Interlaboratory Quality Control in Gynecologic Cytopathology Using the Novel CONQUISTADOR Software

Interobserver Reproducibility in the Latin American Screening Study

M. Alderisio, M.D., **Margherita Branca,** M.D., Ph.D., M.I.A.C., **M. Eržen,** M.D., Ph.D., M.I.A.C., **A. Longatto-Filho,** M.Sc., Ph.D., P.M.I.A.C., **S. Derchain,** M.D., Ph.D., **S. Tatti,** M.D, Ph.D., **S. Vighi,** M.D., **C. Roteli-Martins,** M.D., Ph.D., **L. Leoncini,** M.A., **M. Y. S. Maeda,** M.Sc., **D. M. Montis, R. Gontijo,** M.D., **L. Sarian,** M.D., and **K. Syrjänen,** M.D., Ph.D., F.I.A.C.

Objective

Study Design

Results

Conclusion

Slide exchange and diagnosis comparison are the core components of the interlaboratory QC schemes to maintain and check diagnostic approach and establish standard consensus criteria through consensus opinion.

(Acta Cytol 2007;51:0000–0000)

Keywords: diagnostic accuracy index, kappa, quality control, Pap smear, performance indicators, reproducibility, software.

There is no doubt that organized screening programs are effective in reducing the incidence and

From the Unit of Cytopathology, National Center of Epidemiology, Surveillance and Promotion of Health, National Institute of Health (ISS), Rome, Italy; SIZE Diagnostic Center, Ljubljana, Slovenia; Instituto Adolfo Lutz, São Paulo, Brazil; Life and Health Sciences Research Institute, School of Health Sciences, University of Minho, Braga, Portugal; Universidade Estadual de Campinas, Campinas, Brazil; Gynecology Hospital de Clinicas, Buenos Aires, Argentina; Hospital Leonor M. de Barros, São Paulo, Brazil; and Department of Oncology and Radiotherapy, Turku University Hospital, Turku, Finland.

Dr. Alderisio is _____, Unit of Cytopathology, National Center of Epidemiology, Surveillance and Promotion of Health, National Institute of Health (ISS), Rome, Italy.

Dr. Branca is _____, SIZE Diagnostic Center, Ljubljana, Slovenia.

Dr. Eržen is _____, SIZE Diagnostic Center, Ljubljana, Slovenia.

Dr. Longatto-Filho is _____, Instituto Adolfo Lutz, Sao Paulo, Brazil, and _____, Life and Health Sciences Research Institute, School of Health Sciences, University of Minho, Braga, Portugal.

Dr. Derchain is Universidade Estadual de Campinas, Campinas, Brazil.

Dr. Tatti is First Chair, Gynecology Hospital de Clinicas, Buenos Aires, Argentina.

Dr. Vighi is _____, Gynecology Hospital de Clinicas, Buenos Aires, Argentina.

Dr. Roteli-Martins is _____, Hospital Leonor M. de Barros, São Paulo, Brazil.

__ Leoncini is _____, Unit of Cytopathology, National Center of Epidemiology, Surveillance and Promotion of Health, National Institute of Health (ISS), Rome, Italy.

mortality of cervical carcinoma (CC),[1-3] and the Pap test (when properly done at all steps), remains at the moment the only screening test with established cost-effectiveness.[1-11] Like any other diagnostic procedure, however, Pap smear screening can be compromised at several steps.[4,6,9] Diagnostic reproducibility (measured by κ and weighted κ) and high accuracy are major issues in cytopathology, as discussed in several recent studies.[4,6-29] Indeed, a systematic monitoring of every step of the screening procedure is a fundamental prerequisite to make the program a success.[1-4,6-9]

Our results indicate the necessity for the laboratories to test themselves, periodically analyzing sets of smears that include borderline and controversial lesions in order to increase expertise and achieve, in time, diagnostic reliability also for this kind of lesion, with the aim of continuous QI.

In Italy, these issues have been seriously considered as part of measures taken toward improvement of ongoing regional screening programs.[22-25] We recently developed 3 simple diagnostic variability indices (indices A, B and C) to provide a more easily interpretable measure of the consistency in cytodiagnosis of cervical intraepithelial neoplasia (CIN).[25] Until the recent development of new statistical software, CONQUISTADOR,[30] these indices were calculated manually, which limited their usefulness in the continuous quality monitoring of cytology laboratories. This new software facilitates laboratory quality control (QC) procedures by calculating all the required QC indicators, as recently demonstrated.[30]

To minimize the rate of false negative and false positive diagnoses, each cytopathology laboratory should actively pursue systematic intralaboratory QC programs and also participate in external interactive quality assurance schemes. Slide exchange and comparison of diagnosis is the core component of these interlaboratory QC measures, aimed at maintaining diagnostic standards, and establish consensus criteria through consensus opinion.[25]

As part of the ongoing Latin American Screening (LAMS) study,[31] we designed an external interlaboratory QC program comparing reproducibility of cytologic diagnoses issued by the participating cytology laboratories in Brazil and Argentina. The aim was to assess the diagnostic agreement between different laboratories examining 2 basically different sets of slides (clear-cut and difficult ones) and to what extent this reproducibility can be improved by participation in this type of interlaboratory QC approach. The new CONQUISTADOR software was tested for the first time in true clinical samples and laboratories instead of a simulation study reported before.[30]

For the availability of the free of charge software, see the special note at the end of the paper.

Materials and Methods

Participating Laboratories

This QC study was focused on interlaboratory comparison of cytologic diagnosis and realized by circulating carefully selected conventional Pap smear slide sets among 4 cytopathology laboratories, 3 of which examined the clinical samples derived from the LAMS study.[31] These 3 laboratories in Latin America are (1)

__ Maeda is _____, Instituto Adolfo Lutz, São Paulo, Brazil; Life and Health Sciences Research Institute, School of Health Sciences, University of Minho, Braga, Portugal.

__ Montis is _____, Universidade Estadual de Campinas, Campinas, Brazil.

Dr. Gontijo is _____, Universidade Estadual de Campinas, Campinas, Brazil.

Dr. Sarian is _____, Universidade Estadual de Campinas, Campinas, Brazil.

Dr. Syrjänen is _____, Unit of Cytopathology, National Center of Epidemiology, Surveillance and Promotion of Health, National Institute of Health (ISS), Rome, Italy, and _____, Department of Oncology and Radiotherapy, Turku University Hospital, Turku, Finland.

