

Effectiveness of AutoPap System Location-Guided Screening in the Evaluation of Cervical Cytology Smears

Mark W. Stevens, M.Appl.Sc., C.F.I.A.C.,* Amanda J. Milne, C.M.I.A.C.,
Ian H. Parkinson, Ph.D., Walter W. Nespolon, B.Appl.Sc., C.M.I.A.C.,
Nicola L. Fazzalari, Ph.D., Neeru Arora, M.B.B.S., M.D., F.R.C.P.A.,
and Thomas J. Dodd, M.B.B.S.(Hons), F.R.C.P.A.

Location-guided screening is a feature of the AutoPap primary screening software. Areas of the smear most likely to contain an abnormality are identified for prompt review by the cytotechnologist, thereby facilitating diagnostic accuracy and reducing laboratory workload. A two-armed retrospective study comprising 6,000 conventional smears was undertaken to compare this approach with the current practice of full manual screening of conventional smears. Discrepant diagnoses between the two study arms were subject to an internal discrepancy review process to determine the final truth diagnosis. Analysis of the results show that AutoPap location-guided screening is at least equivalent to current practice when detecting high-grade or suspected high-grade smears. However, the device does not detect low-grade abnormalities, unsatisfactory smears, an endocervical component or organisms, as well as standard screening. The device also assigns a numerical score to each slide, with abnormal smears allocated a higher rank. Slide ranking was found to be of value in triaging abnormal smears for prompt screening and reporting. The performance of the primary screening software was found to be comparable to previous studies, with the majority of abnormal smears being selected by the instrument for manual review.

Diagn. Cytopathol. 2004;31:94-99. © 2004 Wiley-Liss, Inc.

Key Words: cervical smear; AutoPap primary screening system; computer-assisted diagnoses; location-guided screening; automation

The AutoPap primary screening system (APSS; TriPath Imaging, Burlington, NC) was approved by the U.S. Food and Drug Administration (FDA) on 5 May 1998 for in-

tended use in the initial screening of Papanicolaou smears. The system identifies up to 25% of successfully processed slides as normal and requiring no further review (NFR). These slides can be immediately archived. The remaining smears, classified as review, are manually screened with the device identifying at least 15% of smears for quality control manual rescreen.¹

Location-guided screening (AutoPap-GS) is a feature of the primary screening software and allows the laboratory to print a report (PapMap) that shows the areas of smears most likely to contain an abnormality. The PapMap report contains a to-scale rectangle representing the coverslip edges and up to 15 areas most likely to contain an abnormality as open circles (Fig. 1). The screening cytotechnologist screens the slides using these areas or field of views (FOVs). If FOV review indicates no abnormal findings, then the slide is reported as normal; otherwise, the smear is given a full slide review (FSR). The review slides are further ranked by the device from 1 to 5, with the first and second quintiles indicating smears with the highest probability of containing an abnormality. The features of the location-guided software enables the laboratory to improve accuracy, reduce workload, and triage abnormal cases for prompt review and reporting.²

There have been few studies assessing the effectiveness of location-guided screening of conventional smears. Apart from two location-guided screening (LGS) feasibility studies conducted by the manufacturer,^{2,3} only two other published studies have assessed the usefulness of LGS in a laboratory setting.^{4,5}

To appraise this technology, a two-arm retrospective study of 6,000 conventionally prepared smears was undertaken. The primary aims of the study were to determine whether AutoPap LGS-assisted practice (AutoPap-GS) was

Division of Tissue Pathology, Institute of Medical and Veterinary Science, Adelaide, Australia

*Correspondence to: Mark W. Stevens, M.App.Sc., C.F.I.A.C., Division of Tissue Pathology, Institute of Medical and Veterinary Science, P.O. Box 14, Rundle Mall, Adelaide, 5000, Australia.

E-mail: mark.stevens@imvs.sa.gov.au

Received 12 September 2003; Accepted 29 January 2004

DOI 10.1002/dc.20081

Published online in Wiley InterScience (www.interscience.wiley.com).

