

K statistic as a measure of quality control in cervicovaginal cytology

**D. Tamiolakis¹, P. Anastasiadis², D. Karamanidis³, A. Bounovas⁴, K. Romanidis⁵, K. Stellos⁶,
A. Kotini⁷, N. Papadopoulos⁸**

¹Department of Cytology, Regional Hospital of Alexandroupolis; ²Department of Obstetrics & Gynecology, Democritus University of Thrace;

³Department of Obstetrics & Gynecology, Regional Hospital of Alexandroupolis;

⁴Department of Experimental Surgery, Democritus University of Thrace; ⁵Department of Surgery, Democritus University of Thrace;

⁶Medical School, Democritus University of Thrace; ⁷Department of Medical Physics, Democritus University of Thrace;

⁸Department of Histology-Embryology, Democritus University of Thrace (Greece)

Summary

Quality assessment schemes are widespread in most branches of pathology but are uncommon in the more subjective areas of histopathology and cytology. Researchers in many fields have become increasingly aware of the observer as an important source of measurement error. The validity of any method of reporting evidence of an abnormal process in cellular material is based on the degree of correlation with the actual disease process as it exists in the tissue and its reproducibility. Correlations can be tested in retrospective studies in which diagnoses based on cellular evidence are matched against the disease process present in biopsy specimens. Correlations can also be tested by examination of a set of unknown cellular preparations obtained in the presence of proven disease. While reproducibility is indirectly related to correlation, it is meant to imply satisfactory utilization of the method by other groups of cytotechnologists and cytopathologists.

While cytopathology will continue to play an important role as a screening technique for the detection of cancer of the uterine cervix, its usefulness in the study of the early manifestations of the disease process is yet to be realized on a universal basis.

Key words: K statistic; Cervical cytology; Validity; Prevalence.

Introduction

The evaluation of a lab result, is related to the disease involved, and the validity of the test performed. All the above presuppose the precision of the method in use. Precision is a measure of random errors when counting a biologic magnitude. In the literature there are many synonyms describing this conception such as validity, repeatability, stability, reproducibility, etc., often confusing. Despite the necessity of estimating the precision in lab and clinical trials, this is not always the case, resulting in miscellaneous findings, not easily classified.

In this study, we estimated precision in cervicovaginal cytology reporting (Pap-tests). In order to evaluate precision in cervical cytology reporting by two different observers we estimated the K statistic. This coefficient correlates the observed agreement to the haphazard agreement.

Two different microscopists examined retrospectively 2,344 smears (1,172 between the years 1989 and 1995 and 1,172 between the years 1995 and 2001). Smears were randomly selected from the total number of smears examined during the two periods of six years, with a different prevalence each period. In the 1,172 smears of the first period (1989-1995), the number of smears with positive findings was 176 (prevalence: 15.01%), and in the same number of smears examined in the second period (1995-2001), 352 smears were positive (prevalence: 30.03%). When prevalence was 15.01%, the K statistic

was 0.56 (moderate strength of agreement); when prevalence was 30.03%, the K statistic was 0.83 (almost complete strength of agreement).

Conclusively, the K score depends on prevalence and ranges between moderate strength (0.56) when prevalence is low (15.01%) and almost perfect (0.83) when prevalence is high (30.03%).

Materials and Methods

Two thousand three hundred and forty-four Pap smears were retrospectively examined by two microscopists. Smears were selected by chance from the cohort of samples examined during a period of 12 years (1989-2001). These specific smears were divided as follows. From the period of time between the years 1989 and 1995, 1,172 smears were taken and 176 were diagnosed as positive (prevalence: 15.01%); from the period of time between the years 1995 and 2001, an equal number of smears was taken and 352 were diagnosed as positive (prevalence: 30.03%). Positive findings were those consistent with HPV infection, SIL (CIN 1,2,3) and invasive carcinoma (of squamous cell origin or adenocarcinoma). Each of the two observers thoroughly examined all smears one by one and selected a diagnosis according to fixed answers given as:

1. Negative.
2. Findings consistent with HPV infection.
3. Findings consistent with LGSIL (CIN 1).
4. Findings consistent with HGSIL (CIN 2, 3, and Ca in situ).
5. Positive (carcinoma arising from squamous epithelium).
6. Positive (carcinoma arising from glandular epithelium).

To assess quality in cervicovaginal cytology, Kappa coefficient was estimated; K value compares the level of observed agreement to the level of agreement due to chance [1],

K value is estimated by application of the formula, $K = \frac{P_o - P_c}{1 - P_c}$

where P_o is the observed proportion of agreement and P_c is the haphazard proportion of agreement.

Table 1 shows how P_o and P_c values are calculated.

K score for complete agreement is 1, for almost complete agreement 0.81-1, for substantial agreement 0.61-0.80, for moderate agreement 0.41-0.60, for fair agreement 0.21-0.40 and for chance agreement zero [2].