The Latin American Screening Study (LAMS) is funded by European Commission, INCO-DEV Contract #ICA4-CT-2001-10013.

Address correspondence to: Margherita Branca, M.D., Ph.D., M.I.A.C., Unità di Citoistopatologia, Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute, Istituto Superiore di Sanità, Viale Regina Elena, 299, I-00161 Roma, Italy (margherita.branca@fastwebnet.it).

Financial Disclosure: The authors have no connection to any companies or products mentioned in this article.

Received for publication July 11, 2006.

Accepted for publication October 9, 2006.

Istituto A. Lutz, São Paolo, Brazil; (2) Laboratory of Pathology, Universidade Estadual de Campinas, Brazil; and (3) Laboratory of Pathology, Universidad de Buenos Aires, Argentina. In addition, a large cytopathology laboratory in Europe, (4) the Center of Gynecological Cytopathology, Ljubljana (Slovenia) participated in this interlaboratory slide exchange program. All participating centers represent major cytopathology laboratories in university hospitals (2 and 3), in a large research institute (1) or a major private laboratory (4). All laboratories have a sizeable daily workload, and all agreed to participate on a voluntary basis. This program was coordinated by the Cytopathology Unit of the Italian National Institute of Health (Istituto Superiore di Sanità) in Rome.

Confidentiality

Strict anonymity was guaranteed for each participating laboratory. An identification code known by only the study coordinator (M.B.) was attributed to each participating laboratory to enable the recording of their responses completely anonymously.

Selection and Circulation of the Slides

The study design and methods of this study followed the strategy used in our previous study[25] of similar nature that was completed in Italy a few years ago. Instead of using a single set of slides as before,[25] we designed 5 sets (A–E) of slides, including both "clear-cut" smears (as used in certification tests) (sets A–D) and an additional set (E) of "difficult smears," all being positive and also containing borderline lesions and glandular lesions. All slides used were of proven high technical quality.

Building Up the 5 Slide Sets

As the first step, all participating laboratories were asked to provide the study coordinators a series of conventional Pap smears, including all diagnostic categories of the 2001 Bethesda System (TBS 2001). Altogether, we received a pool of 250 such slides from 3 participating laboratories (laboratories 1, 2 and 3). In the coordinator's laboratory, 2 cytopathologists (M.B., M.A.) first examined the total pool of slides, selecting 80 smears with clear-cut diagnosis. These 80 slides were viewed by the third cytopathologist possessing the F.I.A.C. degree (K.S.), who suggested some changes to complete the final selection of the 80 cases, of which a consensus diagnosis was reached.

The 80 slides were divided into 4 sets (A, B, C, and D) of 20 smears, each including in different proportions the 6 diagnostic categories of TBS 2001, except atypical squamous cells (ASCs) and atypical glandular cells (AGCs), to keep the sets including clear-cut cases only. These diagnostic categories are: inadequate smears, negative for intraepithelial lesions or malignancy (infection and reactive and reparative changes), low-grade squamous intraepithelial lesion (LSIL) or CIN 1 (associated or not with human papillomavirus (HPV), high-grade squamous intraepithelial lesion (HSIL) or CIN 2 (associated or not with HPV), HSIL or CIN 3 (associated or not with HPV), and invasive cancer (invasive squamous cancer, adenocarcinoma).

All slides were provided with a new label and a new identification code, wiping out the originals. A reporting form was designed, accompanied by a cover letter with detailed instructions for the participants on how to assess and report the slides. The only clinical information disclosed was the age of the patients. The database was built at National Center of Epidemiology, Surveillance and Promotion of Health, National Institute of Health (ISS), Rome, Italy, to record the responses from the 4 laboratories at various steps of the study.

As the second step, the fifth set (set E) of slides was compiled by the coordinators, using the same approach as in building up the first 4 sets (A–D). From the pool of 250 slides, 2 cytopathologists (M.B., M.A.) selected 20 cases, all being positive and also including borderline lesions, covering ASC, AGC, LSIL or CIN 1 (associated or not with HPV), HSIL or CIN 2 (associated or not with HPV), HSIL or CIN 3 (associated or not with HPV) and invasive cancer (invasive squamous cancer, adenocarcinoma). This selection was reviewed by the third examiner (K.S.), completing the process to reach the consensus diagnosis (used as the reference diagnosis). As the last step, slide set E was circulated among the 4 laboratories, and instructions were given to complete the evaluation as for the sets A–D. In addition, they were asked to fill in a standardized form, containing the list of diagnoses and also a list of recommendations for management, as well as giving their judgment of the diagnostic difficulty of each individual case.

Data Analysis Using CONQUISTADOR Software

As recently reported in this journal,[30] ISS scientists and their collaborators developed a new statistical software (CONQUISTADOR), specifically designed to record cytology data in the routine laboratory and capable of calculating the necessary performance indicators (sensitivity, specificity, PPV, NPV), diagnostic accuracy indices,[25] as well as both normal and weighted κ values. This software is available in 2 versions: the basic version with intralaboratory procedures and an advanced version containing the interlaboratory procedures. Despite these different functions and purpose of use, the basic architecture and platform of these 2 versions is similar.[30]

The basic functions of the CONQUISTADOR

Table I *Summary of Analysis Accuracy for Slide Sets A–D (Clear-Cut Cases)*

				CI	95% CI
No. of slides	73	No. of laboratories	4		
No of reports	272	Mean reports for slide	3.7		
A (True positives)	129	Sensitivity	92.1	86.4–96.0	.004
D (True negatives)	113	Specificity	85.6	78.4–91.1	.005
C (False negatives)	11	Percent false negatives	7.9	4.0–13.6	.001
B (False positives)	19	Percent false positives	14.4	8.9–21.6	.000
		Positive predictive value	87.2	80.7–92.1	.004
		Negative predictive value	91.1	84.7–95.5	.003

software have been previously described.[30] For the basic statistical analysis, all 100 initial diagnoses recorded from the 4 laboratories on the report forms were divided into 6 groups: 0 = diagnosis not made because of inadequate sampling; 1 = within normal limits, infection, and reactive and reparative changes; 2 = atypical squamous cells of undetermined significance (ASCUS) and atypical glandular cells of undetermined significance (AGC-US), LSIL or CIN 1 (associated or not with human papillomavirus (HPV); 3 = CIN 2 (associated or not with HPV); 4 = CIN 3 (associated or not with HPV); 5 = 3 and 4 taken together as HSIL; and 6 = invasive carcinoma (squamous cell carcinoma, adenocarcinoma). Sampling adequacy was limited to 2 categories: adequate and inadequate according to TBS 2001. Recommendations were grouped into 3 categories: 1 = no special recommendation; 2 = early repeated smear; and 3 = colposcopy or biopsy.