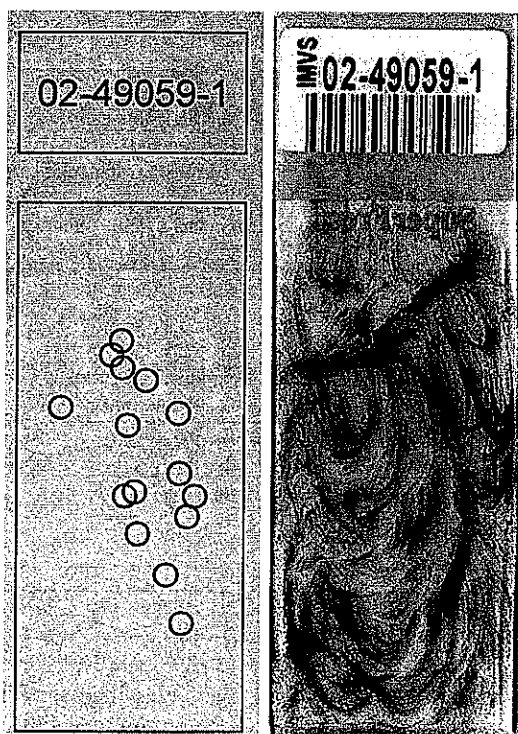


Fig. 1. Conventional smear and PapMap showing FOV locations.

equivalent to the current practice of manual screening of conventional smears when detecting all abnormal (low-grade or worse) smears, unsatisfactory smears, an endocervical component or organisms. The effectiveness of slide ranking in triaging abnormal smears was also assessed.

Materials and Methods

A total of 6,000 consecutive Papanicolaou smears reported by the laboratory from January to February 2000 were retrieved from archive. All slides were visually assessed for suitability and were excluded if they met the following criteria: broken or cracked slides or slides with two coverslips; smears from the vagina or vault; multiple slide cases or smears prepared using monolayer technology. Each slide was allocated a unique barcode label to ensure that the cytotechnologists were effectively masked to the original diagnosis. The slides were processed on the APSS, equipped with location-guided software, and slide results and PapMap tray reports were printed. Slides classified as NFR by the system were archived and PapMap FOV locations were manually traced onto review slides according to company guidelines.³

The cytotechnologists initially screened the FOV locations. An FSR was undertaken if FOV review indicated the presence of abnormal or questionable abnormal cells; an unsatisfactory smear, for example, too few squamous epithelial cells or the presence of an inflammatory infiltrate or

blood; a tumor diathesis; or a questionable hormonal pattern inconsistent with patients age, menstrual cycle, or clinical history.

Approximately 18% of the slides in the review group were finally selected for quality control (QC) review. Because of their ranking, these slides have the highest probability of containing abnormal cells. Following initial screening by the cytotechnologist, who was effectively masked to the final ranking of these smears, QC review slides received an FSR with the FOVs in place. Slides that failed processing due to physical characteristics or insufficient cellularity were categorized as process review (PR) and rerun a second time. A FSR was undertaken if the slide failed repeat processing.

Each slide was allocated a truth, current practice, and AutoPap-GS diagnosis. Truth was taken to be the concordant diagnosis between the current practice and AutoPap-GS diagnosis. Discordant diagnoses were adjudicated by discrepancy panels and in these instances truth was determined by majority agreement.

The AutoPap-GS diagnosis was recorded on worksheets using the Australian Reporting Terminology.⁶ The categories are normal, including benign cellular change; low-grade epithelial abnormality [mild dysplasia, human papillomavirus (HPV), and nonspecific changes in squamous and glandular cells]; high-grade epithelial abnormality (moderate to severe dysplasia and carcinoma); inconclusive, possible high-grade epithelial abnormality; and unsatisfactory. Table I provides a translation between the Bethesda Reporting System⁷ and the Australian terminology. To simplify comparisons between study arms and test for the detection of all abnormal smears, diagnoses were grouped into four main categories: normal; low-grade epithelial abnormality (LGEA); high-grade epithelial abnormality (HGEA), including inconclusive, possibly high-grade; and unsatisfactory.

To determine the performance of AutoPap-GS for the detection of an adequate endocervical component and the presence of organisms or infective agents such as *Candida*, *Trichomonas*, herpes simplex, and *Actinomyces*, the results of the analyses were compared to current practice. An adequate endocervical component was defined as the presence of a minimum of two clusters of well-preserved endocervical glandular and/or squamous metaplastic cells, with each cluster composed of a minimum of five well-preserved cells.⁷ The PapMap is not specifically designed to include fields with endocervical cells. If an endocervical component was not detected in the FOV location, no FSR was performed. For these smears, an endocervical component was reported as not detected. For slides categorized as NR, the machine determination of the presence or absence of an endocervical component was used in keeping with the intended use of the primary screening software.

