



INTEROBSERVER VARIABILITY OF CERVICAL SMEARS BY RESCREENING IN SMALL LABORATORIES OF PARANÁ AND SANTA CATARINA, BRAZIL

Variabilidade interobservadores em esfregaços cervicais em pequenos laboratórios do Paraná e Santa Catarina, Brasil

Ana Paula Weinfurter Lima¹; Jaqueline Plewka¹; Jorge Nunes Basso²; Julia Araújo Torres²; Lelia Maria Philipi²; Domenic Cicchetti³; Aginaldo José do Nascimento¹; Maria Suely Soares Leonart^{1*}

¹Programa de Pós-Graduação em Ciências Farmacêuticas da Universidade Federal do Paraná. Rua Prefeito Lothário Meissner 632, Jardim Botânico, CEP 80210-170, Curitiba, PR, Brazil

²Cytologists

³Department of Biometry, Yale University School of Medicine, New Haven, CT 06512, USA

Recebido em 18/12/2008 - Aceito em 07/07/2009

* Author for correspondence: msue@ufpr.br

RESUMO

A neoplasia cervical constitui um problema de saúde pública em âmbito mundial, estimando-se que seja a terceira mais comum na população feminina. Na realização do exame de Papanicolaou, principal instrumento para a detecção de anormalidades epiteliais, deve-se buscar formas de reduzir as variações, através do aprofundamento e uniformização de critérios morfológicos e da constante revisão de lâminas. Este trabalho visou estudar a variação interobservadores na avaliação citológica de material cervical para a interpretação de lesões pré-neoplásicas e neoplásicas, com a participação de citologistas de laboratórios de pequeno porte nos Estados do Paraná e Santa Catarina. Para tanto, foram avaliados 46 esfregaços de citologia cérvico-vaginal por 4 observadores, em estudo cego, de acordo com seus procedimentos rotineiros. Os resultados foram comparados com um consenso obtido por 2 citologistas experimentados, de acordo com os critérios morfológicos preconizados pelo Sistema Bethesda 2001 para citologia cérvico-vaginal. Um número maior de lâminas foi observado para as classificações NILM, LSIL e HSIL. As percentagens globais de concordância foram NILM, 93,7%; ASC, 15,0%; LSIL, 32,5%, e HSIL, 50,0%. Em todos os pares, as correlações para os resultados dos citologistas foram estatisticamente significativas ($p < 0,05$).

Porém, houve maior concordância dos citologistas participantes entre si do que em relação ao consenso, mostrando uma tendência a subestimação das anormalidades epiteliais. Os resultados obtidos neste trabalho indicam uma boa concordância entre os citologistas, mas sugerem que há necessidade de se uniformizar os critérios morfológicos empregados em diferentes laboratórios, no sentido de reduzir a variabilidade e, desta forma, os índices de falsos resultados em citologia cérvico-vaginal.

PALAVRAS-CHAVE: variações dependentes do observador, esfregaço vaginal, controle de qualidade.

ABSTRACT: Cervical neoplasia constitutes a public health problem worldwide, considered the third most common form of cancer in the female population. In Papanicolaou test, a main instrument for the detection of epithelial abnormalities, it should be searched the reduction on variation analyses, through standardization of morphologic criteria, and constant smear revision programs. The aim of this work was to analyze interobserver variability in the classification of cervical smears for the interpretation of pre- and neoplastic lesions, using cytologists from small size Clinical Laboratories in the Brazilian states of Paraná and Santa Catarina. A 46-slide set of cervical cytology was reviewed by 4 cytologists, in a blind study, using the morphological and interpretative criteria utilized in their daily routine. The results were compared to a consensus by two experienced cytologists according to the 2001 Bethesda System for reporting cervical cytology. A larger number of smears were verified for classifications NILM, LSIL and HSIL. Global concordance percentages were NILM, 93.7%; ASC, 15.0%; LSIL, 32.5%, and HSIL, 50.0%. All the paired correlations were statistically significant ($p < 0,05$), However, higher agreement among the participant cytologists where observed than with the consensus, as a result of subestimation of the epithelial abnormalities. The results obtained in this work indicate a good agreement among cytologists, but it is also suggestive of a necessity of standardization on morphologic criteria for the different laboratories, aiming the reduction on variability and, in such a way, reduction on indices of false results in cervical cytology.

