pp. 178-183 (2013)

Interobserver variability in the interpretation of cervical smears - a must for developing an internal laboratory quality control system

Irina Bălulescu, Mihaela Badea

Micomi Clinic Bucharest (Romania)

Correspondence: Dr. Mihaela Badea e-mail: mihaela.badea@clinicamicomi.ro

Abstract

Interobserver variability means differences among different medical observers who consistently may score patients at different risk levels. Due to the intrinsic degree of subjectivity in cytological interpretation that may lead to false negative results, the interobserver variability and its implication in patient's care justify the need for implementing a routine laboratory system for quality assurance. The objective of this study was to analyse the interobserver variability in classifying cervical precancerous lesions in the specialised cyto-pathology laboratory of Bucharest Micomi Clinic, which receives about 60000 slides/year, in order to improve its quality assurance system. We conducted a retrospective study on 46934 slides evaluated as satisfactory and having been read by at least two qualified professionals, out of which we selected 774 cases (1.64%) with different interpretations. The discordant results have been split into six categories: I) positive at primary lecture, negativated by the pathologist; II) positive with a disagreement of at least one premalignancy level; III) negative at first lecture, positive at review; IV) difficult cases; V) interobserver variability by staff background; VI) third level of re-examination. The highest level of interobsever variability was at the limit between negative results with intense reactive changes and minimum atypical changes (ASC-US category). Consequently, this equivocal category has been included in the list of circumstances that are routinely submitted to a double lecture in the laboratory as a measure to reduce to a minimum level the ineluctable false negative rate of cytology. A relative lower level of discordances was noted on liquid based slides.

Keywords: *cervical cancer screening, quality control, cervical cytology, interobserver variability, atypical squamous cells*

Introduction

Papanicolaou smear test is performed worldwide in order to detect cervical cancer at its earliest stages (pre-cancer), when treatment is most effective and death can be prevented[1].

The microscopic examination and interpretation of histological and cytological specimens is a subjective procedure, highly dependent on the skills and experience of the investigator and the time spent on examination of the cell sample[2-6]. The literature describes an inter- and intra-observer variation and a high variability in percentages of correct cytological diagnoses. This variability has important implications in patient's care and in medical litigations[2,7,8].

Cervical cytology performance is limited by both false positive and false negative results. False negative report occurs when the cytologist fails to detect cancerous or precancerous cells in the smear and is harmful for the patient as it results in a failure to treat the precancerous disease. False positive report is the result of a misinterpretation of a negative smear which is reported as containing abnormal cells. This report causes unnecessary psychological distress and leads to overtreatment.

Received: *May 24, 2013*
Revised: *June 27, 2013*
Accepted: *August 19, 2013*

The quality of the test depends on subsequent steps: adequate sampling, handling and staining of the sample, screening and interpretation of the slide and reporting of the results, as well as the final step of assuring accuracy[2,3,14].

The set of measures designed to ensure the accuracy of interpretation and reporting of cervical smears is termed Quality Control. Internal quality control of cytology screening largely depends on rescreening slides initially screened as negative or inadequate. Procedures must be designed to detect potential false negatives before final results are reported in which case they have the potential to improve patient care as well as individual and laboratory accuracy. A number of approaches are available: rapid reviewing of smears initially reported as negative or inadequate; rapid preview/pre-screening of all smears; random rescreening (full rescreening of a 10% random sample of smears reported as negative or unsatisfactory - CLIA '88); targeted rescreening of specific patient groups known to be at higher risk for cytological abnormalities: history of abnormal bleeding (spotting, intermenstrual, post coital, post menopausal), recurrent vaginal infections, previous abnormal smears, abnormal cervix appearance on colposcopy, history of pre-

cancerous lesions or cervical cancer); seeding abnormal cases into the screening pools; retrospective rescreening of negative cervical cytology specimens from patients with a current high grade abnormality; automated rescreening of smears initially reported as negative[9-13].