Table I. Comparison of Australian and Bethesda Reporting Terminology

<i>Australian terminology</i>	<i>2001 Bethesda system</i>
Negative	Negative
Low-grade epithelial abnormality	Low-grade squamous intraepithelial lesion (includes HPV, mild dysplasia), atypical squamous cells of undetermined significance, glandular atypia (not otherwise specified)
Inconclusive	Atypical squamous cells of undetermined significance (cannot exclude HSIL), glandular atypia favor neoplasia
High-grade epithelial abnormality	High-grade squamous intraepithelial lesion, endocervical adenocarcinoma in situ, carcinoma
Unsatisfactory	Unsatisfactory

An unsatisfactory smear was defined as showing a scant epithelial component (well-preserved and well-visualized squamous epithelial cells covering less than 10% of the slide surface) or if obscuring blood, inflammation or air-drying artifact precluded interpretation of approximately 75% or more of the epithelial cells present.⁷

Upon completion of the study, the AutoPap-GS and current-practice arm diagnosis was compared. Current practice involved an initial full manual screen followed by a manual rescreen of smears from high-risk patients (approximately 20% of cases). The remaining smears were processed on the AutoPap 300 QC device, with manual review of 10% of smears selected as QC review.

Discordant diagnoses from the two study arms were adjudicated by discrepancy panels, comprising three experienced cytology professionals, to determine the diagnostic truth. Truth was determined by a majority agreement of the three panelists (M.W.S., T.J.D., N.A.). Discrepancies in the identification of an endocervical component or organisms were reviewed by an expert cytotechnologist whose assessment was considered to represent truth for the slides. To test the equivalence of the two study arms, 2 × 2 contingency tables were constructed for current practice versus AutoPap-GS, which show agreement or disagreement between the two study arms. The statistical test used to demonstrate the significance of the results was the McNemar's exact test, a conditional binominal test for matched pairs.⁸

The assignment of a numerical score (1-5) to each slide allows the device to rank slides prior to diagnosis. Review slides with a high probability of containing an abnormality are assigned a higher score. The usefulness of slide ranking information to facilitate slide triage and screening was also assessed.

Results

Of the 6,000 smears retrieved from archive, 422 (7.0%) smears were excluded from analysis. These included smears with broken or cracked slides (29); multiple slides or slides with two coverslips (186); vaginal and vault smears (123); monolayer preparations (64); and smears excluded for other reasons, for example, missing slide or cancelled requisition number (20). The remaining 5,583 smears comprised the study population and were processed using the AutoPap-GS

Table II. Final Diagnostic Truth as Determined by AutoPap-GS and Discrepancy Panel Review

Normal	4,331	(94.3%)
LGEA	145	(3.1%)
HGEA	34	(0.7%)
Inconclusive	15	(0.3%)
Unsatisfactory	67	(1.4%)
Total	4,592	

software. Of these, 986 (17.6%) were classified as NFR and immediately returned to archive. PapMaps were manually drawn on the 4,440 (79.5%) of smears classified as review. Eight hundred and eleven or 18% of the review slides were selected for quality control rescreen (QC review). Manual review of these review smears (n = 811) identified seven low-grade smears, one inconclusive, possible high-grade smear, and one unsatisfactory smear. One hundred and fifty-seven (2.8%) smears failed repeat processing due to technical reasons and these process review slides were given a FSR.

Table II shows the final diagnosis of smears in the study population, including resolution of discrepancies between the two study arms. Where there is a discrepancy between AutoPap-GS and current practice, the panel consensus determines the truth for that case. These data are then combined with the results obtained for the AutoPap-GS arm of the study. A total of 194 (4.2%) smears were reported as abnormal. Sixty-seven (1.4%) smears were found to be unsatisfactory due to an inadequate squamous component, excessive inflammation, or blood.

The comparison in the performance of AutoPap-GS and current practice for the detection of all abnormal smears, classified as LGEA or worse, is shown in Table III. The 2 × 2 contingency table shows that AutoPap-GS identified an additional 35 abnormal smears previously reported as normal by current practice, while AutoPap-GS missed an abnormality, as identified by current practice, in 92 smears. The difference in detection of LGEA or worse between the two study arms is statistically significant (McNemar's test, $P < 0.0001$).

The majority of the missed abnormalities, 87/92 or 94.5% of smears, showed only low-grade epithelial changes (Table

Table III. Result Matrix for Identification of All Abnormal Smears (LGEA⁺)*

AutoPap-GS ^a	Current practice		
	LGEA ⁺	Normal	Total
LGEA ⁺	158	35	193 ^b
Normal	92	4,208	4,300
Total	250	4,243	4,493

*n = 194. P < 0.0001.