The simple and weighted κ for individual laboratories, diagnostic categories, management recommendations and judgment of diagnostic difficulty were computed by the CONQUISTADOR software, using the algorithm by Fleiss and Cohen, including the 95% CI.[29,30] In addition, the 3 diagnostic variability indices were calculated, as previously described[25]: index A concerns variability between CIN 1 plus HPV vs. CIN 2; index B: variability between CIN 2 plus CIN 3; index C: variability between CIN 1 vs. CIN 2 + CIN 3 + invasive cancer.[25,30]

Results

All analyses were performed separately for the "clear-cut" samples (slide sets A, B, C and D and the "difficult smears" (slide set E) and presented separately here as well.

Table I summarizes the accuracy analysis of the slide sets A–D. For all laboratories together, sensitivity was 92.1% (95% CI 86.4–96.0) and specificity

Table II *Interlaboratory Agreement in Slide Sets A–D Measured by κ Values*

Comparison	Sample no.	Raw agree	Chance agree	κ Simple	95% CI (κ)	κ Weighted	95% CI (κ_W)
LABS vs. all LABS	6	P0 = 0.67	PE = 0.34	$\kappa = 0.50$	$0.44 < \kappa < 0.56$	$\kappa_W = 0.79$	$0.73 < \kappa_W < 0.84$
LAB02 vs. all LAB	3	P0 = 0.65	PE = 0.33	$\kappa = 0.50$	$0.41 < \kappa < 0.59$	$\kappa_W = 0.82$	$0.75 < \kappa_W < 0.89$
LAB03 vs. all LAB	3	P0 = 0.67	PE = 0.34	$\kappa = 0.50$	$0.42 < \kappa < 0.59$	$\kappa_W = 0.71$	$0.60 < \kappa_W < 0.81$
LAB04 vs. all LAB	3	P0 = 0.68	PE = 0.34	$\kappa = 0.52$	$0.43 < \kappa < 0.60$	$\kappa_W = 0.74$	$0.65 < \kappa_W < 0.83$
LAB05 vs. all LAB	3	P0 = 0.67	PE = 0.34	$\kappa = 0.50$	$0.41 < \kappa < 0.58$	$\kappa_W = 0.82$	$0.75 < \kappa_W < 0.88$
LAB02 vs. other LABS	3	P0 = 0.65	PE = 0.33	$\kappa = 0.50$	$0.41 < \kappa < 0.59$	$\kappa_W = 0.82$	$0.75 < \kappa_W < 0.89$
LAB02 vs. LAB03	68	P0 = 0.63	PE = 0.33	$\kappa = 0.45$	$0.29 < \kappa < 0.61$	$\kappa_W = 0.56$	$0.34 < \kappa_W < 0.79$
LAB02 vs. LAB04	61	P0 = 0.61	PE = 0.33	$\kappa = 0.42$	$0.25 < \kappa < 0.58$	$\kappa_W = 0.61$	$0.40 < \kappa_W < 0.81$
LAB02 vs. LAB05	65	P0 = 0.72	PE = 0.32	$\kappa = 0.60$	$0.46 < \kappa < 0.73$	$\kappa_W = 0.88$	$0.81 < \kappa_W < 0.96$
LAB03 vs. other LABS	3	P0 = 0.66	PE = 0.34	$\kappa = 0.49$	$0.41 < \kappa < 0.56$	$\kappa_W = 0.59$	$0.48 < \kappa_W < 0.70$
LAB03 vs. LAB02	68	P0 = 0.63	PE = 0.33	$\kappa = 0.45$	$0.29 < \kappa < 0.61$	$\kappa_W = 0.56$	$0.34 < \kappa_W < 0.79$
LAB03 vs. LAB04	67	P0 = 0.76	PE = 0.34	$\kappa = 0.64$	$0.50 < \kappa < 0.78$	$\kappa_W = 0.82$	$0.68 < \kappa_W < 0.96$
LAB03 vs. LAB05	72	P0 = 0.63	PE = 0.35	$\kappa = 0.42$	$0.28 < \kappa < 0.56$	$\kappa_W = 0.59$	$0.39 < \kappa_W < 0.78$
LAB04 vs. other LABS	3	P0 = 0.68	PE = 0.34	$\kappa = 0.52$	$0.43 < \kappa < 0.60$	$\kappa_W = 0.74$	$0.65 < \kappa_W < 0.83$
LAB04 vs. LAB02	61	P0 = 0.61	PE = 0.33	$\kappa = 0.42$	$0.25 < \kappa < 0.58$	$\kappa_W = 0.61$	$0.40 < \kappa_W < 0.81$
LAB04 vs. LAB03	67	P0 = 0.76	PE = 0.34	$\kappa = 0.64$	$0.50 < \kappa < 0.78$	$\kappa_W = 0.82$	$0.68 < \kappa_W < 0.96$
LAB04 vs. LAB05	66	P0 = 0.65	PE = 0.35	$\kappa = 0.46$	$0.31 < \kappa < 0.61$	$\kappa_W = 0.71$	$0.56 < \kappa_W < 0.86$
LAB05 vs. other LABS	3	P0 = 0.68	PE = 0.34	$\kappa = 0.54$	$0.47 < \kappa < 0.61$	$\kappa_W = 0.59$	$0.48 < \kappa_W < 0.70$
LAB05 vs. LAB02	65	P0 = 0.72	PE = 0.32	$\kappa = 0.60$	$0.46 < \kappa < 0.73$	$\kappa_W = 0.88$	$0.81 < \kappa_W < 0.96$
LAB05 vs. LAB03	72	P0 = 0.63	PE = 0.35	$\kappa = 0.42$	$0.28 < \kappa < 0.56$	$\kappa_W = 0.59$	$0.39 < \kappa_W < 0.78$
LAB05 vs. LAB04	66	P0 = 0.65	PE = 0.35	$\kappa = 0.46$	$0.31 < \kappa < 0.61$	$\kappa_W = 0.71$	$0.56 < \kappa_W < 0.86$

Table III *Global Agreement in Slide Sets A–D Evaluated by Simple and Weighted κ*

No. of slides	97	No. of laboratories	4
No. of reports	362	Mean no. of reports per slide	3.73
P0 (raw agree)	0.67	PE (chance agree)	0.33
κ Simple	0.51	95% CI for κ	$0.46 < \kappa < 0.56$
κ Weighted	0.80	95% CI for κ_W	$0.76 < \kappa_W < 0.85$

85.6% (95% CI 78.4–91.1). The PPV was 87.2% and NPV was 91.1%. Seven cases were not registered as inadequate.

