Comparison of the Cervical Cytology Test Using the PAPNET Method and Conventional Microscopy

Dalia Weissbrod,[1] Mario Torres,[1] Alfredo Rodríguez,[1]
Irene Ureña,[1] Jesús Estrada,[1]
María Elena Reyes,[1] & Alberto J. Carreto[1]

◆ ◆ ◆

From August 1994 to June 1995, laboratories in 28 Mexican states and the Federal District submitted a total of 10 098 diagnosed Pap test slides to Mexico's National Institute of Epidemiologic Diagnosis and Reference (INDRE) for reexamination by conventional methods and also by the automated PAPNET system in Suffern, New York, U.S.A. The aim was to determine the degree of agreement obtained by these various methods. Most of the slides examined (at least 78%) yielded negative results or merely indicated an inflammatory process; 8% to 14% indicated mild or moderate cervical dysplasia; and 2% to 3% indicated conditions ranging from severe dysplasia to invasive cervical cancer. Comparison of the state laboratory and INDRE diagnoses yielded a Kappa correlation coefficient of 0.62, near the lower limit of agreement, the agreement being poorest in cases where it was necessary to distinguish between degrees of abnormality. Although state laboratory underestimation appeared lowest with respect to cases of atypia and of mild, moderate, and severe dysplasia (between 12% and 20%), these percentages are alarming because it is at these stages that the patient may be treated to prevent evolution to carcinoma. While the Kappa correlation coefficient was better (0.80) when the INDRE and PAPNET diagnoses were compared, PAPNET showed only limited ability to distinguish between various pathologic alterations, and the percentages of underestimates (false negatives) obtained with PAPNET were also high. Overall, the results indicate a need to improve the quality of cervical cytology diagnoses at state public health laboratories in Mexico through stepped-up training and supervision. They also indicate that the use of PAPNET involves greater difficulty than does manual microscopic examination of cervical smears, and that a way still needs to be found to detect and review the false negative results generated by PAPNET before approving use of this technology.

N o sooner had the cervical cytology test devised by George Papanicolaou in the 1950s gained popularity than it became apparent that an automated method for reading test results would be extremely useful. It is not surprising that since then a number of devices capable of performing this type of cytologic screening have been developed. Though many of the initial models have been discontinued, at least six U.S. and Canadian companies are continuing their efforts to perfect the technique.

The first automated system, which was created in the United States in 1956 and given the name "Cytoanalyzer" (1), was designed to detect malignant or premalignant cells by assessing the size and density of the cell nucleus. Unfortunately, the results recorded in the preclinical experiments were not as good as had been anticipated, due to difficulty distinguishing between artificial changes resulting from manipulation of the sample and actual

---
[1] National Institute of Epidemiologic Diagnosis and Reference (*Instituto Nacional de Diagnóstico y Referencia Epidemiológicos*—INDRE), Mexico City, Mexico. Mailing address: Carpio No. 470, Col. Santo Tomás, Del. Miguel Hidalgo, C.P. 11340, Mexico City, Mexico.

malignant cells. The Vickers system, in which the cells were transferred to a transparent wheel-mounted tape that conveyed them through various stains and ultimately through a scanner, constituted the first attempt to improve specimen quality. However, the system was plagued by numerous problems and was eventually discarded.

Research continued, however, and other systems were developed. In the United States, where such initiatives were undertaken with considerable enthusiasm, attention was focused on flow cytometers, which proved unreliable. By the end of the 1970s, enthusiasm had waned; but in Europe and Japan a number of projects aimed at developing computerized methods for conducting image analysis were launched.

It was eventually concluded that such methods faced major problems, since in order to be good a technique needed the ability to distinguish between normal cells, artificial alternations produced by sample manipulation, premalignant cells, and malignant cells. In addition, the International Academy of Cytology established the requirement that any automated system should produce no false negative results (2).

By the mid-1980s precision had improved considerably; but the use of automated devices did not spread through Europe or Japan, since manufacturers feared that the devices would not be accepted by cytopathologists and that marketing problems would result. In the United States, however, the situation was different; a large number of cytologic examinations (some 70 million a year) were carried out using an organized and standardized procedure that facilitated entry of new devices.

