Cytopathology 1993, **4**, 131–137

The evaluation and interpretation of cervical cytology: application of the likelihood ratio concept

R. W. M. GIARD AND J. HERMANS*
*Department of Clinical Pathology, St Clara Hospital, Rotterdam, and Working Party for Clinical Decision Making and *Department of Medical Statistics, State University of Leiden, Leiden, The Netherlands*

Accepted for publication 10 February 1993

GIARD R. W. M. AND HERMANS J. (1993) *Cytopathology* **4**, 131–137

The evaluation and interpretation of cervical cytology: application of the likelihood ratio concept

The Papanicolaou smear (Pap test), used for the detection and prevention of neoplastic lesions of the cervix, is known to have both false negative and false positive results. Proper handling of the diagnostic uncertainty resulting from these errors demands quantification of flaws. Traditionally, sensitivity, specificity and predictive values are used for that aim. In this study another approach is advocated, namely the use of the likelihood ratio. For cervical cytology this ratio is the quotient of the probability of a Pap class within the diseased population to the probability of that same Pap class within the non-diseased group. This approach enables the characterization of each Pap class separately, and is therefore much better for clinical interpretation of the result. It is also a superior approach for quality assessment.

Keywords: cervical cytology, test evaluation, quality control, likelihood ratio, Pap test

Le frottis de Papanicolaou (Pap-Test), utilisé pour la détection et la prévention des lésions néoplasiques du col utérin peut fournir des résultats faussement négatifs et faussement positifs. Une estimation quantitative de ces défauts est nécessaire pour une prise en compte correcte de l'incertitude diagnostique résultant de ces erreurs. La sensibilité, la spécificité et les valeurs prédictives sont traditionnellement utilisées dans ce but. Dans cette étude, une autre approche a été préconisée utilisant le rapport de vraisemblance. En cytologie cervico-vaginale et pour une classe de Papanicolaou donnée, ce rapport est le quotient de la probabilité de cette classe dans la population malade, à la probabilité de cette même classe dans la population non malade. Cette approche permet de caractériser chaque classe de Papanicolaou et de ce fait elle est meilleure pour l'interprétation clinique du résultat cytologique. Elle constitue également une meilleure approche pour l'assurance de qualité.

Unter den gynäkologischen Früherkennungsabstrichen gibt es sowohl falsch negative als auch falsch positive Ergebnisse. Zur Quantifizierung derartiger Störungen dienen üblicherweise di Sensitivität, die Spezifität und die prädiktiven Werte. In der vorliegenden Studie wird ein anderes Verfahren vorgeschlagen, nämlich das Wahrscheinlichkeitsverhältnis. Für die Cervix-cytologie besteht diese aus dem

Correspondence: R. W. M. Giard, Afdeling Klinische Pathologie, St. Clara Ziekenhuis, Olympiaweg 350, 3078 HT Rotterdam, The Netherlands.

Quotienten der Wahrscheinlichkeit einer PAP-Klasse innerhalb einer erkrankten Population zu der einer gesunden Gruppe. Diese erlaubt jede Pap-Klasse einzeln zu charakterisieren, was Vorteile für die klinische Interpretation und die Qualitätskontrolle bietet.

INTRODUCTION

The Papanicolaou smear (Pap test) was welcomed as an auspicious tool in the detection and prevention of neoplastic lesions of the cervix[1]. It was assumed that with its employment, cervical cancer could be discovered and treated in its preneoplastic or early stages. However, though this test has now been applied for many decades, it has not yet fulfilled this promise, and therefore has increasingly come under public and professional scrutiny. False negative test results proved to be an important cause of failed population screening programmes[2] and women suffering from cervical cancer undetected by antecedent smears have instituted legal proceedings against the cytopathologist. Not only has the accuracy of the Pap test been challenged, but also the adequacy of the Papanicolaou classification scheme itself. Recently, a new reporting system was proposed, which was claimed to be more practical and unambiguous[3].

The occurrence in the Pap test of false negative and false positive results is by no means exceptional. No faultless tests exist, so it is a property shared with all other diagnostic tests. To deal adequately with diagnostic uncertainty it is appropriate to quantify the probability of good and false test outcomes retrospectively. Such a test evaluation leads to probabilistic measures, which enable the attending clinician to make a proper interpretation of test results in the individual patient using Bayes' theorem[4].

Traditionally, sensitivity, specificity and predictive values are used for the evaluation of tests. Recently, another approach was suggested for statistical analysis using the likelihood ratio (LR) concept[5]. The LR expresses the likelihood of a test outcome category in patients with disease divided by the likelihood of that outcome category in individuals without disease. For the Pap test, the LR of a defined Pap class is simply the quotient of the proportion of that class within the diseased population to the proportion of that class within the non-diseased group.