Agreement between individual laboratories in slide sets A–D is summarized in Table II, giving both simple and weighted κ values for these interlaboratory comparisons. The global summary of this agreement is given in Table III. Agreement is moderate using simple κ: $\kappa = 0.50$ (95% CI 0.44–0.56), but substantial (to almost perfect) using the weighted κ ($\kappa_w = 0.80$) (95% CI 0.73–0.84).

Table IV *Main Diagnoses and Diagnostic Variability Indices Calculated for Slide Sets A–D*

Slide code	Other diagnoses[a]	HPV	CIN 1	CIN 2	CIN 3	CC	Total	A	A1	B	B1	C	C1
6A	0	4	0	0	0	4	1.00	1.00			1.00	1.00	
7A	0	0	0	0	4	4			1.00	1.00	1.00	1.00	
8A	0	0	0	1	3	4			1.00	1.00	1.00	1.00	
9A	0	3	1	0	0	4	0.75	0.50	1.00	1.00	0.75	0.50	
10A	0	4	0	0	0	4	1.00	1.00			1.00	1.00	
11A	0	0	2	1	0	3	1.00	1.00	0.67	0.33	1.00	1.00	
12A	0	0	0	3	1	4			1.00	1.00	1.00	1.00	
13A	2	1	0	0	0	3	1.00	1.00			1.00	1.00	
17A	0	0	3	1	0	4	1.00	1.00	0.75	0.50	1.00	1.00	
18A	0	3	1	0	0	4	0.75	0.50	1.00	1.00	0.75	0.50	
20A	0	0	0	0	3	3					1.00	1.00	
1B	2	0	0	2	0	4			1.00	1.00	0.50	0.00	
2B	1	0	0	1	2	4			1.00	1.00	0.75	0.50	
3B	0	0	2	2	0	4	1.00	1.00	0.50	0.00	1.00	1.00	
4.B	1	1	2	0	0	4	0.67	0.33	1.00	1.00	0.50	0.00	
6B	0	2	1	1	0	4	0.67	0.33	0.50	0.00	0.50	0.00	
7B	1	1	0	1	1	4	1.00	1.00	1.00	1.00	0.50	0.00	
8B	2	2	0	0	0	4	1.00	1.00			1.00	1.00	
9B	3	1	0	0	0	4	1.00	1.00			1.00	1.00	
10B	1	1	0	0	0	2	1.00	1.00			1.00	1.00	
11B	2	1	0	1	0	4	1.00	1.00	1.00	1.00	0.75	0.50	
12B	0	2	0	1	1	4	1.00	1.00	1.00	1.00	0.50	0.00	
13B	2	1	0	0	1	4	1.00	1.00			0.75	0.50	
15B	2	0	0	2	0	4			1.00	1.00	0.50	0.00	
16B	2	2	0	0	0	4	1.00	1.00			1.00	1.00	
17B	0	3	0	1	0	4	1.00	1.00	1.00	1.00	0.75	0.50	
18B	0	2	0	0	2	4	1.00	1.00			0.50	0.00	
19B	0	2	0	0	2	4	1.00	1.00			0.50	0.00	
20B	0	0	0	0	4	4					1.00	1.00	
5C	0	0	0	2	2	4			1.00	1.00	1.00	1.00	
6C	2	0	0	0	1	3					0.67	0.33	
7C	0	2	1	1	0	4	0.67	0.33	0.50	0.00	0.50	0.00	
9C	3	1	0	0	0	4	1.00	1.00			1.00	1.00	
10C	0	2	0	2	0	4	1.00	1.00	1.00	1.00	0.50	0.00	
17C	1	0	0	1	2	4			1.00	1.00	0.75	0.50	
6D	0	0	0	1	3	4			1.00	1.00	1.00	1.00	
8D	2	2	0	0	0	4	1.00	1.00			1.00	1.00	
9D	0	0	0	4	0	4			1.00	1.00	1.00	1.00	
10D	3	1	0	0	0	4	1.00	1.00			1.00	1.00	
12D	1	0	0	0	2	3					0.67	0.33	
18D	0	0	0	0	4	4					1.00	1.00	
19D	0	0	0	3	1	4			1.00	1.00	1.00	1.00	
20D	1	2	1	0	0	4	0.67	0.33	1.00	1.00	0.75	0.50	
Mean index values							43	0.90	0.80	0.87	0.74	0.82	0.64

[a]Negative for intraepithelial lesion or malignancy but not inadequate smears.
CC = cervical carcinoma, HPV = human papillomavirus.

Table V *Summary of Accuracy Analysis for Slide Set E (Difficult Cases)*

				CI	% CI
No. of slides	20	No. of laboratories	4		
No. of reports	77	Mean no. of reports per slide	3.9		
A (true positives)	57	Sensitivity	74.0	62.8–83.4	95.003
D (true negatives)	0	Specificity	0.0	0.0–0.0	95.000
C (galse negatives)	20	Percent false negatives	26.0	16.6–37.2	95.003
D (false positives)	0	Percent false positives	0.0	0.0–0.0	95.000
		Positive predictive value	100.0	93.7–100.0	95.001
		Negative predictive value	0.0	0.0–16.8	94.998

The diagnostic variability indexes for individual slides in sets A–D are listed in Table IV. The indices A, B and C range from 1 (if all laboratories make the same diagnosis) to 0.5 (if half of the laboratories make 1 diagnosis and half the other 1) and computed using the following equations: A1 = 1 – 2(1 – A), B1 = 1 – 2(1 – B) and C1 = 1 – 2(1 – C). The mean of the index A was 0.90, index B 0.87 and index C (important variability) was 0.82. For these clear-cut sets, the mean A1 index was 0.80, the mean B1 index 0.74 and the mean C1 index was 0.64 (Table IV).

Table V summarizes the accuracy analysis of the slide set E ("difficult slides"). Performance indicators (sensitivity [SE], specificity [SP], positive predictive value [PPV], negative predictive value [NPV]) were markedly lower compared with those calculated for the clear-cut cases in Table I. The PPV, however, was 100%.

