Adenocarcinoma</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>175</td>
<td>17</td>
<td>4</td>
<td>343</td>
<td>130</td>
<td>26</td>
<td>5</td>
<td>700</td>
</tr>
<tr>
<td>2</td>
<td>121</td>
<td>7</td>
<td>0</td>
<td>116</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>259</td>
</tr>
<tr>
<td>3</td>
<td>89</td>
<td>1</td>
<td>2</td>
<td>55</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>154</td>
</tr>
<tr>
<td>4</td>
<td>61</td>
<td>0</td>
<td>1</td>
<td>27</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>43</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>71</td>
</tr>
<tr>
<td>Total</td>
<td>489</td>
<td>25</td>
<td>8</td>
<td>567</td>
<td>149</td>
<td>31</td>
<td>6</td>
<td>1,275</td>
</tr>
</tbody>
</table>

AGC, atypical glandular cells; ASC-H, atypical squamous cells, cannot exclude HSIL; ASC-US, atypical squamous cells of undetermined significance; GS, Becton Dickinson FocalPoint GS Imaging System; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion.

---

**Table 4**

Distribution of Cancer Cases (Reference Diagnosis) by the GS and Control Arms

<table>
<thead>
<tr>
<th>Control Arm</th>
<th>NILM</th>
<th>AGC/AIS</th>
<th>HSIL</th>
<th>Cancer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>GS Arm</td>
<td>NILM</td>
<td>AGC/AIS</td>
<td>HSIL</td>
<td>Cancer</td>
<td>Total</td>
</tr>
<tr>
<td>NILM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AGC/AIS</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>HSIL</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Cancer</td>
<td>1</td>
<td>3</td>
<td>13</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>4</td>
<td>22</td>
<td>22</td>
<td>49</td>
</tr>
</tbody>
</table>

AGC, atypical glandular cells; AIS, adenocarcinoma in situ; ASC-US, atypical squamous cells of undetermined significance; GS, Becton Dickinson FocalPoint GS Imaging System; HSIL, high-grade squamous intraepithelial lesion; NILM, negative for intraepithelial lesion or malignancy.
of agreement of the 2 arms in most satisfactory/NILM cases and referral to an external adjudication process for all abnormal, unsatisfactory, and discrepant cases, as well as for a subsample of satisfactory/NILM cases. Results obtained in each of the study arms were then compared with the reference truth determination for each slide. This process is a standard protocol that has been previously used in FDA PMA clinical trials for automated screening devices.8,26

Accuracy assessments included the ability to correctly classify slides in each of the study arms at the levels of HSIL+, LSIL+, and ASC-US+. In addition, the ability to provide accurate assessments of slide adequacy and to correctly classify cases of invasive carcinoma was studied. Results showed that the GS Imaging System outperformed the manual standard practice arm at the most important levels of abnormality, with statistically significant improvements in sensitivity for the detection of HSIL+ (+19.6%) and LSIL+ (+9.8%). This was achieved with a much smaller decrease in specificity, meaning that a very small number of slides were overcalled in the GS arm to obtain a much larger improvement in sensitivity.

It is interesting that in the HSIL+ category, if final diagnoses of LSIL, ASC-H, and AGC were considered matches to HSIL+ truth as opposed to “false-negative” (which was the required convention of the protocol), this more clinically relevant sensitivity of HSIL+ detection increased to 96.6% in the GS arm compared with 96.1% in the control arm. This is a strong indication that the GS Imaging System is overall very robust at correctly identifying cases that harbor high-grade lesions. In addition, based on the HSIL+ categorical data, the GS arm was also substantially better than the control arm at providing a more precise diagnosis of HSIL+. When considering this tradeoff of sensitivity vs specificity in a screening situation such as the Pap test, calculations of the NPV of the test for HSIL+ become relevant. In this study, the NPV (for not HSIL+) of the GS arm was slightly greater than in the control arm, indicating that the predictive value of a negative test in excluding HSIL+ was improved. Although the NPV of each arm was very high (99.7% for the GS arm and 99.4% for the control arm), a risk-averse consideration of the complementary rate of expected “missed” cases of HSIL+ shows the GS arm to have half (0.3%) the rate of missed HSIL+ of the control arm (0.6%).

In the ASC-US+ category, the GS arm actually identified fewer cases than did the control arm (a non–statistically significant difference). This finding, when combined with the significant improvements in performance for LSIL+ and HSIL+, indicates that the GS Imaging System actually identifies fewer abnormal cases as ASC-US, which also might be considered an improvement in performance, and suggests that some cases designated as ASC-US in the control arm were being correctly classified as benign or reactive (NILM) or as LSIL+/HSIL+ in the GS arm. The ASC-US/SIL ratios were found to be substantially lower in the GS arm than in the control arm of the trial. On the opposite end of the spectrum, the GS arm identified more invasive carcinoma cases as such than did the control arm (69% vs 45%, respectively), although the sensitivity for detection of cancer with abnormal diagnosis at any level was the same at 98% (48/49) in both arms. Most such undercalled cases were classified as HSIL and AGC/adenocarcinoma in situ in both arms of the study. One can speculate that direction to specific high scoring FOVs may make the final classification more specific with the GS Imaging System, but that hypothesis remains untested at present.

