External Quality Control of Cervical Cytopathology: Interlaboratory Variability

Cinara Zago Silveira Ázara Edna Joana Cláudio Manrique
Nádja Lindany Alves de Souza Andryne Rego Rodrigues
Suelene Brito do Nascimento Tavares Rita Goreti Amaral

Unit for External Quality Control, School of Pharmacy, Federal University of Goiás, Goiânia, Brazil

Key Words

Quality control · Cytology · Cervical screening

Abstract

Objective: To compare the variability of screening tests held at laboratories with the Unit for External Quality Control (UEQC), checking the frequency of cases that were discordant, false-positive, false-negative, unsatisfactory or that had a delay in clinical management and diagnostic agreement.

Materials and Methods: The study analyzed 10,053 screening tests from January 2007 to December 2008, including all positive cases, all those that fall under unsatisfactory and at least 10% of negative screening tests. The magnitude of the agreement was analyzed using the kappa coefficient. **Results:** Out of the 10,053 cases analyzed, 7.59% were considered disagreeing, and it was estimated that 1.1% were false-negative. There was a delay in the clinical procedure regarding 2.44% cases. There were 2.82% of cases identified as false-positive and 1.24% as unsatisfactory. The diagnostic agreement was excellent ($\kappa = 0.81$). The agreement of most laboratories concerning screening tests was classified as very good. The agreement of the sample adequacy was reasonable ($\kappa = 0.30$) and the agreement regarding the representation of epithelia was considered excellent. **Conclusion:** Most laboratories showed very good agreement; however, it is worthy of note that to establish the standard-

ization of diagnostic criteria, and enhance the accuracy of screening and improve the quality of cytopathology test results, it is necessary to perform external quality control.

© 2013 S. Karger AG, Basel

Introduction

The cervical cancer screening programs based on screening tests has proven its efficiency in many countries where such programs are organized and mortality rates have been reduced [1]. However, this screening test must present a good diagnostic accuracy [1, 2].

The screening tests show a high percentage of false negatives (FN), with a range of 2–62%. The main causes of the low sensitivity are related to mistakes during the collection of material, the examination of the smear or the interpretation of results [3–5]. Programs of quality control in cytopathology are an alternative that attempts to minimize these mistakes [6, 7].

Internal quality control is a necessary tool in the routine of laboratories in order to reduce mistakes in the assessment and interpretation of screening tests [8, 9]. However, the evaluation of the performance of different laboratories can only be achieved through external quality control [10].

External quality control aims at verifying the technical and diagnostic quality of smears, standardizing results, screening diagnostic difficulties, assessing accordance between source laboratories and the reviewer laboratory, working as a mechanism of diagnostic standardization as well as pointing to continuing education on cytopathology [11].

In Australia, laboratories are subject to regular inspection. There are many requirements ranging from scientific qualifications, ongoing education and limitations on the number of slides screened. Laboratories must also be enrolled in an external quality assurance program. In addition, laboratories reporting cervical cytology must meet specific performance indicators. These are a set of key, quantifiable criteria that allow comparisons of performance of individual laboratories [12].

In Brazil, the laboratories accredited by the Health System to carry out screening tests have to take part in an external quality control via the Unit for External Quality Control (UEQC), a reference laboratory designated by the state.

Overall, the errors in the examination of the smear or the interpretation of results are still a problem faced by laboratories, and it is known that their participation in external quality control programs may be a strategy for ensuring the quality of the screening tests [13, 14]. The objective of this study was to evaluate the variability of the results from the screening tests of laboratories with the results from the UEQC, checking the frequency of disagreeing, false-positive (FP), FN or unsatisfactory cases or those with a delay in clinical management and diagnostic agreement.

Materials and Methods

This is a transversal study carried out by the UEQC at the School of Pharmacy, Federal University of Goiás, Goiânia, Goiás, Brazil, and was approved by the research ethics committee of the institution under protocol No. 117/07.