Results

Cytological findings from 1,172 smears (prevalence 15.01%) and 1,172 smears (prevalence 30.03%) are shown in Tables 2 and 3.

Correlating the findings from Table 2, we found that 227 smears were diagnosed as positive by both observers, 739 smears were diagnosed as negative by both observers, 101 smears were reported positive by the first and negative by the second observer and finally 105 smears were reported negative by the first and positive by the second observer (Table 4).

Correlating the findings from Table 3, we found that 299 smears were reported positive by both observers, 791 smears were reported negative by both observers, 35 smears were reported positive by the first and negative by the second observer, and finally 47 smears were reported negative by the first and positive by the second observer (Table 5).

By applying the formulas (Table 1) when prevalence was 15.01%, P_o was 0.82, P_c was 0.59 and K was 0.56 (moderate agreement); when prevalence was 30.03%, P_o was 0.93, P_c was 0.58 and K was 0.83 (almost complete agreement).

Table 1. — Calculation of P_o and P_c values.

2 nd observer	1 st observer		
Findings	Positive	Negative	Total
Positive	a	c	a+c
Negative	b	d	b+d
Total	a+b	c+d	N=a+b+c+d
$P_o = \frac{a + d}{N}$			
$P_c = \frac{\frac{(a+c)(a+b)}{N} + \frac{(b+d)(c+d)}{N}}{N}$			

Table 2. — Cytological findings in 1,172 smears (prevalence 15.01%).

	–	HPV	LGSIL	HGSIL	SCCa	AdenoCa	Total
1 st observer	844	152	70	53	53	–	1,172
2 nd observer	840	158	70	58	41	5	1,172

Table 3. — Cytological findings in 1,172 smears (prevalence 30.03%).

	–	HPV	LGSIL	HGSIL	SCCa	AdenoCa	Total
1 st observer	833	105	87	93	46	3	1,172
2 nd observer	826	111	84	96	52	3	1,172

Table 4. — Correlation of cytological findings reported from the two observers in 1,172 smears (prevalence 15.01%).

	1 st observer		
2 nd observer	Positive	Negative	Total
Positive	227	105	332
Negative	101	739	840
Total	328	844	1,172

Table 5. — Correlation of cytological findings reported by the two observers in 1,172 smears (prevalence 30.03%).

	1 st observer		
2 nd observer	Positive	Negative	Total
Positive	299	47	346
Negative	35	791	826
Total	334	838	1,172

Discussion

The value of the K statistic has been discussed by Silcocks [3], and he suggests that it is a statistical tool applicable to laboratory quality control which can provide a uniform criteria of repeatability.

In the study by Hicklin *et al.*, use is made of the Pearson correlation coefficient [4]. This is not the ideal statistical method for analysis agreement of opinion because it assumes that the report of one observer influences another. From their study, calculation of the K statistic for the two participating laboratories can be made (value 0.73 if five categories are used or 0.43 if their full 10 are used). The correlation coefficient was also used by Ooms *et al.*, in their study of bladder tumor grading [5]. They also illustrated their results pictorially, which did not show how pathologists differed among themselves as regards a consensus view. Quoting percentage agreements fails to take account of the degree of agreement which could arise by chance [6-8].

Our findings are in concordance with the literature: Horn *et al.* reported K=0.38 (fair agreement) which was raised to 0.68 when arbitrary diagnoses were not included [9]. Hussain *et al.*, also reported (in a study of 10 smears) K=0.79 (substantial agreement) but they admit that the number of cases examined was minimal to provide P_o and P_c values with gravity [10]. Finally, Thomas *et al.* in a study of 140 smears reported K=0.41-0.60 (moderate agreement) [11]. Brown and Brown suggested that all that is required from a cytological report is a comment on the presence of intraepithelial neoplasia or invasion [12]. In the recommendations of the British Society for Clinical Cytology Working Party follow-up of some grades of abnormality can be done cytologically [13]. To avoid completely overloading a colposcopy service some reliance needs to be placed on the grading of cytological abnormalities. We were unable to show complete agreement at the benign/evidence of neoplasia levels (no matter what grade) or at the evidence of intraepithelial neoplasia/invasion levels, so perhaps it is these areas that cervical cytology quality assessment schemes should be focused on in future.

Quality assessment should reflect on the cytopathologist's and technician's performance over the full range of a laboratory's practice; in an observer variation study, at the outset all observers agree that the material is adequate for the diagnosis under study. Overall organization of a cytopathology laboratory, the levels of interest and training of both cytotechnologists and cytopathologists, and the presence or absence of continuing education and quality control programs have a very real effect on the sensitivity of diagnostic cytopathology [14-16]. Important factors which may contribute to laboratory sensitivity are total volume of cases and patient population being screened. Laboratories with a relatively small volume (fewer than 10,000 cases per year) or receiving material from a low risk segment of the population may not have the opportunity to observe a sufficient spectrum of neoplastic disease on a continuing basis to maintain an adequate level of either interest or competence. The average pathologist with responsibility for cytodiagnosis often has not received adequate initial or continuing education in the field of diagnostic cytopathology. This statement is documented by recent studies from an annual cytopathology laboratory testing program in New York State [17]. On initial examination, approximately 15% of pathologists having to screen and interpret cytologic specimens without the aid of cytotechnologists did not meet minimum standards. Laboratory certification by organizations such as the International Academy of Cytology or the American Society of Cytology aid in improving the overall sensitivity of cytopathology laboratories on an individual basis.