In this paper we shall apply this concept for the evaluation and interpretation of cervical cytology and discuss these findings in the light of current criticism of the diagnostic procedure.

STUDY POPULATION AND METHODS

All consecutive smears taken during the year 1988 from women referred by general practitioners to the Department of Obstetrics and Gynaecology of the St Clara Hospital, Rotterdam, are included in this study. The Pap test was consequently applied to a symptomatic population. After sampling with the wooden Ayre spatula, the smears were immediately fixed, stained using the modified Papanicolaou method and screened by an experienced cytotechnologist.

For diagnosis a modified Papanicolaou classification index was used with the following diagnostic descriptions: Pap 1, normal; Pap 2, atypia but benign; Pap 3a, mild or moderate dysplasia; Pap 3b, severe dysplasia; Pap 4, carcinoma *in situ*; Pap 5, invasive cervical cancer.

The cytological diagnoses were systematically encoded by two cytotechnologists and entered in the national registry, especially devised for cervical cytology[6]. Results from the

Table 1. Relation between outcome of the Pap test and final diagnosis based on follow-up data (cytological, histological and/or clinical)

Pap class	No FU	Normal	CIN I	CIN II	CIN III	Sq. carc.	Ad. carc.	Ad. endom.	Oth. mal.	Total
1	0	2555	0	3	2	2	0	3	0	2565
2	0	837	0	8	5	1	0	2	0	853
3a	15	179	39	34	31	1	2	1	0	302
3b	3	12	5	8	27	1	1	4	0	61
4	0	2	0	0	9	5	0	2	2	20
5	0	0	0	0	0	2	0	0	3	5
Total	18	3585	44	53	74	12	3	12	5	3806

No FU, No follow up; CIN, cervical intraepithelial neoplasia; Sq. carc., squamous carcinoma; Ad. carc., adenocarcinoma of the cervix; Ad. endom., adenocarcinoma of the endometrium; Oth. mal., other non-cervical malignancies.

Table 2. Relation between the conclusive diagnosis (dichotomized into two categories) and the outcome of the Pap test

	Pap 1	Pap 2	Pap 3a	Pap 3b	Pap 4	Pap 5	Total
CIN II or worse	7 (0.049)	14 (0.098)	68 (0.478)	37 (0.260)	14 (0.098)	2 (0.014)	142 (1.0)
CIN I or less	2555 (0.704)	837 (0.230)	218 (0.060)	17 (0.0047)	2 (0.00055)	0 (0.000)	3629 (1.0)
Total	2562	851	286	54	16	2	3771

The probability for each Pap class for the given conclusive diagnosis is given in parentheses.

Pap test were then related to the conclusive diagnosis, which was established using histological, cytological or clinical follow-up data. This information was available in the above-mentioned registry. Based on the different clinical significance, the final diagnosis was dichotomized into a group of patients suffering from at least moderate dysplasia (CIN II), and a group with less than moderate dysplasia. The minimal follow-up period for all smears was 2 years.

The results of comparison between outcome and conclusive diagnosis were then summarized in a 6×2 table, from which the probability of a particular Pap class was derived from the proportion of smears given the presence or absence of at least moderate dysplasia (CIN II) of the cervix. The LR for at least significant dysplasia of the cervix for a particular Pap class was calculated as the probability of that Pap class given at least moderate dysplasia divided by the probability of that Pap class in absence of that lesion. For the 95% confidence limits of the LRs, the same methods as for risk ratios were used[7]. From this table, the values for sensitivity

Table 3. Likelihood ratios (LR) derived for each Pap class with 95% confidence limits (95% CL) and their discriminatory power (DP)

Pap class	LR	95% CL	DP
1	0.07	0.03–0.14	1.15
2	0.43	0.26–0.71	0.36
3a	7.97	6.43–9.88	0.90
3b	55.62	32.12–96.34	1.74
4	178.89	41.05–779.69	2.25
5	Infinity	—	—

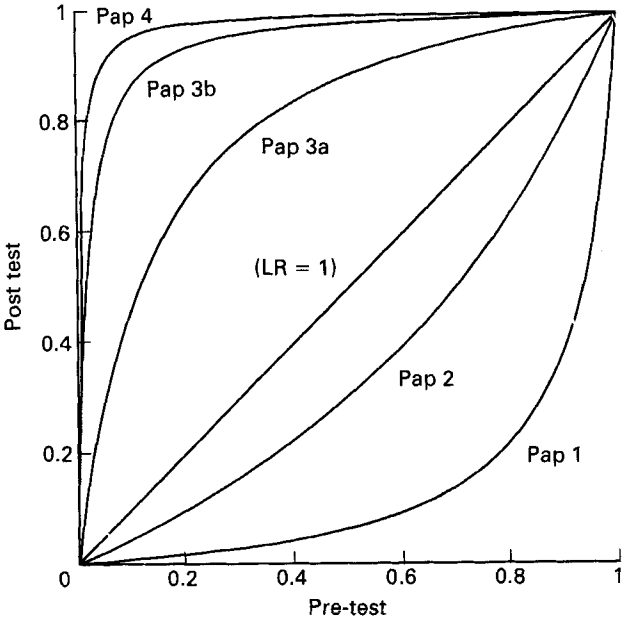

Figure 1. The relation between the prior and post test probability for each Pap class of the presence of at least moderate dysplasia. (Since the likelihood ratio (LR) for Pap class 5 is infinite, this means that the post test probability is always 1.)