In 1987 the controversy surrounding diagnostic quality led the *Wall Street Journal* to publish an extensive article (3) questioning the validity of traditional Pap tests. This in turn led to an increase in the demand for cytotechnologists, as the poor quality of interpretation was attributed to the excessive number of slides that had to be exam-ined daily by each cytotechnologist. At the same time, authorities involved began to reflect more realistically on the quality of diagnoses made using the automated cytologic test, which had only to be similar to the quality of diagnoses obtained using the traditional method (4).

In the field of cytology, a number of morphologic criteria are used to differentiate normal cells from abnormal cells, whose abnormality spectrum is very broad. Application of such criteria is extremely subjective, as a result of which diagnosis is contingent on a large variety of factors— including the capability of the personnel observing and interpreting the cellular changes. It is therefore inevitable that false negative results are produced in the cytologic examination of cervical samples.

Many factors come into play in making a diagnosis following examination of a cervical smear: the preparation of the patient, the technique used to obtain the sample and spread it on the slide, the fixing and staining of the specimen, the thoroughness of the examination performed by the cytotechnologist, and the accuracy of the interpretation made by the cytopathologist. Medical literature has reported that the frequency of false negative results is about 50% in all diagnostic categories of malignant cervical lesions, as a result of errors in sample taking. It is important to distinguish this kind of false negative result from results that are "truly" false negative, these latter being attributable to an interpretation error margin of between 15% and 55% in the case of invasive cancer and between 6% and 45% in the case of *in situ* carcinoma, even when the sample has been obtained using adequate technique (5).

The Mexican Ministry of Health performs some 1 200 000 cytologic examinations of cervical smears annually. In order to promote the proper functioning of screening laboratories and to correct any deficiencies detected, since 1985 the National Institute of Epidemiologic Diagno-

sis and Reference (*Instituto Nacional de Diagnóstico y Referencia Epidemiológicos*—INDRE) has conducted a quality control program. The research study reported here was performed in 1994 and 1995 using the PAPNET system, which was employed by INDRE as an instrument to provide quality control in reading some 10 098 slides submitted for review by 28 states and the Federal District. The cytotechnologists and cytopathologist responsible for the study received prior training and were certified by the manufacturer of the automated system. Among other things, the study established the frequencies at which normal, premalignant, and malignant cells were being detected in the various Mexican states using conventional microscopy—as well as the frequencies of their detection during quality control work conducted by or through INDRE using both the automated PAPNET method and traditional microscopy.

As this suggests, the aim of our work was to compare the results of cervical smear analyses in state laboratories using conventional microscopy with the results of analyses conducted by INDRE for quality control purposes, using both automated and conventional methods, in order to determine the degree of agreement between these three sets of data.

MATERIALS AND METHODS

The PAPNET apparatus is a computerized device that, by means of optical density analysis, examines cervical smear slides and stores abnormal images on a cassette. This automated system detects suspect images in samples of cervical exudate mounted on conventional slides that have been fixed and stained by the Papanicolaou method. It has the ability to distinguish between normal and abnormal images (i.e., between artificial changes produced by manipulation, extracellular microorganisms, cells infected by intracellular micro-

organisms, dysplastic cells, and cancerous cells). In the process, the slides themselves are not altered in any way, retaining their integrity for subsequent microscopic examination (6).

In examining each slide, the system uses abnormality-detecting algorithmic procedures to select 128 images (64 frames, each containing two images amplified either 200 or 400 times), which are stored on a magnetic tape. For our study the system employed two stations at different locations. One, the examination station, was located at the offices of the Neuromedical Systems Company in Suffern, New York. This station received the slides, identified by bar codes, and the corresponding readings. Using the aforementioned optical density system, the slides were analyzed and abnormal images were stored on cassettes. The other, the verification station, was located at INDRE. Its equipment included a 200 megabyte computer with a 21-inch color monitor and inkjet printer. This station received the box with the images and diagnoses of the encoded slides, which were reviewed initially by a cytotechnologist and subsequently by a cytopathologist.