The study included technical professionals responsible for 14 laboratories accredited by Health System Care, appointed by the Municipal Secretariat of Health of the city of Goiânia. The sample consisted of conventional cervical smears and results selected by the System of Information on Cervical Cancer [11], entering the screening test results of the source laboratories, including all positive, unsatisfactory and at least 10% of the negative smears performed by laboratory per month from January 2007 to December 2008, totaling 10,053 cases analyzed. The laboratories review 10% of negative smears as a method of internal quality control.

The method by which smears were reviewed by the UEQC consists of detailed manual screening by experts of all smears from the laboratories accredited by the Brazilian Health System, which are

defined in this study as source laboratories. First, the smears and results were checked, and when a nonconformity was identified (identification of smears not matching the result, broken smears, smears with no result or results without a smear), the laboratory was informed and the material was returned.

After being checked, all the smears and results classified as negative, unsatisfactory or with any alteration (atypias, low- and high-grade squamous intraepithelial lesion, carcinomas and adenocarcinomas) by the source laboratory were sent to the first reviewers.

The smears considered accordant after the first review were approved and considered as a final diagnosis. The smears considered discordant after the first review, were referred to a second review; if there was any consensus between the 2 reviews, then this was considered a final diagnosis. In the case of disagreement between the reviews, the smear was referred to a consensus meeting in which a final diagnosis was defined.

The cases considered to be disagreeing were the ones with a change in clinical procedure, according to the criteria established by the Ministry of Health/National Cancer Institute.

The results were considered FN, FP and a delay in clinical management as follows:

- FN: initial diagnosis of unsatisfactory smears, negative cytology and review classified as a diagnosis of atypical squamous cells of undetermined significance (ASC-US), ASC cannot exclude a high-grade squamous intraepithelial lesion (ASC-H), a low-grade squamous intraepithelial lesion (LSIL), a high-grade squamous intraepithelial lesion (HSIL), HSIL with features suspicious for invasion (HSIL-suspicious of invasion), squamous cell carcinoma, atypical glandular cells not otherwise specified (AGC-NOS), AGC favor neoplasia (AGC-NEO), adenocarcinoma in situ and invasive adenocarcinoma.
- FP: initial diagnosis of ASC-US, ASC-H, LSIL, HSIL, HSIL-suspicious of invasion, squamous cell carcinoma, AGC-NOS, AGC-NEO, adenocarcinoma in situ and invasive adenocarcinoma and the review classified as negative or unsatisfactory.
- A delay in clinical management: an initial diagnosis of ASC-US or LSIL and review classified as a more serious diagnosis of ASC-H, HSIL, HSIL-suspicious of invasion, squamous cell carcinoma, AGC-NOS, AGC-NEO, adenocarcinoma in situ and invasive adenocarcinoma.
- The clinical managements for cervical cytology diagnosis were established in accordance with the Ministry of Health recommendations, as specified below [15].
- Negative for intraepithelial lesion or malignancy: follow the routine cytological screening.
- ASC-US and LSIL: repeat cytology screening tests in 6 months. If changes persist or there is a worse diagnosis, send patient for colposcopy and biopsy.
- ASC-H and HSIL: perform colposcopy and biopsy.
- Squamous cell carcinoma, AGC-NOS, AGC-NEO, adenocarcinoma in situ, invasive adenocarcinoma: perform colposcopy. In case of lesions, perform biopsy. If not, perform conization.

The Bethesda System [16] was used to evaluate the adequacy of the sample and the classification of the cytology result.

The disagreeing results were sent to the source laboratory, which used the UEQC diagnosis, in the case of disagreement with the final diagnosis. When the source laboratory agreed with the final diagnosis, it had to reissue the result, mentioning that the review was confirmed by the UEQC. When the source laboratory disagreed with the final diagnosis, a consensual meeting took place

Table 1. Frequency of screening tests from the source laboratory and from the UEQC