The highest incidence of interobsever variability is in the ASC-US category, which has been problematic for pathologists, clinicians, and patients since its inclusion into the Bethesda system (1988). The problem lies in the fact that cytologists, regardless of their professional level, may not always reach distinct diagnosis, in black and white, on a cytologic smear. According to the Bethesda System, atypical squamous cells (ASC) is a term for which the definition has evolved over time. Generally, ASC is used to designate squamous epithelial cells in cervicovaginal cytologic samples that appear abnormal and have features suggestive but not a fully diagnostic of squamous cell dysplasia[1]. The proportion of ASC-US results in different cytologic laboratories is variable, depending on the type of population, and on cytotechnologists' experience. It is considered that for a low-risk population the proportion of the results ASC-US must be below 5%. It is also accepted that in laboratories serving populations with high risk (colposcopic clinics, clinics for sexually transmitted diseases) the percentage of ASC-US results may be higher, provided that the ratio ASC/SIL is not greater than 2-3:1[1,16,17,18].

Materials and methods

The retrospective study was conducted in Micomi, a multidisciplinary Clinic specialized in the cervical pathology located in Bucharest (running the largest Romanian cytology laboratory with over 60000 slides/year) on 774 cases with discrepant results selected from 46934 cytology slides recorded in the database during 9 months of activity (November 2010 - March 2011 and February 2012 - May 2012). The admission criteria were a minimum double lecture of the slides done by qualified persons with different or similar medical background and the classification of slides as satisfactory for evaluation.

The samples were received from Micomi Colposcopy department and other gynecology offices in Bucharest and around the country and were either conventional (18560, respectively 39.5%) or liquid based (18374, respectively 60.5%).

The laboratory used the 2001 Bethesda System terminology for reporting cervical cytology interpretation.

All slides were firstly read by one of the 4 cytotechnologists (biologists) who selected all the positive slides, as well as the cases with high risk for cytological abnormalities, despite their first result (previous abnormal cytology, CIN treated conservatory, history of cervical cancer, colposcopic abnormalities, high risk HPV persistent infection, abnormal bleeding - intermenstrual, post coital, post menopausal, macroscopic suspicion, history of recurrent cervical/vaginal infections). The positive and high risk cases were then submitted to a pathologist for a second lecture, as part of the routine validation process of the laboratory, before releasing the results.

The special activity of the study concerned the negative results after the first reading, who usually were double screened in a compulsory minimum level of 10%. All negative slides included in the study were submitted for a second lecture (100% review).

The discordances between different interpretations were analyzed and classified in 6 categories: I - positive results for squamous lesions after primary screening, negativated by the pathologist; II - positive results for squamous lesions at both lectures with disagreement of at least one premalignancy level; III - negative results at the first reading, positive on the second reviewing (false negative results); IV - variability in interpretation by cellular line; V - interobserver variability by medical staff background (same medical training: pathologist/pathologist and different educational basis: biologist/pathologist); V. - third level of re-examination and discussion of difficult cases (internal continuous medical education).

Results

The different interpretation of cytological slides was recorded in 1.64% of cases. These selected 774 cases with different interpretation were conventional 333 (C - 43%) and liquid based 441 (LBC - 57%). The structure of interobserver variability in this group:

Category I - 398 positive results for squamous lesions after primary screening, negativated by the pathologist, with subgroups I. a. (Table 1) and I. b. (Table 2).

Table 1 Number of ASC-US cases after primary screening, negativated by the pathologist

Diagnosis	Biologist 1	Biologist 2	Biologist 3	Biologist 4	Pathologist Validation	Total	LBC/C
ASC -US (Nov 2010)	14	6	21	-	Negative	41	22/19
ASC - US (Dec 2010)	4	1	12	-	Negative	17	9/8
ASC - US (JAN 2011)	16	15	13	-	Negative	44	21/23
ASC -US (Feb 2011)	20	6	21	-	Negative	47	35/12
ASC - US (March 2011)	16	8	17	26	Negative	67	40/27
ASC - US (Feb 2012)	4	2	10	1	Negative	17	10/7
ASC - US (March 2012)	17	4	24	9	Negative	54	30/24
ASC - US (April 2012)	10	7	18	18	Negative	53	42/11
ASC-US (May 2012)	13	5	22	10	Negative	50	24/26
TOTAL	**114**	**54**	**158**	**64**		**390 Cases**	**233/157**