From the standpoint of reproducibility, the method of reporting cellular evidence of disease has now been satisfactorily employed by numerous cytotechnologists and pathologists trained in several different laboratories.

There are those who are somewhat critical of the data presented by certain laboratories with respect to the high level of specificity of diagnosis reported. For example, Koss [18] has stated: "We have to differentiate between what I call the closed system and open system. The closed system is that in which a single laboratory compares its cytologic results with its own histologic interpretation. There the accuracy may be very high. In what I call the open system, in which different observers in different laboratories compare their results, the accuracy is not nearly satisfactory". In response to this type of comment, the results of our study would appear appropriate. It is towards this goal upon which the philosophical foundations of diagnostic cytopathology are built.

References

- [1] Cohen J.: "A coefficient of agreement for nominal scale". *Educational and Psychological Measurement*, 1960, 20, 37.
- [2] Landis J. R., Koch G. G.: "The measurement of observed agreement for categorical data". *Biometrics*, 1977, 33, 159.
- [3] Silcocks P. B. S.: "Measuring repeatability and validity of histological diagnosis-a brief review with some practical examples". *J. Clin. Pathol.*, 1983, 36, 1269.
- [4] Hicklin M. D., Watts J. C., Plott H. E. *et al.*: "Retrospective evaluation of gynaecological diagnosis. 1. Reproducibility using an experimental diagnostic scale". *Acta Cytologica*, 1984, 28, 58.
- [5] Ooms E. C. M., Anderson W. A. D., Alons C. L., Boon M. E., Veldhuizen R. W.: "Analysis of the performance of pathologists in the grading of bladder tumors". *Hum. Pathol.*, 1983, 14, 140.
- [6] Husain O. A. N., Butler E. B., Evans D. M. D., MacGregor J. E., Yule R.: "Quality control in cervical cytology". *J. Clin. Pathol.*, 1974, 27, 935.
- [7] Langley F. A.: "Quality control in histopathology and diagnostic cytology". *Histopathology*, 1978, 2, 3.
- [8] Evans D. M. D., Shelly G., Clearly B., Baldwin Y.: "Observer variation and quality control of cytodiagnosis". *J. Clin. Pathol.*, 1974, 27, 945.
- [9] Horn P. L., Lowell D. M., Livolsi V. A., Boyle C. A.: "Reproducibility of the cytologic diagnosis of HPV infection". *Acta Cytol.*, 1984, 962.
- [10] Husain O. A. N., Butler E. B., Woodford F. P.: "Combined external quality assessment of cytology and histology opinions: a pilot scheme for a cluster of five laboratories". *J. Clin. Pathol.*, 1984, 37, 993.
- [11] Thomas G. D. H., Head C., Thorogood J.: "Quality assessment in cervical cytology: a pilot study". *J. Clin. Pathol.*, 1988, 41, 215.
- [12] Brown L. J. R., Brown L.: "Cervical screening". *Lancet*, 1986, 2, 925.
- [13] Evans D. M. D., Hudson E. A., Brown C. L., Boddington M. M., Hughes H. E., Mackenzie E. F. D.: "Management of women with abnormal cervical smears: supplement to terminology in gynaecological cytopathology". *J. Clin. Pathol.*, 1978, 40, 530.
- [14] Wied G. L.: "Quality control mechanism for cytology program". *Acta Cytol.*, 1965, 9, 407.
- [15] Woodworth F. E.: "Organization of the laboratory". In: Keebler and Reagan: "A Manual of Cytotechnology". (American Society of Clinical Pathologists), Chicago, 1975, 251.
- [16] Woodworth F. E.: "Organization of the laboratory". In: Wied, Koss and Reagan. "Compendium on Diagnostic Cytology". (Tutorials of Cytology), Chicago, 1976, 594.
- [17] Collins D. N., Kaufmann W., Clinton W.: "Quality evaluation of cytology laboratories in New York State: expanded program, 1971-1973". *Acta Cytol.*, 1974, 18, 404.
- [18] Koss L. G.: "Precancerous changes of the epithelia of the uterine cervix". In: Wied, Koss and Reagan. "Compendium on Diagnostic Cytology". (Tutorials of Cytology), Chicago, 1976, 99.

Address reprint requests to:

N. PAPADOPOULOS

Asso. Professor in Histology-Embryology
Democritus University of Thrace
Dragana, 68100 Alexandroupolis (Greece)