Agreement between individual laboratories in slide set E is summarized in Table VI, giving both simple and weighted κ values for these interlaboratory comparisons. The global summary of this agreement is given in Table VII. Agreement is moderate using simple κ: κ = 0.45 (95% CI 0.34–0.57) and does not markedly improve using the weighted κ (κ_w = 0.53) (95% CI 0.36–0.69).

The diagnostic variability indexes for individual slides in set E are listed in Table VIII. The mean of the index A was 0.95, index B 1.00 and index C (important variability) was 0.80. For these difficult cases, the mean A1 index was 0.90, the mean B1 index 1.00 and the mean C1 index was 0.61 (Table VIII).

Finally, all 5 sets were analyzed together. The performance indicators for the 5 sets are summarized in Table IX. Sensitivity was 85.7% (95% CI 80.3–90.1) and specificity 85.6% (95% CI 78.4–91.1). The interlaboratory reproducibility for all slide sets is shown in Table X, where simple κ is 0.51 (95% CI 0.46–0.56)

Table VI *Interlaboratory Agreement in Slide Set E Measured by κ Values*

Comparison	Sample no.	Raw agree	Chance agree	κ Simple	95% CI (κ)	κ Weighted	95% CI (κ_W)
LABS vs. all LABS	6	P0 = 0.67	PE = 0.33	κ = 0.45	0.34 < κ < 0.57	κ_W = 0.53	0.36 < κ_W < 0.69
LAB02 vs. all LAB	3	P0 = 0.75	PE = 0.32	κ = 0.52	0.35 < κ < 0.69	κ_W = 0.69	0.45 < κ_W < 0.93
LAB03 vs. all LAB	3	P0 = 0.63	PE = 0.34	κ = 0.46	0.31 < κ < 0.62	κ_W = 0.26	0.01 < κ_W < 0.52
LAB04 vs. all LAB	3	P0 = 0.75	PE = 0.32	κ = 0.52	0.35 < κ < 0.69	κ_W = 0.69	0.45 < κ_W < 0.93
LAB05 vs. all LAB	3	P0 = 0.55	PE = 0.33	κ = 0.32	0.17 < κ < 0.48	κ_W = 0.45	0.23 < κ_W < 0.67
LAB02 vs. other LABS	3	P0 = 0.75	PE = 0.33	κ = 0.52	0.35 < κ < 0.69	κ_W = 0.69	0.45 < κ_W < 0.93
LAB02 vs. LAB03	19	P0 = 0.68	PE = 0.32	κ = 0.53	0.28 < κ < 0.79	κ_W = 0.46	–0.01 < κ_W < 0.92
LAB02 vs. LAB04	19	P0 = 1.00	PE = 0.32	κ = 1.00	0.55 < κ < 1.00	κ_W = 1.00	0.55 < κ_W < 1.00
LAB02 vs. LAB05	18	P0 = 0.56	PE = 0.31	κ = 0.35	0.09 < κ < 0.61	κ_W = 0.64	0.28 < κ_W < 1.00
LAB03 vs. other LABS	3	P0 = 0.63	PE = 0.34	κ = 0.46	0.31 < κ < 0.62	κ_W = 0.26	0.01 < κ_W < 0.52
LAB03 vs. LAB02	19	P0 = 0.68	PE = 0.32	κ = 0.53	0.28 < κ < 0.79	κ_W = 0.46	–0.01 < κ_W < 0.92
LAB03 vs. LAB04	19	P0 = 0.68	PE = 0.53	κ = 0.64	0.28 < κ < 0.79	κ_W = 0.46	–0.01 < κ_W < 0.92
LAB03 vs. LAB05	19	P0 = 0.53	PE = 0.25	κ = 0.42	–0.06 < κ < 0.56	κ_W = –0.03	–0.43 < κ_W < 0.38
LAB04 vs. other LABS	3	P0 = 0.75	PE = 0.32	κ = 0.52	0.35 < κ < 0.69	κ_W = 0.69	0.45 < κ_W < 0.93
LAB04 vs. LAB02	19	P0 = 1.00	PE = 0.32	κ = 1.00	0.55 < κ < 0.58	κ_W = 1.00	0.55 < κ_W < 1.00
LAB04 vs. LAB03	19	P0 = 0.68	PE = 0.53	κ = 0.64	0.28 < κ < 0.79	κ_W = 0.46	–0.01 < κ_W < 0.92
LAB04 vs. LAB05	18	P0 = 0.56	PE = 0.31	κ = 0.35	0.09 < κ < 0.61	κ_W = 0.64	0.28 < κ_W < 1.00
LAB05 vs. other LABS	3	P0 = 0.55	PE = 0.33	κ = 0.32	0.17 < κ < 0.48	κ_W = 0.45	0.23 < κ_W < 0.67
LAB05 vs. LAB02	18	P0 = 0.56	PE = 0.31	κ = 0.35	0.09 < κ < 0.61	κ_W = 0.64	0.28 < κ_W < 1.00
LAB05 vs. LAB03	19	P0 = 0.53	PE = 0.25	κ = 0.42	–0.06 < κ < 0.56	κ_W = –0.03	–0.43 < κ_W < 0.38
LAB05 vs. LAB04	18	P0 = 0.56	PE = 0.31	κ = 0.35	0.09 < κ < 0.61	κ_W = 0.64	0.28 < κ_W < 1.00

Table VII *Global Agreement in Slide Set E Evaluated by Simple and Weighted κ*

No. of slides	20	No. of laboratories	4
No. of reports	77	Mean no. of reports per slide	3.85
P0 (raw agree)	0.67	PE (chance agree)	0.33
κ Simple	0.45	95% CI for κ	$0.34 < \kappa < 0.57$
κ Weighted	0.53	95% CI for κ_W	$0.36 < \kappa_W < 0.69$

and weighted κ = 0.80 (95% CI 0.76–0.85), which is the lower limit of almost perfect agreement. The diagnostic variability indices for the entire sets of slides are listed in Table XI, all exceeding 0.80, except CI index, which is 0.63.