Table 2 Number of cases with positive results other than ASC-US, negativated by the pathologist

Diagnosis	Biologist 1	Biologist 2	Biologist 3	Biologist 4	Pathologist Validation	Total	LBC/C
LSIL	4	1	1		Negative	6	4/2
ASC-H	1	1			Negative	2	1/1
TOTAL						**8 Cases**	**5/3**

Subgroup I. a. ASC-US results negativated

Subgroup I. b. Positive results other than ASC-US, negativated

From all cases negativated by the pathologist, the largest percentage was in the ASC-US category (390/398 cases), a small percentage of the results being interpreted LSIL (6/398 cases) and ASC-H (2/398 cases).

The smears have been analysed and discussed on the basis of the morphological criteria and clinical context with the whole team using a multi headed microscope, in the weekly meetings. After discussions, it was found that on these smears were changes that sham a true lesion (differential diagnostic): koilocyte versus pseudokoilocyte, reactive cellular changes due to fungal infection, *Trichomonas vaginalis*, drying artefacts and reparatory changes which could be interpreted as premalignant depending on skill and experience of each cytotechnologist.

Category II - interobserver variability in interpretation of positive results for squamous lesions. This category includes 100 cases in which there was a disagreement of at least one malignancy level as LSIL vs. ASC-US, ASC-US vs. ASC-H and vice-versa, ASC-US vs. LSIL (Table 3).

Category III - 84 negative results at the first reading - positive on the second reviewing (false negative results) (Table 4).

As well, most of the discordances (81/84 cases) were at the limit between a negative result and a result with minimum atypical changes (category ASC-US). There were only three discordances out of 84 cases with more than one level of malignancy (LSIL, ASC-H).

Category IV - variability in interpretation of the lesions depending on the cellular line (49 cases), with three variants: IV. a. - results with glandular changes at the first lecture (primary screening), negatives at the second one; IV. b. - results with abnormalities on both cellular lines (squamous and glandular) at first lecture, validated as squamous lesion; IV. c. - results with glandular changes at the first read, validated as squamous lesion (Table 5).

Table 3 Interpretation differences regarding the positivity level

Diagnosis	Biologist 1	Biologist 2	Biologist 3	Biologist 4	Pathologist Validation	Total	LBC/C
ASC-US	13	4	5	2	LSIL	24	13/11
	2	1	2	3	ASC-H	8	3/5
LSIL	22	4	4	7	ASCUS	37	16/21
	2		1		ASC-H	3	3/0
ASC-H		1	3		HSIL	4	2/2
	4	2	4	5	ASCUS	15	7/8
HSIL	1		2	5	ASC-H	8	4/4
				1	ASCUS	1	0/1
TOTAL						**100 Cases**	**48/52**

Table 4 Structure of false negative results

Diagnosis	Biologist 1	Biologist 2	Biologist 3	Biologist 4	Pathologist Validation	Total	LBC/C
Negative	43	7	21	10	ASCUS	81	42/39
Negative	1	1			LSIL	2	2/0
Negative	1				ASC-H	1	1/0
TOTAL						**84 Cases**	**45/39**

Table 5 Interpretation differences regarding glandular changes

Diagnosis	Biologist 1	Biologist 2	Biologist 3	Biologist 4	Pathologist Validation	Total	LBC/C
Glandular lesion	15	1	4		Negative	20	14/6
Lesion glandular + squamous	12	3		7	Squamous lesion	22	11/11
Squamous lesion		1	3		Glandular lesion	4	1/3
Glandular lesion	1			1	Squamous lesion	2	2/0
Glandular lesion AGCem	1				Glandular lesion AGC NOS	1	0/1
TOTAL						**49 Cases**	**28/21**

Our results are in concordance with the relevant literature, ascertaining the difficulty in correctly identifying glandular lesions in cytology and accurately classifying them as being squamous, endocervical, or endometrial.