To conduct our study of diagnostic agreement between results obtained by the states, PAPNET, and INDRE, state laboratories attached to the Ministry of Health were invited to send a random sample of all the slides examined, preferably reflecting, in proper proportion, all the diagnoses made. The number of slides submitted by each laboratory was therefore dependent on its volume of work during its period of participation in the project, which lasted from August 1994 to June 1995. The states of Chihuahua, Oaxaca, and Querétaro opted to refrain from participating.

In all, 10 098 slides were received that had been prepared and subjected to conventional microscopy by technicians working at state public health laboratories in 28 Mexican states and the Federal District. The

slides and their diagnoses were subsequently forwarded to INDRE, where specialized personnel subjected each slide to examination via the automated PAPNET system and compared the results to those obtained by conventional microscopy. The information obtained was analyzed using the EPI-INFO package (version 6.0) and the SPSS package (version 5.1).

RESULTS

Table 1 shows the numbers and percentages of cervical smear slides obtained from each state and the Federal District. Of the 10 098 slides, the largest numbers came from the Federal District (1571) and the states of México (1184), Veracruz (802), Sinaloa (684), and Tabasco (632); the smallest numbers came from the states of Puebla (48), Jalisco (50), Morelos (68), Michoacán (80), and Guerrero (86).

Table 2 shows the diagnoses issued by the states, the PAPNET system, and INDRE (the latter following microscopic review). As is evident, more than 78% of the analyses from all sources generated negative results or revealed an inflammatory process, followed in order of frequency by mild and moderate dysplasias. According to the INDRE results, 3.6% of the samples contained insufficient material or were otherwise inadequate. It is also evident that the overall discrepancy between diagnoses made by the states and INDRE was greater than the overall discrepancy between diagnoses made by INDRE and PAPNET. In addition, PAPNET detected relatively few positively cancerous cases—a category encompassing all lesions unmistakably cancerous, including ones which, due to technical problems such as poor sample taking (for example, an excessive content of blood from hemorrhagic diatheses or an insufficient number of cells), cannot be properly classified with regard to their status. In this same vein, PAPNET more frequently classified samples as being insufficient or in-

adequate, and also as being characterized by cellular atypia.

In calculating agreement between the positive and negative diagnoses made by the states and INDRE, a Kappa correlation coefficient of 0.62, around the lower limit of concordance, was obtained for the overall analysis. With respect to particular diagnoses, the greatest diagnostic agreement (93.0%) was obtained with slides classified as normal or exhibiting inflammatory changes. In contrast, agreement was low-

Table 1. Numbers and percentages of Pap test slides sent to the central laboratory at the National Institute of Epidemiologic Diagnosis and Reference (INDRE) by public health laboratories of 28 states and the Federal District.

Origin of slides	No.	%
Aguascalientes	183	1.8
Baja California Norte	226	2.2
Baja California Sur	109	1.1
Campeche	154	1.5
Coahuila	97	1.0
Colima	319	3.2
Chiapas	201	2.0
Federal District	1 571	15.6
Durango	151	1.5
Guanajuato	436	4.3
Guerrero	86	0.9
Hidalgo	392	3.9
Jalisco	50	0.5
México	1 184	11.7
Michoacán	80	0.8
Morelos	68	0.7
Nayarit	526	5.2
Nuevo León	393	3.9
Puebla	48	0.5
Quintana Roo	180	1.8
San Luis Potosí	403	4.0
Sinaloa	684	6.8
Sonora	409	4.0
Tabasco	632	6.3
Tamaulipas	103	1.0
Tláxcala	118	1.2
Veracruz	802	7.9
Yucatán	201	2.0
Zacatecas	292	2.9
Total	10 098	100

Table 2. Cervical smear diagnoses (absolute numbers and percentages) made by the automated PAPNET system, INDRE, and public health laboratories of 28 states and the Federal District.