KEYWORDS: Papanicolaou smears, quality control, interobserver variation

INTRODUCTION

Cervical neoplasia constitutes a public health problem, worldwide. It represents 10% of all the malignant tumors in women, being the second most lethal neoplasia. However, it can be prevented through accessible technologies that allow the cure of 100% of the cases diagnosed when the pre- or neoplastic lesions are confined to the epithelial tissue (CEBES, 2001). In developing countries, morbidity remains high due to low quality service standards for screening and prevention of the cervical cancer (PALO et al., 2002; AGUILAR-PEREZ, et al., 2003).

Although cervical exfoliative cytology has made some headway in the detection of precursory lesions (UTAGAWA et al., 2000; PINHO and MATOS, 2002), the implementation and upgrade of quality control programs remain necessary. Errors can occur when cellular changes are missed or misinterpreted by the cytologist, even though this is less common than errors due to problems of sampling, fixation or inadequate clinical information (LEMAY and MEISELS, 1999; ROMBACH et al., 1987).

Rescreening of cervical smears can be used as a tool for laboratory control. Intra- and interobserver agreement can determine the reliability of the interpretation criteria and disclose the possibility of improving diagnostic consistency (ROMBACH et al., 1987; MITCHELL et al., 1988). The results of Interobserver variability studies can provide guidelines for decreasing observer discrepancies, thereby leading to a high degree of homogeneity in the cytological results obtained by different

professionals and laboratories (SOLOMON et al., 2002).

This study was carried out to assess interobserver variability in the classification of cervical smears for the interpretation of pre- and neoplastic lesions, using data provided by cytologists from small size clinical laboratories in the Brazilian states of Paraná and Santa Catarina.

MATERIAL AND METHODS

A total of 46 satisfactory cervical smears, randomly selected, were screened by 6 cytologists of clinical laboratories in the states of Paraná and Santa Catarina, after the patients' informed consent, and approval by the Ethical Committee for Research Involving Humans, Setor de Ciências da Saúde, Universidade Federal do Paraná. The observed classifications for the smears were, negative for intraepithelial lesion or malignancy (NILM), atypical squamous cells (ASC), low grade squamous intraepithelial lesion (LSIL), high grade squamous intraepithelial lesion (HSIL), invasive squamous cell carcinoma, atypical glandular cells (AGC), adenocarcinoma in situ (AIS), invasive adenocarcinoma and mixed alterations.

The smears were screened, in a blind study, by 4 cytologists (A,B,C, and D), with minimal of 4 years of experience, of 4 distinct laboratories, using the morphological and interpretative criteria utilized in their daily routine plus a consensual review (E) done by 2 experienced cytologists of Clinical Laboratory of Pharmaceutical Sciences Postgraduate Program of Universidade Federal do Paraná, according to the 2001 Bethesda System for

reporting cervical cytology (MITCHELL et al.,1988). The results were analyzed for interobserver variability.

The data were analyzed as non-parametric data for Spearman rank order correlations, percentage of global concordance, and for analyses of principal components. The Intraclass correlation coefficient (ICC), for more than 2 raters (BARTKO, 1966; BARTKO, 1974; PORTNEY and WATKINS, 1993) was performed with the use of the on-line statistical package of Hong Kong University, 2008, at the following website:

http://department.obg.cuhk.edu.hk/researchsupport/IntraClass_correlation.asp. Statistical significance was considered for $p < 0.05$, and all the statistical calculations were made using the Excel spreadsheet (Microsoft), and the statistical package Statistica 8.0 (Statsoft).

RESULTS AND DISCUSSION

The difficulty posed by this type of study is that an unequivocal gold standard is not

available. Some studies use consensus results obtained by a group of specialists (COCCHI et al., 1996; KATO et al., 1995) . In this work we used a review interpretation accomplished by detailed consensual analysis and discussion of an entire set of smears by two cytologists based on well-established morphological criteria.

Interobserver variability has important implications for patient care, diagnostic errors and medical litigations (GUPTA et al., 2001). In fact, the major problem is not to evaluate whether a smear is positive or negative, but to establish the correlation between lesion stage and the observed cell abnormalities.