Source laboratory		UEQC											Total	%
		ASC				Squamous lesions			AGC		Glandular lesions			
		A	B	C	D	E	F	G	H	I	J			
ASC	A	518	26	1	2	1	0	1	0	0	1	551	5.48	
	B	125	7,091	36	28	21	12	0	4	2	1	7,320	72.81	
	C	4	127	228	53	65	31	0	1	0	0	509	5.06	
	D	2	75	26	144	15	77	3	0	1	0	343	3.41	
Squamous lesions	E	1	21	101	19	560	139	0	0	2	0	843	8.38	
	F	0	2	4	22	7	327	12	1	1	1	377	3.76	
	G	0	4	0	0	0	3	25	0	0	0	32	0.32	
AGC	H	0	21	1	3	1	4	0	5	1	0	36	0.36	
	I	0	9	2	10	0	6	0	0	4	0	31	0.31	
Glandular lesions	J	0	2	0	0	1	4	1	0	1	2	11	0.11	
Total		650	7,379	399	281	671	603	42	11	12	5	10,053	100	
%		6.46	73.4	3.97	2.79	6.67	6.0	0.42	0.11	0.12	0.05	100		

Kappa = 0.81 (0.80 a 1.0); diagnostic agreement = 92.41%. Figures in bold type are the numbers of cases with the same cytology result of UEQC and source laboratory.

A = Unsatisfactory; B = negative; C = ASC-US; D = ASC-H; E = LSIL; F = HSIL; G = HSIL-suspicious of invasion/carcinoma (squamous cell carcinoma); H = AGC-NOS; I = AGC-NEO; J = adenocarcinoma in situ/adenocarcinoma (endocervical adenocarcinoma).

between the source laboratory and the UEQC, in which the final diagnosis was defined.

The data analysis used the application SAS for Windows [17]. The magnitude of agreement among tests performed by source laboratories and by the UEQC was measured by the kappa coefficient, weighted with its respective 95% confidence intervals (CIs), depending on the need to assign different weights to the disagreements. The agreement level was classified as follows: <0: worst agreement, 0–0.2: bad agreements, 0.2–0.4: reasonable agreements, 0.4–0.6: good agreements, 0.6–0.8: very good agreements and 0.8–1.0: excellent [18].

Results

Out of the 10,053 cases analyzed, the source laboratories classified 5.06% as ASC-US, 3.41% as ASC-H, 8.38% as LSIL, 3.76% as HSIL, 0.14% as HSIL-suspicious of invasion, 0.36% as AGC-NOS and 0.31% as AGC-NEO. The UEQC classified 3.97% as ASC-US, 2.79% as ASC-H, 6.67% as LSIL, 6.0% as HSIL, 0.23% as HSIL-suspicious of invasion, 0.11% as AGC-NOS and 0.12% as AGC-NEO (table 1).

Out of the 10,053 cases analyzed, 7.59% were considered disagreeing. It was estimated that 1.1% were esti-

mated to be FN, 0.37% of which were reclassified as ASC-US, 0.22% as LSIL, 0.30% as ASC-H, 0.13% as HSIL, 0.01% as HSIL-suspicious of invasion, 0.07% as AGC and 0.02% as adenocarcinoma in situ. Out of the 2.44% cases considered to have had a delay in clinical management, 0.84% were initially classified as ASC-US, 0.53% were reclassified as ASC-H, 0.31% as HSIL and 0.01% as AGC. Out of the 1.50% initially classified as LSIL, 0.19% were reclassified as ASC-H, 1.38% as HSIL and 0.02% as AGC. The study considered 2.82% of the cases to be FP and 1.24% to be unsatisfactory. The diagnostic agreement between the source laboratories and the UEQC was excellent (kappa = 0.81) (table 1).

The agreement of the sample adequacy was reasonable (kappa = 0.30). Regarding the representation of squamous, glandular and metaplastic epithelia, the agreement was excellent with kappa (0.84, 0.93 and 0.91, respectively; table 2).

It was observed that the diagnostic agreement between the laboratories and the UEQC was considered excellent for four laboratories, very good for eight and good for two (table 3).