Diagnosis	INDRE		PAPNET		States	
	No.	%	No.	%	No.	%
Normal or inflammatory changes	8 572	84.9	8 508	84.3	7 964	78.9
Mild dysplasia	595	5.9	577	5.7	1 074	10.6
Moderate dysplasia	256	2.5	213	2.1	349	3.5
Severe dysplasia	87	0.9	77	0.8	139	1.4
Carcinoma *in situ*	0	—	0	—	2	0.0
Squamous carcinoma *in situ*	51	0.5	57	0.6	101	1.0
Invasive squamous carcinoma	72	0.7	41	0.4	44	0.4
Adenosquamous carcinoma	2	0.0	1	0.0	1	0.0
Microinvasive carcinoma	0	—	0	—	9	0.1
Invasive endocervical adenocarcinoma	0	—	0	—	2	0.0
Carcinoma (unspecified)	46	0.5	16	0.2	27	0.3
Endometrial adenocarcinoma	1	0.0	1	0.0	2	0.0
Atypia	52	0.5	105	1.0	1	0.0
Endocervical atypia	4	0.0	3	0.0	5	0.0
Inadequate or insufficient sample	360	3.6	495	4.9	31	0.3
No result	0	—	3	0.0	347	3.4
Total	10 098	100	10 098	100	10 098	100

est when it was necessary to distinguish between degrees of abnormality: carcinoma (of unspecified status), 13.3%; invasive carcinoma, 51.4%; and moderate dysplasia, 63.1% (Table 3).

The highest percentages of underestimation were found for diagnoses of carcinoma (of unspecified status) and invasive carcinoma. Although the underestimation observed was lower in cases of atypia or mild dysplasia (11.5%), moderate dysplasia (19.8%), and severe dysplasia or *in situ* car-

cinoma (13.2%), the numbers are nevertheless alarming, because it is in these stages that the patient may be treated to prevent evolution to carcinoma (6).

When the diagnoses made by INDRE and PAPNET were compared, a Kappa correlation coefficient of 0.80, located around the upper limit of concordance, was obtained. As Table 4 shows, in percentage terms there was good agreement (98.8%) regarding positive and negative diagnoses (the latter including cells found to exhibit

Table 3. Degree of agreement (%) between the diagnoses made by the states using conventional microscopy and those made by INDRE, showing the percentages of apparent underestimation and overestimation relative to the INDRE diagnosis in each case.

Diagnosis	Agreement (%)	Underestimation (%)	Overestimation (%)
Normal or inflammatory changes	93.0	0.0	6.9
Atypia or mild dysplasia	74.1	11.5	14.4
Moderate dysplasia	63.1	19.8	17.1
Severe dysplasia or carcinoma *in situ*	78.5	13.2	8.3
Carcinoma (unspecified)	13.3	68.9	17.8
Invasive carcinoma	51.4	48.6	0.0

Table 4. Degree of agreement (%) between the diagnoses made by the automated PAPNET system and those made by INDRE, showing the percentages of apparent underestimation and overestimation relative to the INDRE diagnosis in each case.

Diagnosis	Agreement (%)	Underestimation (%)	Overestimation (%)
Normal or inflammatory changes	98.8	0.0	1.2
Atypia or mild dysplasia	79.6	18.0	2.4
Moderate dysplasia	71.1	25.6	3.3
Severe dysplasia or carcinoma *in situ*	71.8	26.6	1.6
Carcinoma (unspecified)	26.3	68.4	5.3
Invasive carcinoma	60.3	39.7	0.0

inflammatory processes), but percentage agreements indicating PAPNET's ability to distinguish between various pathologic alterations were noticeably lower; in addition, the percentages of underestimation recorded using this automated system were quite high (6).

As Table 5 indicates, PAPNET detected fewer specific infectious processes than did traditional microscopy employed by either INDRE or the states.

Table 6 presents an overall account by patient age group of the INDRE diagnoses. As may be seen, severe dysplasia or *in situ* carcinoma was detected at a relatively high frequency on slides from women 25–54 years old, while invasive carcinoma was diagnosed primarily on slides from women 35 and over.

DISCUSSION

Judging from the results obtained in this study, there is a need to improve the quality of the cervical cytology diagnoses issued by state public health laboratories in Mexico. It is particularly important that the ability to correctly identify the various abnormalities of cellular morphology be improved, in order to minimize over- and

Table 5. Infectious disease agents detected on cervical smears examined by INDRE, PAPNET, and state laboratories.