A similarly designed FDA PMA clinical trial for the ThinPrep Imaging System analyzed results in a similar manner. In that study, the results were substantially different from those in the present trial. The ThinPrep Imaging System trials showed statistically significant increases in the detection sensitivity for ASC-US+ cases (+6.4%) but a non–statistically significant decrease in detection sensitivity for LSIL+ cases (−0.5%) and a non–statistically significant increase in detection sensitivity for HSIL+ cases (+5.8%).26,27 The absolute sensitivities for the detection of LSIL+ and HSIL+ cases were lower in the ThinPrep Image trial at 79.2% and 79.9%, respectively, compared with the GS Imaging System sensitivities of 86.1% and 85.3%, respectively. The ThinPrep Imaging System achieved statistically significant improvement in the specificity of HSIL+ detection, meaning that it improved the performance of the system for determining what is not HSIL+. Subtle methodological differences between the trials may account for some of the data noted; however, a consistently clear message is that the GS Imaging System is better than standard manual screening for the detection of the more important abnormal slides containing SIL and does not achieve this superior performance from the detection of increased numbers of “equivocal” atypical cases.

There are a number of features intrinsic to the GS Imaging System that may account for the improved performance over manual screening. First and foremost is the device’s ability to score each slide based on the likelihood of abnormality being present. This has been a tenet of the operation of prior Slide Profiler–based systems, such as the original QC and primary screening applications. Data from a variety of studies show that the degree of abnormality is roughly (although not absolutely) proportional to the final score, meaning slides with higher levels of abnormality generally are given the highest scores. This allows for ranking of slides into hierarchical categories, and when this information on slide ranking is given to screeners, the higher a priori probability of abnormality being present increases the sensitivity of the procedure.

In the GS Imaging System, the hierarchical ranking procedure extends to individual FOVs. This means that the first FOV reviewed in the screening process has the highest likelihood of containing abnormality, and previous studies have
The random rescreening of negative slides has been shown to be an inefficient task. In 10% random rescreening protocols, the maximum number of false-negative cases that can be identified is 10% of the total. In addition, the sensitivity for the detection of false-negative slides while rescreening "negative" slides has been shown to be as low as 25%.28 Looked at in this manner, a 10% random rescreening procedure would be expected to identify as few as 2.5% of all the false-negative cases in any population. Directed rescreening has the advantage of looking at only slides given a high a priori probability of being false-negative. In previous studies, QC rescreening of this population was shown to identify 4 times the number of total false-negative slides at any level of abnormality and 7 times the number of HSIL+ false-negative cases.1,2 Review of the slide ranking data in the present and several prior studies shows very clearly that HSIL+ cases are consistently ranked in the top quintiles of BD FocalPoint Slide Profiler–screened slides, therefore making review of any case in the highest rankings originally called negative a much more fruitful procedure than is a random process.21,22

An additional safety net has also been built into the algorithm for GS Imaging System slide categorization that is very robust for the identification of high-grade lesions that present in low-cellularity slides or slides in which cellularity is obscured by diathesis or blood. Cancer cases are well documented to sometimes present in this manner,29,30 and, hence, FOV-only screening may not be optimal for correct triage or final interpretation. In premenopausal patients, the low squamous cellularity and rarity of malignant cells are often attributed to excessive blood, inflammation, and necrotic debris associated with the invasive tumor. The latter elements may form the dominant components of the sample, hence diluting the number of malignant cells in the final preparation. In postmenopausal patients, squamous cellularity is often limited as a result of poor sample collection from atrophic epithelium. Regardless of the etiology, samples of low or inadequate squamous cellularity must be considered at increased risk for harboring a significant lesion and, thus, deserve additional attention during the screening process. Such cases, presenting with low squamous cellularity, are routinely labeled as such and FOVs are not generated. Instead, slides with low cellularity are always flagged for a full manual screening. In a previous study using the BD FocalPoint Slide Profiler with LGS, about a quarter of invasive carcinomas (18/69)30 presented with low squamous cellularity, as did about a third (18/49) in the present study, and all such cases were classified appropriately and received a full manual screening.

Another item of significant importance in the practice of cytology is the ability to correctly classify slides as satisfactory or unsatisfactory. In this trial, the GS arm labeled more truth-determined unsatisfactory cases as such than did the control arm, by 19%. Truth-determined unsatisfactory slides were correctly classified 91.3% of the time in the GS arm vs 73.9% of the time in the control arm. As has been well-documented for BD SurePath slides, the overall unsatisfactory rates were very low overall, at 0.22% for the GS arm and 0.16% for the control arm.

Overall, the results of the clinical trial confirm the robust nature of the BD FocalPoint Slide Profiler concept for the hierarchical ranking of slides, the identification of the adequacy of a specimen, and the ability to reliably identify low-cellularity specimens at greatest risk for containing invasive carcinomas. In addition, the BD FocalPoint GS Imaging System shows robust FOV hierarchical ranking. When used in its intended mode of presentation of FOVs to screening cytologists with full manual screening only when potential abnormality or adequacy issues are present, the system leads to significant improvement in the precise detection of the most important cervical cytology lesions, namely the HSIL, LSIL, and invasive carcinomas, without a concomitant increase in equivocal cases of squamous atypia.

In the post–human papillomavirus vaccine implementation era, the number of high-grade squamous abnormalities present in any given population of patients is expected to diminish.31 As the prevalence of true abnormal cases in the overall population decreases, the specificity of detection (positive predictive value) at any level of sensitivity will necessarily decrease. In addition, experiments in other visual screening exercises have clearly demonstrated that the sensitivity of visual detection of events (and, thus, the predictive value of a negative result) is decreased as the events become rarer.32 The ability to increase the prevalence of “abnormals” in a screening exercise will, therefore, be necessary to maintain the current level of positive and negative predictive value in cervical cytology. Stratification of slides and of slide FOVs by the BD FocalPoint GS Imaging System is one method of increasing the prevalence of abnormal events presented to observers and, therefore, the predictive value of the overall cervical cytology examination, which will be important to maintaining quality and patient safety in the postvaccine era.

Of course, standard use of the device in clinical laboratories will be necessary to further document its role in a routine practice setting, but the data presented herein provide validation for initial introduction as a superior system compared with the current manual screening standard of practice.
References