and specificity were also derived. Sensitivity was defined as the sum of the proportions of Pap classes 3a to 5 given CIN II or more, specificity as the sum of the proportions of Pap class 1 and 2 given CIN I or less. For these parameters, the 95% confidence limits were calculated for binomial proportions.

The post test probability of the presence of at least moderate dysplasia for a patient with a particular Pap class, P(CIN II | Pap class), depends on both the LR of that class, LR(Pap class), and the prior probability of that grade of dysplasia.

Using Bayes' theorem it is calculated using the following formula[8]:

$$P(\geqslant \text{CINII} \mid \text{Pap class}) = \frac{\text{prior} \times \text{LR(Pap class)}}{(1 - \text{prior}) + \text{prior} \times \text{LR(Pap class)}}$$

Finally, for each Pap class the discrimination power (DP) was expressed as the absolute value of the 10-logarithm of the likelihood ratio:

$$DP = |^{10}\log(LR)|$$

thus permitting the comparison of the different classes. An indifferent class (LR = 1) has a DP = 0. A class with a likelihood ratio of 0.0001 will have the same discriminatory power of DP = 4 as a class with a likelihood ratio of 10 000.

RESULTS

In this retrospective study all 3873 Pap tests taken during 1988 were reviewed. Of this total, 67 smears (1.7%) were inadequate and accordingly not included. The relationship between the outcome of the 3806 technically adequate smears and the conclusive diagnosis is summarized in Table 1.

For statistical evaluation of the Pap test employed in this group, 17 cases with non-cervical malignancies were discarded, as were the 18 cases where adequate follow-up data were lacking. The data from the remaining 3771 cervical smears were transformed to the format of Table 2, where the final outcome is presented in two categories because of its different clinical consequences. The proportion of each Pap class given the final outcome of presence or absence of significant cervical lesion is derived from the numbers in this table. Using the quotients of these proportions for each Pap class, the LRs were calculated, and from these the discriminatory power (Table 3). Considering Pap 3a or higher as 'positive' test outcome, the sensitivity was 121/142 = 0.85 (95% confidence interval: 0.79–0.91) and the specificity 3392/3629 = 0.93 (95% confidence interval 0.92–0.94).

Given LRs derived for each Pap class and using Bayes' theorem, the relationship between the pre- and post test probability of clinically significant cervical lesions was calculated and depicted in Figure 1.

DISCUSSION

The traditional method for the evaluation of a diagnostic test is the determination of sensitivity and specificity. To calculate these test characteristics for a test with multiple outcome categories demands the combination of several outcomes to, finally, two categories. With cervical cytology Pap 1 and 2 are considered 'normal' and Pap 3a, 3b, 4 and 5 'abnormal'. The use of LRs for the evaluation of cervical cytology is much more realistic. Instead of combining categories, every outcome is now assessed individually. With the use of Bayes' theorem, this particular test result is now interpretable and the clinician can decide what to do. What does this method show for the different Pap classes?

In our data, based on the LR, there is a clear-cut difference between the diagnostic properties of the two 'normal' results Pap 1 and Pap 2 (see Figure 1). This important information is lost by disregarding their difference by combining them into one outcome category, 'normal'. For the 'abnormal' results, the gain of information (the difference between pre- and post test probability of moderate or severe dysplasia) for the class Pap 3a is less than with Pap 4 (see Figure 1). This indicates that every category has a different information content, hence the combination of outcome categories is to be avoided.

Good discrimination of a Pap class might be defined as having a discriminatory power of at least 1 (implying an LR of $\geqq 10$ or $\leqq 0.1$). When the six different results are compared on the

basis of their DP, Pap 2 and Pap 3a do not fulfil this requirement, and have meagre discrimination abilities. The clinical significance of these observations might be a matter of further discussion; in any case, combining different categories should be avoided.