Discussion

Issues related to QC in cytopathology laboratories have attracted considerable interest in the recent literature.[4,6,7-29] To minimize the rate of false negatives and false positives (the 2 fundamental diagnostic errors), it is necessary that cytopathology laboratories actively pursue different intralaboratory QC measures and also participate in external QC programs. In a previous paper,[25] we listed several features as essential elements in these interlaboratory QC measures. Slide exchange and diagnosis comparison is the core component of the interlaboratory QC schemes to maintain and check diagnostic approach and establish standard consensus criteria through consensus opinion. To be effective and adequate, this necessitates a setup of appropriate software for filing data and calculation of accuracy and reliability measures. Such calculations should include statistical analysis of the reliability and accuracy, using the reference diagnosis as baseline.[25,30]

Until recently, a single software program that could compute all the necessary indices, performance indicators and reproducibility tests was unavailable. To fill this gap, we recently developed new statistical software (CONQUISTADOR) that was originally tested for performance in a simulation study using virtual laboratories and virtual samples.[30] The program was developed in a standard Microsoft environment and operates with Windows '98 (Microsoft Inc., Redmond, Washington, U.S.A.) and later systems. It may be used on a single personal computer (workstation mode) and in a client and server environment. The software interfaces with Microsoft Office and, in particular, with Excel procedures (used for data import and export and printing) and with the Microsoft Access database, which the software accesses with standard Structured Query Language (SQL), via a fast, efficient Active Data Project (ADO) Provider.[30] The software also delivers full compatibility in the alternative use of the Microsoft SQL Server relational database, whose use, on the other hand, would be suitable only in the case of massive applications, which are mainly of the interlaboratory type. In addition to all performance indicators (SE, SP, PPV and NPV) and their 95% CI, this new software also calculates the recently introduced diagnostic accuracy indices,[25] as well as normal and weighted κ values.[29,30] In the present study, this software was used, for the first time, in analysis of true clinical samples and using real laboratories as raters to test the interlaboratory reproducibility of conventional Pap smears.

Indeed, diagnostic reproducibility and accuracy are the 2 major issues in cytopathology and can be addressed by a number of procedures and programs for quality assurance (QA). Practical and theoretical limitations, as well as benefits of external quality assurance

Table VIII *Main Diagnoses and Diagnostic Variability Indices Calculated for Slide Set E*

Slide code	Other diagnoses[a]	HPV CIN 1	CIN 2	CIN 3	CC	Total	A	A1	B	B1	C	C1
2E	1	0	0	0	3	4					0.75	0.50
3E	1	2	1	0	0	4	0.67	0.33	1.00	1.00	0.75	0.50
4E	0	3	0	1	0	4	1.00	1.00	1.00	1.00	0.75	0.50
6E	1	3	0	0	0	4	1.00	1.00			1.00	1.00
8E	0	4	0	0	0	4	1.00	1.00			1.00	1.00
12E	1	3	0	0	0	4	1.00	1.00			1.00	1.00
13E	1	0	0	0	1	2					0.50	0.00
15E	0	0	0	2	2	4			1.00	1.00	1.00	1.00
16E	0	2	0	2	0	4	1.00	1.00	1.00	1.00	0.50	0.00
17E	0	3	0	1	0	4	1.00	1.00	1.00	1.00	0.75	0.50
19E	1	0	0	0	3	4					0.75	0.50
20E	1	0	0	1	2	4			1.00	1.00	0.75	0.50
Mean indexes values						12	0.95	0.90	1.00	1.00	0.80	0.61

[a]Negative for intraepithelial lesion or malignancy but not inadequate smears.
CC = cervical carcinoma, HPV = human papillomavirus.

Table IX *Summary of the Accuracy Analysis for All 5 Slide Sets (A–E)*

				CI	95% CI
No. of slides	93	No. of laboratories	4		
No. of reports	349	Mean no. of reports per slide	3.8		
A (True positives)	186	Sensitivity	85.7	80.3–90.1	.003
D (True negatives)	113	Specificity	85.6	78.4–91.1	.005
C (False negatives)	31	Percent false negatives	14.3	9.9–19.7	.003
B (False positives)	19	Percent false positives	14.3	8.9–31.6	.000
		Positive predictive value	90.7	85.9–94.3	.003
		Negative predictive value	78.5	70.9–84.9	.004

(EQA) schemes for cervical cytology, have been extensively discussed in the recent literature.[4,6-8,31-40] Several field studies have consistently shown that EQA settings improve the quality of diagnostic performance and enable participating laboratories to detect major deficiencies in specific areas of cervical cytology practice.[32-40] On one hand, data provided by performance evaluation systems are subject to potential bias.[32,34,36] According to EQA principles, any EQA system for cytology laboratories should assess the performance of each participant individually, as well as that of the laboratory as a whole.[37] Circulated slide sets among the laboratories is the core of the interlaboratory QC and quality improvement (QI), and it is the only method to evaluate interobserver and interlaboratory reproducibility.[17,20,23,24,39,40]

In the present study, we used slide sets at different levels of diagnostic difficulty: 4 sets of clear-cut cases (n = 80) and 1 set of difficult cases (n = 20). When the interlaboratory agreement in these 2 levels was compared, not unexpectedly, both regular κ and, particularly, weighted κ were markedly lower in the difficult cases slide set (set E), compared with the routine cases (sets A–D) (Tables III and VII). This is not unexpected, because the former set included only positive cases, with all categories and borderline smears as well. The κ statistic is considered the most rigorous and reliable index, because it takes into account the extent of random agreement. It can be computed in 2 forms, simple κ and weighted κ.[27] The weighted κ, which can be used in cases in which the diagnostic categories can be ranked in an ordinal scale, takes into account not only the presence of disagreement but also its extent; that is, it gives more weight to a disagreement between CIN 1 and invasive cancer than between CIN 1 and CIN 2.[28,29] Taking into consideration the diagnostic difficulty of set E, the weighted κ of 0.53, which implicates a moderate agreement, must be considered a good achievement (Table VII). For the entire series of 5 slide sets, the weighted κ was 0.8 (Table X), which is the lower limit of the almost perfect ranking, indicating excellent interlaboratory agreement among these 4 diagnostic laboratories.