In addition, most patients with cytologic diagnosis of glandular lesions that we could follow during the study (14/49) had more high grade squamous lesions on conization (10/14) than true glandular diseases.

Category V - interobserver variability analysed by medical background.

- V. a. Personnel with the same educational basis/medical training: pathologist/pathologist (Tables 6a and 6b).

Table 6a Interobserver variability by staff educational background: pathologist/pathologist

Category type	Pathologist 1	Pathologist 2	Pathologist 3 Validation	Total	LBC/C
Category I	27	18	Final Result	45	31/14
Category II	15	7		22	10/12
Category III	5	2		7	3/4
Category IV	3	2		5	3/2
TOTAL				**79 Cases**	**47/32**

Table 6b Examples of positive results variation between three pathologists

No. discordant cases	Pathologist 1	Pathologist 2	Pathologist 3
1	HSIL	ASC-H	HSIL
2	ASC-US	LSIL+K	ASC-US
1	ASC-H	LSIL+K	ASC-H
2	LSIL	LSIL	ASCUS
2	N	LSIL	ASC-US
1	ASC-H	LSIL+K	ASC-H
1	ASC-US	N	ASC-US
1	N	ASC-US postmen	N
1	HSIL	ASC-H+AGCec	ASC-H
TOTAL 12 cases - 6 LBC/6 C			

Table 7 Interobserver variability by staff educational background: *biologist/pathologist* (examples)

No. discordant cases	Biologist	Pathologist 1/2	Pathologist 3 Final Diagnosis
1	ASC-H	LSIL+K	ASC-H
10	ASC-US	Negative	ASC-US
4	ASC-US	LSIL	ASC-US
1	ASC-US	LSIL	Negative
23	Negative	ASC-US	Negative
2	ASC-US postmen	Negative	ASC-US Postmen
1	LSIL	LSIL	ASC-US
1	LSIL	Negative	ASC-US
1	Negative	ASC-H	ASC-US
1	LSIL	Negative	ASC-US
1	ASC-US	ASC-H	ASC-US
1	ASC-US	ASC-US	Negative
1	AGCec	ASC-US	Negative
1	Negative	AGCec	Negative
1	ASC-US	AGCec	Negative
1	AGCec	AGCec	ASC-US
1	ASC-H	ASC-US+AGCec	ASC-US+AGCec
TOTAL 52 cases - 29 LBC/23 C			

- V. b. with different educational basis: biologist/pathologist (Table 7).

Category VI - discussion of difficult cases – continuing medical education of the personnel. Every week, smears with discordant results between the first reader and pathologist are reviewed, individually, by all biologists and pathologists and after that are discussed on the basis of the morphological criteria and clinical context in collective meetings, using a multi headed microscope. After discussions, in most cases the first reader and the other agree with the final diagnosis given by the pathologist, but still remains a small number of cases in which the percentage of interobserver agreement varies (due to the intrinsic degree of subjectivity in cytological interpretation).

For example, within a week 40 discordant cases were selected; after reviewing the smears, in 24 cases everyone agrees with the final result given by the pathologist in the validation process, including first reader who has reconsidered the diagnosis.

Of the 16 cases remaining, 10 have been on the borderline between negative and ASC-US and 5 cases of variations for the interpretation of positive results: 2 cases LSIL vs ASC-US, 2 cases ASC-H vs ASC-US, 1 case of AGCec +ASC-US vs negative and 1 case of AGC ec negativated.

In this cases the percentage of interobserver agreement of the results has varied between 14.3% and 85.7%.

Discussion

Even if statistically the level of discordancy in slides evaluation is a very low one, the absolute figure means 774 different medical attitudes for as many patients. In our laboratory, the discordancy rate corresponds to 3-4 such situations per every working day, that is each independent reader can over-diagnose or miss about 3 cases per week. Fortunately, at the sensible inferior limit of ASC-US, the false negative rate is almost 5 times smaller than the over-diagnosis.