^aAs determined by AutoPap-GS and discrepancy panel review.

^bExcludes one smear diagnosed as unsatisfactory on current practice.

Table IV. Result Matrix for Identification of LGEA Smears (n = 145)*

AutoPap-GS ^a	Current practice		
	LGEA	Normal	Total
LGEA	106	31	137 ^b
Normal	87	4,208	4,295
Total	193	4,239	4,432

*P < 0.0001.

^aAs determined by AutoPap-GS and discrepancy panel review.

^bExcludes seven high-grade/suspected high-grade smears and one smear diagnosed as unsatisfactory on current practice.

Table V. Result Matrix for Identification of HGEA and Suspected High-Grade Smears (n = 49)*

AutoPap-GS ^a	Current practice		
	HGEA	Normal	Total
HGEA	27	4	31 ^b
Normal	5	4,208	4,213
Total	32	4,212	4,244

*P = 1.0.

^aAs determined by AutoPap-GS and discrepancy panel review.

^bExcludes 18 smears diagnosed as LGEA on current practice.

IV). AutoPap-GS identified an additional 31 low-grade smears, which is statistically significantly different to the detection rate for current practice (P < 0.0001).

AutoPap-GS was equivalent to current practice for the identification of high-grade and suspected high-grade smears (P = 1.0; Table V). AutoPap-GS detected four high-grade or suspected high-grade smears that were classified as normal by current practice, whereas there were five smears identified as high-grade or suspected high-grade by current practice that were classified as normal by AutoPap-GS.

Table VI summarizes the performance of AutoPap-GS in identifying smears unsatisfactory for assessment. AutoPap-GS detected 15 unsatisfactory smears, which were classified as satisfactory by current practice, whereas there were 32 smears detected as satisfactory by AutoPap-GS, which were classified as unsatisfactory by current practice. Therefore, the two study arms were not equivalent in identifying unsatisfactory smears (P = 0.019).

The two study arms were not equivalent for the detection of an endocervical component (P < 0.0001; Table VII).

Table VI. Result Matrix for Detection of an Unsatisfactory Smear (n = 67)*

AutoPap-GS ^a	Current practice		
	Unsatisfactory	Satisfactory	Total
Unsatisfactory	52	15	67
Satisfactory	32	4,493	4,525
Total	84	4,508	4,592

*P = 0.019.

^aAs determined by AutoPap-GS and discrepancy panel review.

Table VII. Result Matrix for the Presence or Absence of an Endocervical Component*

AutoPap-GS ^a	Current practice		
	Present	Absent	Total
Present	4,254	186	4,440
Absent	458	666	1,124
Total	4,712	852	5,564

*P < 0.0001.

^aAs determined by AutoPap-GS and discrepancy panel review.

Table VIII. Result Matrix for the Presence or Absence of an Organism or Infectious Agent

AutoPap-GS ^a	Current practice		
	Present	Absent	Total
Present	110	25	135
Absent	160	5,288	5,448
Total	270	5,313	5,583

*P < 0.0001.

^aAs determined by AutoPap-GS and discrepancy panel review.

AutoPap-GS failed to detect this component in 458 smears detected by current practice, whereas AutoPap-GS identified an endocervical component in 186 smears not detected by current practice.

The study population comprised 295 smears with evidence of organisms or an infectious agent. Fungal organisms morphologically consistent with *Candida* was found in 274 (92.9%) smears; *Trichomonas vaginalis*, 13 (4.4%) smears; cellular changes associated with herpes simplex virus, 3 (1.0%) smears; and bacteria morphologically consistent with *Actinomyces*, 5 (1.7%) smears. Table VIII summarizes the performance of AutoPap-GS in identifying the presence of an organism or infectious agent. AutoPap-GS detected 25 smears with organisms (*Candida*, 18 smears; *Trichomonas*, 3 smears; herpes simplex virus, 2 smears; and *Actinomyces*, 2 smears) that were not detected by current practice, whereas AutoPap-GS failed to detect organisms in 160 smears detected by current practice (*Candida*, 153 smears; *Trichomonas*, 4 smears; and *Actinomyces*, 3 smears). Therefore, the two study arms were not equivalent in identifying organisms or infectious agents (P < 0.0001).