Table I shows the distribution of the results produced by the cytologists. A larger number of smears were verified for classifications NILM, LSIL and HSIL. Cytologist C tended to underestimate negative epithelial abnormalities (NILM), while cytologist E tended to super-estimate atypical squamous cells (ASC)

Table I. Frequency of reported epithelial abnormality classification by the cytologists for 46 cervical smears

Cytologist	NILM	ASC	LSIL	HSIL	Adeno-carcinoma	Mixed Alterations
A	17	2	6	3	1	1
B	19	1	3	4	0	1
C	13	1	7	6	0	1
D	16	1	3	2	0	2
E*	19	5	10	6	1	4

E* - consensus results of 2 cytologists. Mixed alterations - squamous + glandular cell alterations. NILM - intraepithelial lesion or malignancy; ASC - atypical squamous cells; LSIL - low grade

squamous intraepithelial lesion; HSIL - high grade squamous intraepithelial lesion.

Plot of the inter-observer correlation on analyses of principal components (Figure 1) shows higher agreements for cytologists A, B, C, and D. The correlation matrix of smear interpretations for the cytologists is shown in Table II. All the correlations were significant. Lower values of correlations were obtained with consensus E.

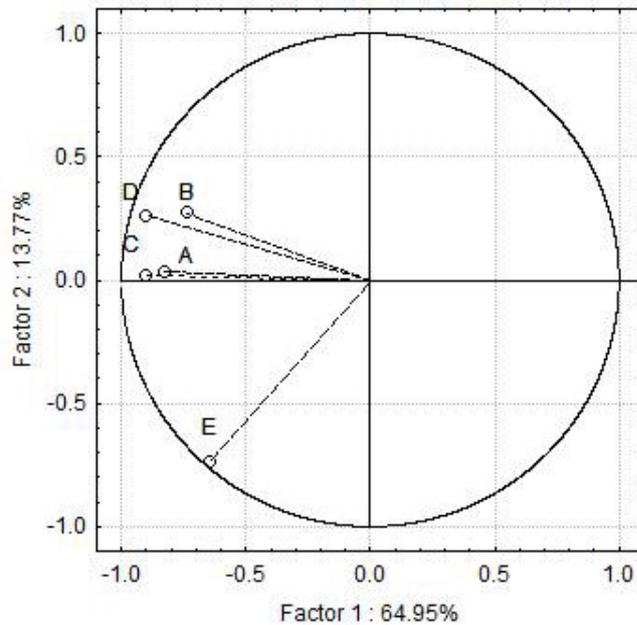


Figure 1. Plot of the correlation cycle for variable factor scores on analyses of principal components

Projection of the variables on the factor-plane (1 x 2)

Table II. Correlation matrix of evaluations of epithelial abnormality classification by cytologists in 46 cervical smears

Cytologist	A	B	C	D
A				
B	0.629			
C	0.649	0.619		
D	0.705	0.643	0.677	
E*	0.421	0.507	0.511	0.457

Spearman Rank Order Correlations. Correlations are significant at $p < 0.05000$.

E* - consensus results of 2 cytologists.

The intra-class correlation coefficients (ICCs), using the "on-line" statistical package of the Hong Kong University, 2008 were 0.6122 for cytologists A-D, and 0.4714 for cytologists A-E. According to the criteria of Cicchetti and Sparrow, 1981, ICCs can be classified for levels of clinical significance as follows: poor, < 0.40 ; fair, 0.40 to 0.59 ; good, 0.60 to 0.74 ; excellent, 0.75 to 1.00 .

10, 21, 34, 39 and 44. This resulted from a tendency of some cytologists in underestimating the interpretation in relation to the consensus. In this case, an important implication is that if the abnormal smear is falsely classified as negative, the patient can be lost to follow-up clinical care (HINDMANN, 1987).