Recently the problem of negative cytology in the face of cervical cancer or severe dysplasia has gained much attention[9–11]. The false negative rate with histologically proven cancer may be as high as 63%. The three main reasons for a false negative test are sampling error, screening detection error, and rapid development of cancer. The problem of overdiagnosis leading to unnecessary treatment has in addition been recognized. The effectiveness of the Pap test has therefore been questioned. Every effort should certainly be taken to reduce error, but this will not lead to a perfect Pap test. As already stated, no perfect test exists, and as a consequence of this maxim, every test outcome must be examined in the light of other clinical information. In other words, the diagnostic importance of a negative finding must be carefully analysed[12].

Diagnostic tests are used for various applications, mainly establishment or exclusion of disease. For the exclusion of a clinically significant dysplastic cervical lesion, it is shown in Figure 1 that Pap 2 does not achieve that aim. A woman belonging to a high risk group presenting with recurrent contact bleeding, estimated to have a prior probability of cervical cancer of 0.35, still has a chance of 0.19 of harbouring moderate dysplasia or more when the result of the smear is Pap 2. Conversely, an asymptomatic woman with a low risk profile (say a pre-test probability of less than 0.10), however, can be almost certain of her health with the result being Pap class 1.

If the test is to be criticised for missing the lesion, it should be remembered that we are dealing with a screening test, which by definition is intended to discriminate between women possibly harbouring disease and women probably being disease-free; it is not an exclusion test[13]. After a positive result, the true disease state must be established with a confirmatory test. This stresses the need to define clearly the purposes of the test and evaluate it from this perspective. For screening reasons, the test has proved to be a valuable tool, but quality control is cardinal for its success.

From the perspective of quality control, the likelihood ratio concept is a powerful statistical tool in the evaluation, and above all the comparison, of different laboratories, as was demonstrated in a recent study of aspiration cytology of the breast[14]. Not only is it suitable for the comparison of different laboratories using the same classification scheme, but also for comparing the discriminatory power of a new and an old system of classification on the same population.

For the introduction of the Bethesda system, several theoretical arguments have been put forward, which have been questioned[15]. Is this new classification scheme really better? Do the newly defined outcome categories have a better discriminatory power? The likelihood ratio approach may give an answer by showing this.

Our analysis shows that the likelihood ratio concept is a better way of interpreting a cytology result than evaluating the sensitivity or specificity or predictive value of the test, and is a good measure of the quality of the report. It therefore deserves the attention of every cytopathologist.

REFERENCES

1 Koss L. The Papanicolaou test for cervical cancer detection. A triumph and a tragedy. *JAMA* 1989; **261:** 737–43.

2 Chamberlain J. Failures of the cervical cytology screening program. *Br Med J* 1984; **289:** 853–4.

3 National Cancer Institute Working Group. The 1988 Bethesda system for reporting cervical/vaginal cytological diagnoses. *JAMA* 1989; **262:** 931–4.

4 Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making*. Boston: Butterworths 1988: 67–100.

5 Giard RWM, Hermans J. The interpretation of diagnostic cytology with likelihood ratios. *Arch Pathol Lab Med* 1990; **114:** 852–4.

6 Voojs GP, Casparie-van Velsen I, Peters F, Beck HL. National Registry of cervical cytologic diagnoses in the Netherlands. *Acta Cytol* 1989; **33:** 825–30.

7 Morris JA, Gardner MJ. Calculating confidence intervals for relative risks, odds ratios, and standardized ratios and rates. In: Gardner MJ, Altman DG, eds. *Statistics with Confidence. Confidence Intervals and Statistical Guide-lines*. London: BMA Press 1989: 50–63.

8 Albert A. *Multivariate Interpretation of Clinical Laboratory Data*. New York: Marcel Dekker 1987: 75–97.

9 Morell ND, Taylor JR, Snyder RN, Ziel HK, Saltz A, Willie S. False-negative cytology rates in patients in whom invasive cervical cytology subsequently developed. *Obstet Gynecol* 1982; **60:** 41–45.

10 Attwood ME, Woodman CBJ, Luesly D, Jordan JA. Previous cytology in patients with invasive carcinoma of the cervix. *Acta Cytol* 1985; **29:** 108–10.

11 Mitchell H, Medley G, Giles G. Cervical cancers diagnosed after negative results on cervical cytology: perspective in the 1980s. *BMJ* 1990; **300:** 1622–6.

12 Gorry GA, Pauker SG, Schwartz WB. The diagnostic importance of a normal finding. *N Eng J Med* 1978; **298:** 486–9.

13 Fletcher RH, Fletcher SW, Wagner EH. *Clinical Epidemiology, the Essentials*. Baltimore: William & Wilkings 1988: 159–69.

14 Giard RWM, Hermans J. The value of aspiration cytology of the breast: a statistical review of the medical literature. *Cancer* 1992; **69:** 2104–10.

15 Voojs GP. Does the Bethesda system promote or endanger the quality of cervical cytology? *Acta Cytol* 1990; **34:** 455–6.