Table 2. Comparison of sample adequacy and epithelial representation between source laboratory and from the UEQC

Source laboratory	UEQC			
	Satisfactory	Obscuring factors	Unsatisfactory	Total
Satisfactory	7,415	1,859	120	9,394
Obscuring factors	33	51	18	102
Unsatisfactory	18	21	518	557
Total	7,466	1,931	656	10,053
kappa = 0.30 (0.33–0.38)				
Epithelial representation	Squamous cells			Total
	No	Yes		
Squamous cells				
No	527	40		567
Yes	137	9,349		9,486
Total	664	9,389		10,053
kappa = 0.84 (0.82–0.87)				
	Glandular cells			Total
	No	Yes		
Glandular cells				
No	2,887	155		3,042
Yes	114	6,897		7,011
Total	3,001	7,052		10,053
kappa = 0.93 (0.93–0.94)				
	Metaplastic cells			Total
	No	Yes		
Metaplastic cells				
No	8,045	189		8,234
Yes	76	1,743		1,819
Total	8,121	1,932		10,053
kappa = 0.91 (0.90–0.92). The figures in bold type are the numbers of cases with the same cytology result of UEQC and source laboratory.				

Discussion

The results of this study showed that the agreement between results from screening tests in most source laboratories and the UEQC was very good; however, among the disagreeing results, the FP prevailed, followed by cases with a delay in clinical management.

Among cases considered to be FN, 0.59% would have to repeat the screening test in 6 months and 0.51% would

have to undergo colposcopy. In a study by Pereira et al. [14], out of the 67,954 revised screening tests, 1.67% were considered FN.

As a measure to minimize the FN results and ensure their improvement, studies have shown that the rapid review of 100% of results previously classified as negative is an efficient and cost-effective method [19, 20].

In addition, 2.44% of the cases reviewed had a delay in clinical management, i.e. all these cases should have been sent immediately for colposcopy. For such disagreeing cases, whether they were FN or had a delay in clinical management, a new result was reissued, locating the patients who initially received inadequate results and sending them for recommended clinical management and proper treatment.

Among cases considered as FP, 1.52% did not go on to repeat the screening tests in six months and 1.30% did not undergo colposcopy. This minimized the loss for both patients and the Health System, because once the lesion had been diagnosed, the patient would be sent to perform the recommended clinical management, and so unnecessary spending was avoided. A study has shown results similar to this one, in which most cases were reclassified as negative, avoiding unnecessary costs in repeating the screening tests and in the follow-up [21].

Moreover, the cytology criteria of the source laboratories and the UEQC were similar in this study with an excellent rate of agreement (92.41%).

These results were consistent with Maeda et al. [22], who found excellent levels of agreement between the source laboratory and reviewer (86.62%), and observed no FN cases, showing the applicability of external quality control in the public health system and meeting the expectations of quality required by the Ministry of Health.

It was observed that the majority of disagreeing cases were related to borderline cytology results (ASC and ASC-H). This interlaboratory and interobserver variability has been found in other studies. Confortini et al. [23] showed a low interlaboratory reproducibility in the category ASC-US (kappa = 0.34). Gatscha et al. [24] showed in their review that, out of 632 smears first diagnosed with atypical squamous, only 200 cases (32%) showed interobserver agreement, which demonstrated the low reproducibility of this diagnostic category. Smith et al. [25], Juskevicius et al. [26] and Stoler and Schiffman [27] showed that rates of atypical cells and squamous intraepithelial lesions can be influenced by the rigidity in the adoption of morphological criteria and by the degree of experience in the interpretation of cytological specimens, often without knowing the factors affecting the interobserver reproducibility.

Table 3. Evaluation of agreeing, unsatisfactory, FN and FP cases and those with a delay in clinical management