For these same reasons, the conventional performance indicators (SE, SP, PPV, and NPV) were also different in the 2 diagnostic difficulty categories (Tables I and V). In a meta-analysis of 62 studies concerning the performance of cytology toward histology, SE and SP ranged from 11% to 99% and 14% to 97%, respectively.[38] Overall, the performance of Pap smears in this study, as shown by the rate of false negatives and positives (SE 85.7% and SP 85.6%), was somewhat better than that reported in the literature.[38]

The newly introduced diagnostic accuracy indices[25] were not remarkably different between the sets A–D and set E (Tables IV and VIII). As previously explained, these indices describe the ability to differentiate between the different diagnostic categories.[25] In this study, the ability to discriminate reliably between CIN 2 and CIN 3 was high (mean variability of index B was 0.89). Recently, Husain et al[36] found a crude agreement of 87% and a nonweighted κ of 0.79 in a comparison of the reporting patterns of 5 laboratories using 3 main diagnostic categories (benign, CIN, and malignant).[36] In our approach, using the diagnostic categories of TBS 2001, including ASC and AGC, a

Table X *Global Agreement in All Slide Sets (A–E) Evaluated by Simple and Weighted κ*

No. of slides	97	No. of laboratories	4
No. of reports	362	Mean no. of reports per slide	3.73
P0 (raw agree)	0.67	PE (chance agree)	0.33
κ Simple	0.51	95% CI for κ	$0.46 < \kappa < 0.56$
κ Weighted	0.80	95% CI for κ_W	$0.76 < \kappa_W < 0.85$

Table XI *Diagnostic Variability Indices Calculated for All Slide Sets (A–E)*

No. of slides	55	No. of laboratories	4
No. of reports	211	Mean no. of reports per slide	3.84
Mean A index	0.91	Mean A1 index	0.83
Mean B index	0.89	Mean B1 index	0.78
Mean C index	0.82	Mean C1 index	0.63

crude agreement of 0.67% was obtained among the 4 laboratories, with nonweighted κ of 0.45 (Table VII). For these difficult cases, this reflects good diagnostic accuracy, which was not markedly better for routine cases (Table III) when nonweighted κ was used (0.51).

In evaluating the results of studies like this, it should be borne in mind that cytology and histology, like most diagnostic technologies, are subject to considerable intraobserver and interobserver variation. This is shown, for example, by the wide variation of the diagnostic categories of TBS among different laboratories.[25,39] Alarmingly, interobserver agreement as low as 35% has been obtained among 5 experienced cytopathologists assessing 20 smears using TBS.[19] Compared with that, the results of the present exercise are significantly better, however. For the entire set of 100 slides (including both the clear-cut and difficult cases), the raw agreement was 67%, with regular κ = 0.51 and weighted κ = 0.80. The latter, in particular, indicates that the interlaboratory agreement among these 4 diagnostic laboratories is substantial (in fact, varying from substantial to almost perfect).

In conclusion the present study shows that it is possible to achieve high interlaboratory reproducibility in the assessment of slide sets similarly as is regularly used in aptitude or proficiency tests,[41] if TBS 2001 criteria are strictly followed. Our results also indicate the necessity for the laboratories to test themselves, periodically analyzing sets of smears that include borderline and controversial lesions in order to increase expertise and achieve, in time, diagnostic reliability also for this kind of lesion, with the aim of continuous QI. It must be stressed finally that the expertise of the cytopathologist plays a fundamental role.

Acknowledgments

This study has been supported by the European Commission, INCO-DEV Programme (Contract ICA4-CT-2001-10013) for the LAMS study, which is gratefully acknowledged.

References

1. IARC Working Group on Cervical Screening and UICC Project Group on Evaluation of Screening Programs for Cancer of the Uterine Cervix. In Screening for Cancer of the Uterine Cervix. Edited by M Hakama, AB Miller, NE Day. Lyon, IARC Press, 1986, pp 147–160
2. Franco E, Syrjänen K, de Wolff C, Patnick J, Ferenczy A, McGoogan E, Bosch X, Singer A, Munoz N, Meheus A, Monsonego J: New developments in cervical cancer screening and prevention. Cancer Epidemiol Biomarker Prev 1996;5:853–856
3. Miller AB, Nazeer S, Fonn S, Brandup-Lukanow A, Rehman R, Cronje H, Sankaranarayanan R, Koroltchouk V, Syrjänen K, Singer A, Onsrud M: Report on Consensus conference on Cervical Cancer Screening and Management. Int J Cancer 2000; 86:440–447
4. Koss LG: The Papanicolaou test for cervical cancer detection: A triumph and a tragedy. JAMA 1989;261:737–743
5. Koss LG: Diagnostic accuracy in cervico-vaginal cytology. Arch Pathol Lab Med 1993;117:1240–1242
6. Coleman D, Day N, Douglas G, Farmery E, Lynge E, Philip J, Ponti AN, Ronco G, Segnan S: European guidelines for quality assurance in cervical cancer screening. Eur J Cancer Suppl 4 1993;29A:S1–S38
7. Wied GL, Keebler CM, Rosenthal DL, Schenck V, Somrak TM, Vooijs GP (editors): Compendium on quality assurance, proficiency testing and workload limitations in clinical cytology. *In* Tutorials of Cytology. Chicago, 1995, pp 1–356
8. Syrjänen KJ: Quality assurance in the cytopathology laboratories of the Finnish Cancer Society. *In* Compendium of Quality Assurance, Proficiency Testing and Workload Limitations in Clinical Cytology. Edited by GL Wied, CM Keebler, DL Rosenthal. Chicago, Tutorials of Cytology, 1995; pp 134–141
9. Krieger PA, McGoogan E, Vooijs GP, Amma NS, Cochand-Priollet B, Colgan TJ, Davey DD, Geyer JW, Goodell RM, Grohs DH, Gupta SK, Jones BA, Koss LG, Mango LJ, McCallum SM, Nielsen M, Robinowitz M, Sauer T, Schumann JL, Syrjänen KJ, Suprun HZ, Topalides T, Wertlake PT, Whittaker J: Quality assurance/control issues: IAC task force summary. Acta Cytol 1998;42:133–140
10. Mody DR, Davey DD, Branca M, Raab SS, Schenck UG, Stanley MW, Gordon WR, Arbyn M, Beccati D, Bishop JW, Collaco LM, Cramer SF, Fitzgerald P, Heinrich J, Jhala NC, Montanari G, Kapila K, Naryshkin S, Suprun HZ: Quality assurance and risk reduction guidelines. Acta Cytol 2000;44:496–507
11. Branca M, Morosini PL, Marsan C, Coleman D: Continuous quality improvement in laboratories which undertake cervical cytology: Leonardo da Vinci-Cytotrain Project 1996-2002. PHARMIT Edizioni Scientifiche, Rome 2002; pp 1–39
12. Evans DMD, Shelley G, Cleary S, Baldwin Y: Observer variation and quality control of cytodiagnosis. J Clin Pathol 1974;27: 945–950
13. Klinkhamer PJJM, Vooijs GP, de Haan AFJ: Intraobserver and interobserver variability in the quality assessment of the cervical smears. Acta Cytol 1989;33:215–218
14. Horn Pl, Lowell DM, LiVolsi VA, Boyle CA: Reproducibility of the cytologic diagnosis of human papillomavirus infection. Acta Cytol 1985;29:692–694
15. Klinkhamer PJ, Vooijs GP, de Haan AF: Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears. Acta Cytol 1988;32:794–800
16. Yobs AR, Plott AE, Hicklin MD, Coleman SA, Johnston WW, Ashton PR, Rube IF, Watts JC, Naib ZM, Woods RJ et al: Retrospective evaluation of gynecologic cytodiagnosis: II. Interlaboratory reproducibility as shown in rescreening large consecutive samples of reported cases. Acta Cytol 1987;31:900–910
17. Confortini M, Biggeri A, Cariaggi MP, Carozzi MF, Minuti PM, Russo A, Palli D: Intralaboratory reproducibility in cervical cytology: Results of the application of a 100-set slide set. Acta Cytol 1993;37:49–54
18. Sherman ME, Schiffman MH, Lorincz AT, Manos MM, Scott DR, Kuman RJ, Kiviat NB, Stoler M, Glass AG, Rush BB: Toward objective quality assurance in cervical cytopathology: Correlation of cytopathologic diagnoses with detection of high-risk human papillomavirus types. Am J Clin Pathol 1994;102:182–187
19. Young NA, Naryshkin S, Atkinson BF, Ehya H, Gupta PK,