The category of positive results for squamous lesions after primary screening, negativated by the pathologist had a total of 390 cases with an interpretation of ASCUS and only 8 cases of LSIL and ASC-H. These results are comparable with the relevant literature, the fact that ASC-US diagnosis is one of the less reproducible cytological interpretations being well known. Unlike the rest cytological positive results, which named abnormalities at the level of individual cells, the conclusion ASC-US refers to the global appearance of the smear, amounting cellular abnormalities regarded from the quantitative point of view to their severity and in a certain clinical context (regarding relevant pathological antecedents, hormonal status and patient's age), specified by the clinician at the request of this exam.

Consequently, the equivocal category between intense reactive changes and minimum atypical changes has been included in the list of circumstances that are routinely submitted to a double lecture in the laboratory of the Micomi Clinic as part of the internal quality assurance system.

In our study, the decision between ASC-US and variants of the negative results was taken by a pathologist after reviewing these smears and considering the patient's individual clinical context.

By another perspective, the variability in interpreting these kinds of aspects is a measure of first readers' prudence. The usual step following an ASC-US result is the HPV testing, which can indicate the individual observer's level of accuracy and prudence. These correlations are reported in a separate article. The double reading of the limit ASC-US/negative with inflammation restricts the overuse of viral testing especially in young women and saves some of the inherent false negative results.

The category of interobserver variability in interpretation of positive results for squamous lesions included 100 cases in which there was a disagreement of at least one malignancy level as LSIL vs. ASC-US, ASC-US vs ASC-H and vice-versa, ASC-US vs. LSIL.

Another category was that of negative results at the first reading - positive on second review (84 cases). Most of the differences in interpretation were again at the limit between a negative result and a result with minimum atypical changes and only three cases were with a disagreement of at least one level of malignancy (LSIL, ASC-H).

The category of interobserver variability in interpretation of the lesion depending on cellular line had a total of 49 cases. Frequently, the difficulty in proper identifying glandular lesions lies not only in differentiating lesions from the reactive benign changes, but also in the classification with accuracy of the origin of cell as squamous, endocervical, or endometrial. In addition, the analysis of clinical cases showed that most patients with cytologic diagnosis of glandular lesions had in fact high grade squamous lesions, not true glandular abnormalities.

Regarding the type of cytological slides included in the selected and analysed sample of 774 slides, we found a relative better concordance in smears' interpretation in the liquid based type. Liquid based cytology (LBC) is a thin-layer or monolayer slide preparation technology that has been introduced as a potential solution to overcome the shortcomings of conventional Pap smears in cervical cancer screening. The sample is collected in a similar way to the conventional Pap smear, using a broom-type device, however rather than smearing the sample onto a microscope slide, the head of the device is rinsed or broken off into a vial of preservative fluid (immediate fixation), so that most or all cervical cells are retained. Samples are transported to the laboratory where they are mixed to disperse the cells, cellular debris, such as blood or mucus, is removed and a thin layer of cervical cells is deposited on a microscope slide, which is then stained.

Conclusion

The highest interobserver variability incidence was at the limit between a negative result with reactive changes and a result with minimum atypical changes (ASC-US category), as an expression of the inherent subjectivity in the assessment of individual cases. It is not surprising that ASC-US diagnosis has a low reproductibility and that there is a large interobserver variability regardless of the length of the training period or of the level of expertise. Our results are comparable with the relevant literature, where ASC-US diagnosis is one of the less reproducible cytological interpretations. Consequently, the equivocal category between intense reactive changes and minimum atypical changes has been included in the list of circumstances that are routinely submitted to a double lecture in the laboratory of the Micomi Clinic as part of the internal quality assurance system.

The main purpose of quality assurance activities in the cytopathological laboratory should be the maintenance, monitoring and continuous improvement of diagnostic accuracy, with the reduction to a minimum level of false negative reports rates. Double screening of slides finds more lesions than one time screening alone. So, multiple screening can be used in order to increase sensitivity. Still, for economic reasons a one time screening of one single slide remains the standard of patient care in most settings in our country.

Besides the protocol for rescreening circumstances, the internal quality assurance system of the cytology laboratory integrates in patients' risk evaluation the colposcopy and biopsies results and HPV status. These aspects are presented in a distinct communication.