Plots of case factor scores on analyses of principal components (Figure 2) show some major disagreements in smears numbered 4, 7,

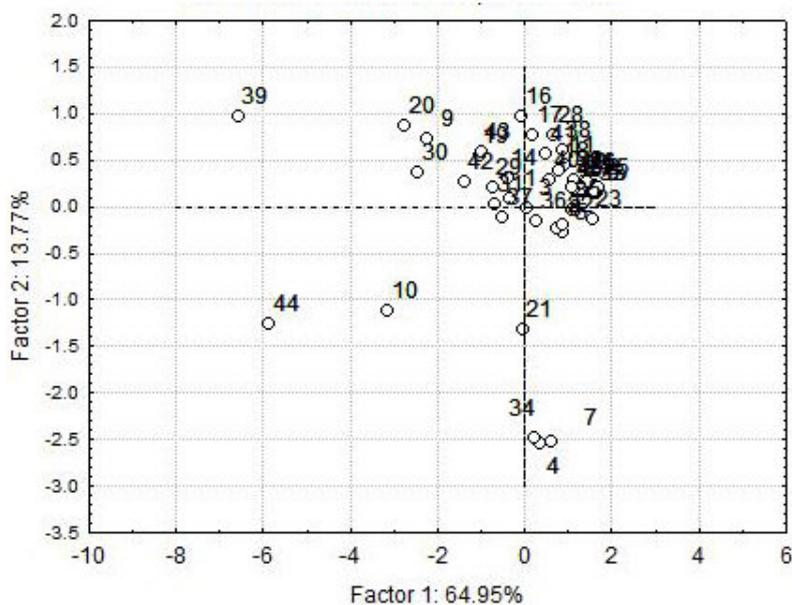


Figure 2. Plot of the projection case factor scores on analyses of principal components

Projection of the cases on the factor-plane (1 x 2)

Table III shows the percentage of global concordance with the consensus for the 4 frequent results of cytological categories. The ASC category presented the lowest percentage of global concordance. Such lower concordance score for the cytomorphologic interpretations of each cytologist in relation to consensus is in agreement with the results observed for the ASC category in the study of Cocchi et al., 1996. Their data indicated that when one of the

laboratories interpreted a smear as ASCUS, 40.3% of the paired evaluations indicated HSIL or carcinoma.

It was also observed that the highest percentage of global concordance occurred for negative evaluations. This is in agreement with data for negative classification found in the work of Gupta et al., 2001, whose medium accuracy was better for benign alterations and LSIL than for HSIL.

Table III. Percentage of global concordance for the cytological classification groups NILM, ASC, LSIL and HSIL, related to the consensus

Evaluations	Global concordance (%)
NILM	93.7
ASC	15
LSIL	32.5
HSIL	50

NILM - intraepithelial lesion or malignancy; ASC - atypical squamous cells; LSIL - low grade squamous intraepithelial lesion; HSIL - high grade squamous intraepithelial lesion. The data are average of 4 paired comparisons with the consensus.

In the LSIL group, one degree of disagreement can represent a change in group classification and can account for the lower percentages found, compared with the HSIL category, with more classification possibilities (sub-groups).

In order to estimate screening error, Graaf et al., 1987 rescreened smears originally

classified as Papanicolaou classes I or II from women, whose cytological results were consistent with moderate, severe dysplasia, carcinoma in situ or invasive cancer, three years later. The initial diagnosis proved to be underestimated in 17% of the smears. Gatscha et al., 2001 analyzed interobserver variation on rescreening of 632 ASCUS cases, a diagnostic

category that is considered ill-defined. They achieved full or partial agreement in the ASCUS category in 45% of the cases.

In places where the prevalence of uterine cervix lesions is low, rescreening of cervical smears can be very useful for quality control for important diagnostic decision - making, despite the questionable benefit of false-negative detection (CONFORTINI et al., 1993).

In a study of interobserver variation in the diagnosis and grading of dyskaryosis in cervical smears between specialist and non-specialist cytopathologists (O'SULLIVAN, et al., 1994), it were observed major differences between specialist and non-specialist cytopathologists in the diagnosis and grading of cervical smears and in the recommended management of patients with abnormal smears. Thus, they point out the importance in the exploration of possible strategies for standardizing the reporting of cervical smears, accreditation in cytopathology for non-specialist consultants, and the value of participation in external quality assessment schemes.