Kline TS, Luff RD: Inter-observer variability of cervical smears with squamous-cell abnormalities: A Philadelphia study. Diagn Cytopathol 1994;11:352–357

20. Branca M, Duca PG, Riti MG, Rossi E, Leoncni L, Turolla E, Morosini PL; the National Working Group for Extemal Quality Control in Cervical Screening: Reliability and accuracy of reporting cervical intraepithelial neoplasia (CIN) in 15 laboratories throughout Italy: Phase 1 of a national programme of external quality control in cervical screening. Cytopathology 1996;7:159–172
21. Raab SS, Snider TE, Potts SA, McDaniel HL, Robinson RA, Nelson DL, Sigman JD, Thomas PA: Atypical glandular cells of undetermined significance: Diagnostic accuracy and interobserver variability using selected cytological criteria. Am J Clin Pathol 1997;107:299–307
22. Ciatto S, Cariaggi MP, Minuti PM, Confortini M, Palli D, Pas L, Schenck U, Bonaccorsi A, Corradis G, Olivati S, Pieri L, Carretti D, Cocchi V: Interlaboratory reproducibility in reporting inadequate cervical smears: A multicentre multinational study. Cytopathology 1996;7:386–390
23. Cocchi V, Sintoni C, Carretti D, Sama D, Chiari U, Segaia V, DeLazer AL, Grilli N, Papaleo R, Ghirardini C, Bucchi L: External quality assurance in cervical-vaginal cytology: Interlaboratory agreement in the Emilia-Romagna region of ltaly. Acta Cytol 1996;40:480–488
24. Ronco G, Montanari G, Aimone V, Parisio F, Segnan N, Valle A, Volante R: Estimating the sensitivity of cervical cytology errors of interpretation and test limitation. Cytopathology 1996; 7:151–158
25. Branca M, Morosini PL, Duca PG, Verderio P, Giovagnoli MR, Riti MG, Leoncini L: Reliability and accuracy in reporting CIN in 14 laboratories: Developing new indices of diagnostic variability in an interlaboratory study. Acta Cytol 1998;42: 1370–1376
26. Stoler MH, Schiffman M: Interobserver reproducibility of cervical cytologic and histologic interpretations: Realistic estimates from the ASCUS-LSIL Triage study. JAMA 2001;285: 1500–1505
27. Fleiss JL: Statistical Models for Rates and Proportions. New York, J Wiley, 1981, pp 212–225
28. Landis JR, Koch GG: The measurement of observer agreement for categorical date. Biometrics 1977;33:159–174
29. Fleiss JL, Cohen J: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educ Psych Meas 1973;33:613–619
30. Branca M, Morosini P, Severi P, Erzen M, Di Benedetto C, Syrjänen K: New statistical software for intralaboratory and interlaboratory quality control in clinical cytology: Validation in a simulation study on clinical samples. Acta Cytol 2005;49:398–404
31. Syrjänen K, Naud P, Derchain SM, Roteli-Martins C, Longatto-Filho A, Tatti S, Branca M, Eržen M, Serpa-Hammes L, Matos J, Gontijo R, Sarian L, Lima TP, Maeda MYS, Lörincz A, Dores GB, Costa S, Syrjänen S: Comparing PAP smear cytology, aided visual inspection, screening colposcopy, cervicography and HPV testing as optional screening tools in Latin America: Study design and baseline data of the LAMS study. Anticancer Res 2005;25:3469–3480
32. Penner DW: An overview of the College of American Pathologists' program in surgical pathology and cytopathology. Acta Cytol 1989;33:439–442
33. Bonfiglio TA: Quality assurance in cytopathology: Recommendations and ongoing quality assurance activites of the American Society of Clinical Pathologists. Acta Cytol 1989;33:431–433
34. Yobs AR, Swanson RA, Lamotte LC Jr: Laboratory reliability of the Papanicolaou smear. Obstet Gynecol 1985;65:235–244
35. Thomas GDH, Head C, Thorogood J: Quality assessment in cervical cytology: A pilot study. J Clin Pathol 1988;41:215–219
36. Husain O, Butler EB, Woodford FP: Combined external quality assessment of cytology and histology opinions: A pilot scheme for a cluster of five laboratories. J Clin Pathol 1984; 37:993–1001
37. Collins DN, Kaufmann W, Clinton W: Quality evaluation of cytology laboratories in New York State: Expanded program—1971-1973. Acta Cytol 1974;18:404–413
38. Fahey MT, Irwig L, Macaskill P: Meta-analysis of Pap test accuracy. Am J Epidemiol 1995:141:680–689
39. O'Sullivan JP, Ismail SM, Barnes WS: Interobserver variation in diagnosis and grading of dyskaryosis in cervical smears: Specialist cytopathologist compared with non-specialist. J Clin Pathol 1994:47:515–518
40. Palli D, Confortini M, Biggeri A, Russo A, Cariaggi MP, Carozzi MF: A quality control system involving peer review of abnormal cervical smears. Cytopathology 1993;4:17–25
41. Rosenthal DL: Criteria and Standards for Cytology Proficiency Testing in Quality Assurance, Proficiency Testing and Workload Limitations in Clinical Cytology: Tutorial of Cytology. Edited by GL Wied, et al. Chicago, 1995, pp 238–240