The ranking of abnormal smears by the APSS is shown in Table IX. Of the 241 abnormal smears, classified as LGEA

Table IX. Triage of Abnormal Smears

Quintile	LGEA ⁺		HGEA	
1	104	(43.1%)	37	(57.8%)
2	52	(21.6%)	10	(15.6%)
3	37	(15.3%)	6	(9.4%)
4	32	(13.2%)	8	(12.5%)
5	16	(6.6%)	3	(4.7%)
Total	241	64		

or worse, 156 (64.7%) smears were ranked in the first and second quintile. Thirty-seven (15.3%) smears were ranked in the third quintile, 32 (13.2%) in the fourth quintile, and 16 (6.6%) in the fifth quintile. Forty-seven (73.4%) high-grade or suspected high-grade smears were ranked in the first and second quintile. Six (9.4%) smears were ranked in the third quintile, eight (12.5%) in the fourth quintile, and three (4.7%) in the fifth quintile.

The performance of the APSS in categorizing slides as NFR was also assessed. Of 986 smears classified as NFR, 975 were normal. The remaining 11 smears were subject to discrepancy review to determine the final diagnostic truth. These included seven smears reported as LGEA and one smear as inconclusive, possibly of high-grade. None of these smears showed evidence of disease progression at 6- and 12-month follow-up. Three smears were reported as unsatisfactory due to an inadequate squamous component and were found to be normal on repeat cytology. Forty-six smears showed evidence of an organism or infectious agent and this finding would have gone unreported if the laboratory had been using the software in routine cytology practice.

Discussion

The effectiveness of the APSS has been validated in numerous clinical trials with final FDA approval granted in 1998 for its intended use in the initial screening of cervical smears.¹ To improve laboratory accuracy and reduce workload, a location-guided method was developed as an additional feature of the primary screening software. Few independent studies have assessed this enhancement since the completion of initial feasibility studies by the company in 1997 and 1998.^{2,3} A subsequent laboratory-based study by Huang et al.⁴ demonstrated statistically superior performance of AutoPap-GS in the detection of abnormal slides; however, this study involved an enriched sample of abnormal smears not relevant to intended use of the device. The aim of the present study was to assess the effectiveness of AutoPap-GS compared to current practice.

Analysis of the results show that while AutoPap-GS is equivalent to current practice for the detection of high-grade and suspected high-grade abnormalities, the device performed less well in the detection of smears with a low-grade abnormality. A common feature of low-grade smears is the

lower prevalence of abnormal cells. For example, smears showing HPV changes may contain only a few scattered koilocytes, which may not be readily identified by the location software. Lee et al.² tested triage sensitivity over a range of different slide populations and found that sensitivities between low prevalence abnormal slides and the overall abnormal slide population was not statistically different. In a study of the AutoPap primary screening software without LGS, Alasio et al.⁹ observed decreased sensitivity for slides having a low prevalence of abnormal cells. It was postulated that low prevalence smears combined with inflammation causing a decrease in nuclear staining may influence the detection of cells by the device. These findings contrast with an earlier study assessing the detection capability of the device for smears with a low prevalence of abnormal cells.¹⁰ Further studies may be required to assess the sensitivity of AutoPap-GS for low prevalence abnormalities.

The inclusion of atypical squamous cells of undetermined significance (ASC-US) smears in the LGEA category may explain in part the poor results obtained for low-grade abnormalities. The AutoPap-GS is primarily designed for the detection of dysplastic abnormalities or worse. However, the diagnosis of ASC-US can be subjective and it was of interest to test the effectiveness of the device to detect these atypias. AutoPap-GS was not equivalent to current practice for the identification of unsatisfactory smears, an endocervical component or the presence of organisms or infective agents.

A smear is called unsatisfactory if an inadequate squamous-cell component is present or if a dense inflammatory infiltrate or excessive blood obscures cell detail. Review of FOV locations underestimated smear adequacy in a significant number of smears. Smear adequacy was often considered optimal on the basis of FOV review but subsequent FSR, as part of the discrepancy resolution process, indicated an inadequate smear. FOV review also underestimated the presence of an endocervical component and these findings highlight the importance of assessing all available information on a smear for a determination of both smear adequacy and the presence of an endocervical component.

AutoPap-GS was not equivalent to current practice for the identification of organisms or infective agents, although a more confident conclusion may be obtained by studying a larger sample. Detection of organisms by FOV review was favored if the agent was accompanied by cellular atypia, for example, herpes simplex, or for organisms eliciting an inflammatory cellular response. Detection by FOV review was further enhanced if the organism was widely distributed across the slide, for example, *Trichomonas vaginalis*. *Candida* and *Actinomyces* often manifest as a patchy presence on the slide and, in the absence of appreciable cellular changes, may not be identified in the FOV locations.