Source laboratory	Reviewed smears		Unsatisfactory		FN		FP		A delay in clinical management		kappa (CI)
	n	%	n	%	n	%	n	%	n	%	
1	564	5.61	1	0.17	2	0.35	7	1.24	40	7.09	0.82 (0.77–0.86)
2	4,041	40.19	52	1.28	44	1.09	68	1.68	109	2.70	0.84 (0.82–0.85)
3	875	8.70	12	1.37	10	1.14	50	5.71	22	2.51	0.74 (0.69–0.78)
4	79	0.78	2	2.53	2	2.54	4	5.06	4	5.06	0.65 (0.49–0.81)
5	318	3.16	13	4.08	8	2.52	8	2.52	6	1.89	0.59 (0.48–0.69)
6	476	4.73	12	2.52	3	0.63	12	2.52	1	0.21	0.77 (0.69–0.84)
7	176	1.75	7	3.97	0	0	14	7.96	15	8.53	0.57 (0.46–0.67)
8	739	7.35	8	1.08	5	0.68	30	4.06	7	0.95	0.80 (0.75–0.85)
9	295	2.93	1	0.33	5	1.70	19	6.44	6	2.03	0.80 (0.73–0.86)
10	200	1.98	2	1.0	3	1.50	7	3.50	2	1.00	0.68 (0.54–0.81)
11	273	2.71	7	2.56	4	1.47	3	1.10	3	1.10	0.85 (0.79–0.91)
12	1,201	11.94	3	0.24	9	0.75	40	3.33	14	1.17	0.87 (0.84–0.89)
13	434	4.31	5	1.15	7	1.61	5	1.15	8	1.84	0.74 (0.65–0.83)
13	382	3.79	0	0	8	2.09	16	4.19	8	2.09	0.70 (0.62–0.79)
Total	10,053	100	125	1.24	110	1.1	283	2.82	245	2.44	0.81 (0.80–0.82)

Diagnostic disagreeing = 7.59%.

Nevertheless, this study showed that concerning the adequacy of the sample, the agreement was considered reasonable ($\kappa = 0.30$). The majority of disagreeing cases is related to obscuring factors that partially affect the analysis of the smear. Our results were consistent with those of Cocchi et al. [28], who found a low interlaboratory agreement concerning the adequacy of the sample that showed κ variation between 0.01 and 0.29 and epithelial abnormalities ranging between 0.53 and 0.78.

This study found that 1.24% of smears first classified as negative were reclassified as unsatisfactory. So sample collection was repeated, according to instructions by the Bethesda System and Ministry of Health, avoiding a possible FN [16, 25].

External quality control enables the evaluation of laboratories via review of negative, positive and unsatisfactory selected smears, helping to reduce FN and FP results and to improve the accuracy of screening tests.

The results of this study showed the importance of external quality control and serve to support the implementation of continuing education for professionals to establish the standardization of diagnostic criteria, improving the quality of results of cytopathology screening in cancer of the cervix.

Acknowledgement

I would like to thank the technical professionals of the UEQC for cooperation in the screening.

Disclosure Statement

The authors have no conflicts of interest to declare.

References

- 1 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.
- 2 Mandelblatt J, Lawrence W, Womack SM, Jacobson D, Yi B, Hwang YT, Gold K, Barter J, Shah K: Benefits and costs of using HPV testing to screen for cervical cancer. *JAMA* 2002; 287:2372–2381.