Lee et al., 2003 showed that telecytology can be used as an alternative method for the cytologic diagnosis of cervical smears, particularly in quality assurance programs. Diagnostic accuracy and interobserver reproducibility by pathologists and cytotechnologists using intraclass correlation coefficient (ICC), revealed good interobserver agreement for the first (0.72) and second (0.64) glass slide diagnoses and the first (0.72) and second (0.60) digital image diagnoses. The kappa values for intraobserver variation between first and second glass slide diagnoses and first and second digital image

diagnoses showed moderate to excellent agreement. They suggested that digital images are suitable substitutes for glass slides.

Souza, et al., 2004 analyzed the intra and intercytopathologists disagreement in 50 cervical slides of Pap smears, previously diagnosed as atypical skin cells with undetermined significance (ASCUS). They observed a high degree between the different analyses of the same cytopathologist, varying from 7.8 to 74.4% according to the Kappa test. The study also confirms the existence of subjectivity in the analysis of the ASCUS category and imprecision in the analysis criteria by the same examiner.

In a study of Alderisio et al., 2007, assessing interlaboratory quality control in gynecologic cytopathology and interobserver reproducibility in the Latin American screening, they showed that high interlaboratory reproducibility was obtained for sets of clear-cut cases, while more interlaboratory variation was evident in the difficult samples.

In this work all the participants, except one, include rescreening procedures in, at least, 10% of the smears in their internal quality control systems. Only two participants have access to external quality control programs and, in general, all keep close contact with the clinicians, monitoring the histological and colposcopic results of the patients.

The results obtained in this study indicate very good agreement among the cytologists. However, it is necessary to search for new forms of reducing variability by applying uniform cytological criteria and efficient quality control systems, to improve patient care regarding detection of an important pathology such as uterine cervical cancer.

REFERENCES

AGUILAR-PEREZ, J.A.; LEYVA-LOPES, A.G.; ÂNGULO-NÁJERA, D.; SALINAS, A.; LAZCANO-PONCE, E.C. Tamizaje em câncer cervical: conocimiento de la utilidad y uso de citología cervical em México. *Rev. Saúde Pública*, v.37, p.100-6, 2003.

ALDERISIO, M.; BRANCA, M.; ERZEN, M.; LONGATTO-FILHO, A.; DERCHAIN, S.; TATTI, S.; VIGHI, S.; ROTELI-MARTINS, C.; LEONCINI, L.; MAEDA, M.Y.; MONTIS, D.; GONTIJO, R.; SARIAN, L.O.; SYRJANEN, K. Interlaboratory quality control in gynecologic cytopathology using the novel CONQUISTADOR software. Interobserver reproducibility in the Latin American screening study. *Acta Cytol.* v.51, n.6, p.872-81, 2007.

BARTKO, J.J. Corrective note to: "the intraclass correlation as a measure of reliability." *Psychological Reports*, v.34, p.418, 1974.

BARTKO, J.J. The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, v.19 p.3-11, 1966.

CEBES - Centro Brasileiro de Estudos de Saúde. Curitiba: a saúde de braços abertos. Rio de Janeiro: CEBES, 2001.

CICCHETTI, D,V.; SPARROW, S,S. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, v.86, p.127-137, 1981.

COCCHI, V.; SINTONI, C.; CARRETI, D.; SAMA, D.; CHIARI, U.; SEGALA, V.; DELAZER, L.; GRILLI, N.; PAPAEO, R.; GHIRARDINI, C.; BUCCHI, L. External quality assurance in cervical/vaginal cytology: interlaboratory agreement in the Emiglia Romana region of Italy. *Acta Cytol.*, v.40, n.480-8, 1996.

CONFORTINI, M.; BIGGERI, A.; CARIAGGI, M.P.; CAROZZI, F.M.; MINUTI, P.A.; RUSSO, A.; PALLI, D. Intralaboratory reproducibility in cervical cytology: results of the application of a 100-slide set. *Acta Cytol.*, v.37, p.49-54, 1993.

GATSCHA, R.M.; ABADI, M.; BABORE, S.;, CHHIENG, D.; MILLER, M.J.; SAIGO, P.E. Smears Diagnosed as ASCUS: Interobserver Variation and Follow-Up; *Diagnostic Cytopathology*, v.25, n.2, p.138-140, 2001.