The complex internal quality assurance system supposes a sustained activity of continuous training of the specialised multidisciplinary team. Gynaecologists should be informed about all measures undertaken by the cytopathology laboratory they collaborate with in order to increase the accuracy of diagnostic and patients' risk evaluation. ■

References

1. Koss L.G., Melamed M.R., Koss' diagnostic cytology and its histopathologic bases, Lippincott Williams & Wilkins, 5 Edition, USA, 2006, 1:184-240.
2. Wiener H.G., Klinkhamer P., Schenck U., Arbyn M., Bulten J., Bergeron C. & Herbert A., European guidelines for quality assurance in cervical cancerscreening: recommendations for cytology laboratories, 2007, Cytopathology, 18, 67–78.
3. Wiener H.G., Klinkhamer P., Schenck U., Bulten J., Bergeron C. & Herbert A., Laboratory guidelines andquality assurance practices for cytology, In Arbyn M., Anttila A., Jordan J., Segnan N., Ronco G., Wiener H. - European Guidelines for Quality Assurance in Cervical Cancer Screening, 2007, Ed. Office of Official Publ EU, Luxembourg.
4. Klinkhamer P.J., Vooijs G.P., De Haan A.F., Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears, 1988, Acta Cytol, 32:794–800.
5. 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., Ratcliffe N.A., Newcombe R.G., Inter- and intra-observer variation in the reporting of cervical smears: specialist cytopathologists versus histopathologists, 1996, Cytopathology, 7:78–89.
6. Koss L.G., The Papanicolaou test for cervical cancer detection. A triumph and a tragedy, 1989, JAMA, 261:737–743.
7. Stoler M.H., Toward objective quality assurance: the eyes don't have it, 2002, Am J Clin Pathol, 117:520–522.
8. Stoler M.H., Schiffman M.A., Interobserver reproducibility of cervical cytologic and histologic interpretations, 2001, JAMA, 285:1500–1505.
9. Arbyn M. & Schenck U., Detection of False Negative Pap Smears by Rapid Reviewing: A Metaanalysis, 2000, Acta Cytologica, 44: 949 – 957.
10. Arbyn M., Schenck U., Ellison E. & Hanselaar A., Metaanalysis of the accuracy of rapid prescreening relative to full screening of pap smears, 2003, Cancer, 99, 1: 9 – 16.
11. Baker A. & Metcher D.M., Rapid cervical cytology screening, 1996, Cytopathology, 2:229-301.
12. Faraker C.A. & Boxer M.E., Rapid review (partial rescreening) of cervical cytology. Four years experience and quality assurance implications, 1996, J. Clin. Pathol, 49, 7:587–591.
13. Faraker C.A. & Boxer M. E., Rapid review (partial rescreening) of cervical cytology. Four years experience and quality assurance implications, 1997, J. Clin. Pathol., 50, 1:87.
14. Gill G.W., Vigilance in Cytoscreening. Looking without seeing. Advance for Medical Laboratory Professionals, 8, 15:14-15, 21, 1996.
15. Houliston D.C., Boyd C.M., Nicholas D.S. & Smith J.A., Personal performance profiles: a useful adjunct to quality assurance in cervical cytology, 1998, Cytopathology, 9:162–170
16. Badea M. & Virtej P., Sinopsis de patologie cervicală preinvazivă, 2003, Ed. Infomedica, București, 262 p.
17. Grenko R.T., Abendroth C.S., Frauenhoffer E.E, Ruggiero F.M., & Zaino R. J., Variance in the Interpretation of Cervical Biopsy Specimens Obtained for Atypical Squamous Cells of Undetermined Significance, 2000, Am J Clin Pathol, 114:735-740.
18. McGrath C.M., ASCUS in Papanicolaou Smears Problems, Controversies and Potential Future, Directions, 2002, Am J Clin Pathol; 117(Suppl 1):S62-S75.
19. Garcea L.R, DiMaio D., The Papillomaviruses, 2007, Springer, New York, USA, 419 p.