Slide ranking results indicate that the majority of high-grade and suspected high-grade smears were ranked in the first and second quintiles (73.4%), with the remaining smears distributed over the other quintiles, including the fifth. Similar results were obtained by previous investigators.^{9,11,12} In our study, a wider spread of smears across ranks was apparent for all abnormal smears, including those smears showing a low-grade abnormality. This may be related to a number of factors compromising the ability of the device to classify an abnormality accurately, for example, the presence of inflammation and/or a low prevalence of abnormal cells. The results of this study support the manufacturer's claim that abnormal smears are assigned a higher numerical score.

The performance of the APSS in identifying up to 25% of successfully processed slides as normal was assessed and results were found to be comparable to those of previous larger studies.^{1,11,12} With the exception of one case, reported as inconclusive, possibly high-grade abnormality, all severe abnormalities were selected by the instrument for review. Confirmation of an abnormality was not obtained on subsequent follow-up for the inconclusive smear or for the seven smears reported as LGEA. Three smears reported as unsatisfactory and 46 smears showing evidence of an organism or infective agent were included in the NFR subset. Although the Papanicolaou smear is primarily a test for precancerous changes, the identification of organisms is an add-on benefit with this information useful in the overall management of the patient.¹³

Location-guided screening has the potential to improve diagnostic accuracy and reduce laboratory workload. In our study, the device was effective in detecting high-grade and suspected high-grade smears but failed to perform as well with low-grade abnormalities, unsatisfactory smears, and the detection of infectious entities and an endocervical component. Slide ranking information, however, facilitated the triage of abnormal smears for prompt attention.

Due to the labor-intensive transfer of PapMaps to slides, the current system is difficult. The development of an automated microscope incorporating an electronic location

capability may improve diagnostic accuracy and further streamline laboratory operations.

Acknowledgments

The authors thank the laboratory clerical and technical staff for valuable assistance and TriPath Imaging for lending an AutoPap system for research purposes.

References

1. Wilbur DC, Prey MU, Miller WM, Pawlick GF, Colgan TJ. The AutoPap system for primary screening in cervical cytology: comparing the results of a prospective, intended-use study with routine manual practice. *Acta Cytol* 1998;214:214-220.
2. Lee JSJ, Kuan L, Oh S, Patten FW, Wilbur DC. A feasibility study of the AutoPap system location-guided screening. *Acta Cytol* 1998;42:221-226.
3. NeoPath AutoPap primary screening system. Location guided screening supplement: operator's manual no. 750-04512-01, rev. A. Burlington, NC: TriPath Imaging; 1999.
4. Huang TW, Lin TSM, Lee JSJ. Sensitivity studies of the AutoPap system location-guided screening of cervical-vaginal cytologic smears. *Acta Cytol* 1999;43:363-368.
5. Chang AR, Lin WF, Chang A, Chong KS. Can technology expedite the cervical cancer screening process? a Hong Kong experience using the autopap primary screening system with location-guided screening capability. *Am J Clin Pathol* 2002;117:437-443.
6. National Pathology Accreditation Advisory Council. Requirements for gynaecological (cervical) cytology. Canberra: Australian Government Publishing Service; 1997. p 6-7.
7. Kurman RJ, Solomon D. The Bethesda system for reporting cervical/vaginal cytologic diagnoses. New York: Springer-Verlag; 1994. p 12-29.
8. Fleiss JL. Statistical methods for rates and proportions, 2nd ed. New York: John Wiley and Sons; 1981. p 1-50.
9. Alasio LM, Alphantery C, Grassi P, et al. Performance of the AutoPap primary screening system in the detection of high-risk cases in cervicovaginal smears. *Acta Cytol* 2001;45:704-708.
10. Lee JSJ, Wilhelm P, Kuan L, et al. AutoPap system performance in screening for low prevalence and small cell abnormalities. *Acta Cytol* 1997;41:56-64.
11. Bibbo M, Hawthorne C, Zimmerman B. Does use of the AutoPap assisted primary screener improve cytologic diagnosis. *Acta Cytol* 1999;43:23-26.
12. Bibbo M, Zimmerman B. Performance of the AutoPap primary screening system at Jefferson University Hospital. *Acta Cytol* 1999;43:27-29.
13. Wilbur DC, Prey MU, Miller WM, et al. AutoPap system detection of infections and benign cellular changes: results from primary screener clinical trials. *Diagn Cytopathol* 1999;21:355-358.