GRAAF, Y.V.D.; VOOIJS, G.P.; GAILLARD, H.L.J.; GO, D,M,D,S. Screening errors in cervical cytologic screening. *Acta Cytol.*, v.31, p.434-438, 1987.

GUPTA, D.K.; KOMAROMY-HILLER,,G.; RAAB, S.S.; NATH, M.E.. Interobserver and intraobserver variability in the cytologic diagnosis of normal and abnormal metaplastic squamous cells in Pap smears. *Acta Cytol.*, v.45, p.697-703, 2001.

HINDMAN, W.M. A proposal for quality control in gynecologic cytology. *Acta Cytol.*, v.31, p.384-385, 1987.

KATO, I.; SANTAMARIA, M.; RUIZ, P.A.; ARISTIZABAL, N.; BOSCH, F.X.; SANJOSÉ, S.; MUÑOZ, N. Interobserver variation in cytological and histological diagnoses of cervical neoplasia and its epidemiologic implication. *J. Clin. Epidemiol.*, v.48, p.1167-1174, 1995.

LEE, E.S.; KIM, I.S.; CHOI, J.S.; YEOM, B.W.; KIM, H.K.; HAN, J.H.; LEE, M.S.; LEONG, A.S. Accuracy and reproducibility of telecytology diagnosis of cervical smears. A tool for quality assurance programs. *Am. J. Clin. Pathol.* v.119. n.3, p.356-60, 2003.

LEMAY, C.; MEISELS, A. 100% rapid (partial) rescreening for quality assurance. *Acta Cytol* 1999;43:86-88.

MITCHELL, H.; MEDLEY, G.; DRAKE, M. Quality control measures for cervical cytology laboratories. *Acta Cytol.*,v.32, p.288-292, 1988.

O'SULLIVAN, J.P.; ISMAIL, S.M.; BARNES, W.S.; DEERY, A.R.; GRADWELL, E.; HARVEY J.A.; HUSAIN, O.A.; KOCJAN, G.; MCKEE, G.; OLAFSDOTTIR, R. Interobserver variation in the diagnosis and grading of dyskaryosis in cervical smears: specialist cytopathologists compared with non-specialists. *J. Clin. Pathol.* v.47, n.6, p.515-8, 1994.

PALO, G.D.; CHANEN, W.; DEXEUS, S. *Patologia e tratamento do trato genital inferior* 1 ed. Belo Horizonte: Editora Médica e Científica; 2002.

PINHO, A.A.; MATTOS, M.C.F.I. Validade da citologia cérvico-vaginal na detecção de lesões pré-neoplásicas e neoplásicas de colo de útero. *J. Bras Patol Med Lab.*, v.38, p.225-231, 2002.

PORTNEY, L.G.; WATKINS. M.P.. *Foundations of Clinical Research. Applications and Practice* Norwalk: Connecticut: Appleton & Lange, ISBN 0-8385-1065-5 p.509-516, 1993.

ROMBACH, J.J.; CRANENDONK, R.; VELTHIUS, F.J.J.M. Monitoring laboratory performance by statistical analysis of rescreening cervical smears. *Acta Cytol.*, v.31, p.887-894, 1987.

SOLOMON, D.; DAVEY, D.; KURMAN, R.; MORIARTY, A.; O'CONNOR, D.; PREY, M.; RAAB, S.; SHERMAN, M.; WILBUR, D.; WRIGHT, T. Jr.; YOUNG, N. Forum Group Members; [Bethesda 2001 Workshop](#). The 2001 Bethesda System: terminology for reporting results of cervical cytology. *JAMA*, v.287, n.16, p.2140-1, 2002.

SOUZA, J.H.K.; KALIL, I.V.; LEITE, J.M.; GEBER, S. Evaluation of cervical slides previously diagnosed as ASCUS: interassay and interobserver comparison. *Rev. Bras. Ginecol. Obstet.* v.26, n.3, p.233-240, 2004.

UTAGAWA, M.L.; PEREIRA, S.M.M.; CAVALIERE, M.J.; SHIRATA, N.K.. Lesões precursoras de câncer do colo uterino em adolescentes: impacto em saúde pública. *Folha Médica*, v.119, p.55-8, 2000.