- 3 Franco R, Amaral RG, Montemor EBL, Montis DM, Morais SS, Zeferino LC: Factors associated with false-negative cervical cytopathological results. *Rev Bras Ginecol Obstet* 2006;28:479–485.
- 4 Zahniser DJ, Sullivan PJ: Cytyc corporation. *Acta Cytologica* 1996;40:37–44.
- 5 Gill GW: Blinded review of Papanicolaou smears. *Cancer Cytopathol* 2005;105:53–56.
- 6 Takkanen J, Geagea A, Neiminen P, Anttila A: Quality improvement project in cervical cancer screening: practical measures for monitoring laboratory performance. *Acta Obstet Gynecol Scand* 2003;82:82–88.
- 7 Ribeiro AA, Santos SCD, Silva SRRS, Nascimento MA, Fonseca-Carvasan GA, Carneiro MAS, Rabelo-Santos M, Rabelo-Santos S: Endocervical component in conventional cervical smears: influence on detection of squamous cytologic abnormalities. *Diagn Cytopathol* 2007;35:209–212.
- 8 Sood N, Singh V: Evaluation of 100% rapid rescreening of cervical smears. *Indian J Pathol Microbiol* 2009;52:495–497.
- 9 Tavares SBN, Souza NLA, Manrique EJC, Albuquerque ZBP, Zeferino LC, Amaral RG: Comparison of the rapid prescreening, 10% random review, and clinical risk criteria as methods of internal quality control in cervical cytopathology. *Cancer Cytopathol* 2008;114:165–170.
- 10 Branca M, Morosini P, Duca P, 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.
- 11 Ministry of Health, National Cancer Institute: Manual of Quality Management for Cytopathology Laboratories. Rio de Janeiro, 2012.
- 12 Annabelle Farnsworth. Screening for the prevention of cervical cancer in the era of human papillomavirus vaccination: an Australian perspective. *Acta Cytol* 2011;55:307–312.
- 13 Salvetto M, Sandiford P: External quality assurance for cervical cytology in developing countries. Experience in Peru and Nicaragua. *Acta Cytol* 2004;48:23–31.
- 14 Pereira SMM, El Ramos D, Yamamoto LS, Shirata NK, di Loreto C, Ferraz MG, Longatto Filho L: External quality control of cervical cytopathologic and the reflex in the health public laboratory. DST-J bras Doenças Sex Transm 2006;18:172–177.
- 15 Brazilian nomenclature for cervical cytology reports and guidelines. *Rev Bras Ginecol Obstet* 2006;28:486–504.
- 16 Solomon D, Nayar R: Bethesda System for Reporting Cervical or Vaginal Cytologic Diagnoses, ed 2. Rio de Janeiro, Revinter, 2005.
- 17 SAS/STAT software changes and enhancements through release 8.2. Cary, SAS Institute, Inc., 1999–2001.
- 18 Landis JR, Koch GC: The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
- 19 Arbyn M: Detection of false negative Pap smears by rapid reviewing. *Acta Cytol* 2000;44:949–957.
- 20 Amaral RG, Zeferino LC, Hardy E, Westin MC, Martinez EZ, Montemor EB: Quality assurance in cervical smears: 100% rapid rescreening versus 10% random rescreening. *Acta Cytol* 2005;49:244–248.
- 21 Sebastião APM, Noronha L, Scheffel DLH, Garcia MJ, Carvalho NS, Collaço LM, Bleggi-Torres LF: Study of undetermined atypias in relation to prevalence and disagreement percentile in cases of the Cervical Cancer Screening Program of Paraná, Brazil. *J Bras Patol Med Lab* 2004;40:431–438.
- 22 Maeda MYS, Di Loreto C, Barreto E, Cavaliere MJ, Utagawa ML, Sakai YI, Corrêa RO, Adura PJD, Marzola VO: Siscolo-quality control system in the health public laboratories: preliminary study. *J Bras Patol Med Lab* 2004;40:425–429.
- 23 Confortini M, Carozzi F, Dalla Palma P, Ghiringhelo B, Parisio F, Prandi S, Ronco G, Ciatto S, Montanari G: Interlaboratory reproducibility of atypical squamous cells of undetermined significance report: a national survey. *Cytopathology* 2003;14:263–268.
- 24 Gatscha RM, Abadi M, Babore S, Chhieng D, Miller MJ, Saigo PE: Smears diagnosed as ASCUS: interobserver variation and follow-up. *Diagn Cytopathol* 2001;25:138–140.
- 25 Smith AE, Sherman ME, Scott DR, Tabbara SO, Dworkin L, Olson J, Thompson J, Faser C, Snell J, Schiffman M: Review of the Bethesda System Atlas does not improve reproducibility or accuracy in the classification of atypical squamous cells of undetermined significance smears. *Cancer* 2000;90:201–206.
- 26 Juskevicius R, Zou KH, Cibas ES: An analysis of factors that influence the ASCUS/SIL ratio of pathologists. *Am J Clin Pathol* 2001;16:331–335.
- 27 Stoler MH, Schiffman M: Atypical squamous cells of undetermined significance-low-grade squamous intraepithelial lesion triage study (ALTS) group: interobserver reproducibility of cervical cytology and histology interpretations; realistic estimates from the ASUS-LSIL Triage Study. *JAMA* 2001;285:1500–1505.
- 28 Cocchi V, Sintoni C, Carreti D, Sama D, Chiari U, Segala 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 Italy. *Acta Cytol* 1996;40:480–488.