Disease agent	INDRE		PAPNET		States	
	No.	%	No.	%	No.	%
Trichomonas vaginalis	455	4.5	396	3.9	478	4.7
Candida albicans	387	3.8	249	2.5	404	4.0
Human papillomavirus	348	3.4	224	2.2	722	7.1
Trichomonas vaginalis plus human papillomavirus	17	0.2	12	0.2	34	0.3
Candida albicans plus human papillomavirus	10	0.2	6	0.1	25	0.2
Candida albicans plus *Trichomonas vaginalis*	5	0.0	5	0.0	11	0.1
Other causative agents	3 767	37.3	3 754	37.2	3 080	30.5
None	5 109	50.6	5 452	54.0	5 344	52.9
Total	10 098	100	10 098	100	10 098	100

Table 6. INDRE diagnoses of the cervical smears examined, by patient age group.

Age (in years)	Normal or inflammatory changes	Atypia* or mild dysplasia	Moderate dysplasia	Severe dysplasia or *in situ* squamous carcinoma	Carcinoma (of indeterminate status)	Invasive carcinoma[†]	Inadequate or insufficient sample	Total No.	Total %
≤15	22	3	0	0	0	0	1	26	0.3
16–24	1 712	145	37	12	1	0	62	1 969	19.5
25–34	2 963	192	74	54	11	3	105	3 402	33.7
35–44	2 106	178	80	35	14	10	78	2 501	24.8
45–54	936	68	32	27	8	21	51	1 143	11.3
55–64	373	31	11	2	8	19	27	471	4.7
≥65	132	6	13	4	3	16	10	184	1.8
Unknown	328	24	9	4	1	6	26	398	3.9
Total	8 572	647	256	138	46	75	360	10 094	100.0

* Does not include endocervical atypia.
† Includes invasive squamous carcinoma, adenosquamous carcinoma, microinvasive carcinoma, and endometrial adenocarcinoma.

underestimating of the various pathologic states and thereby pave the way for adequate preventive and therapeutic measures.

In order to obtain correct diagnoses from examination of cervicovaginal smears, cytotechnologists should be highly trained and supervised by a physician cytologist, a pathologist, or both. For this reason, various types of supervision and training should be systematically carried out, together with courses designed to improve skills in the taking, fixing, staining, and mounting of samples, as well as other technical aspects of cervical smear testing.

Our results showed that the use of PAPNET led to various technical problems, including lower resolution of images when the device providing the greatest degree of magnification (400x) was used, with consequent loss of definition or disappearance of the cell from the field of view. Also, in order to ensure adequate quality it was necessary to have at least 10% of the surface of the slide covered with cellular material.

Another significant criterion for classifying a slide sample as satisfactory (that is, a criterion affirming that it came from the endocervical canal) is the presence of groups of columnar cells or cells exhibiting squamous metaplasia, neither of which can be detected by PAPNET. On 30% of the slides, PAPNET detected large numbers of inflammatory cells (including polymorphonuclear leukocytes and lymphocytes but no epithelial cells) in 64 of the 128 frames examined. Likewise, when the slide had been poorly fixed, PAPNET was unable to identify artificial changes produced by manipulation of the sample, such as "coffee grounds," impeding observation of cell nuclei. Particles of talc and other similar elements also confound the PAPNET system because of a shortage of epithelial cells with normal morphology in the images.

When the sample has been poorly spread on the slide (it may be thick, as when dealing with a microbiopsy), it is impossible to make an interpretation, as the superimposition of cells prevents detection of abnormality. In cases of invasive carcinoma, the cells tend to be poorly preserved. However, an expert observer can detect them under the microscope, something that cannot be done with the PAPNET system because the cells cannot be seen via the monitor. In addition, the specificity of PAPNET increases at the expense of its sensitivity (the number of false positive results), and as a result a high number of false atypias is diagnosed.

Review of each slide by the PAPNET system typically takes from 45 seconds up to 2 minutes for the most difficult cases. In our study the process took up to 8 minutes per case because of the various steps involved in using PAPNET (obtaining properly taken samples and affirming their adequate fixing, staining, and mounting). It was also necessary to consider the time needed to place standard-sized coverslips over the specimens; bar-encode each slide; package the samples appropriately; send the slides by courier to the central station in New York; await arrival of the interpretation of each tape and slide (in three to five days); and verify the number of slides that were misdirected or arrived broken. Beyond that, it was necessary to consider the time taken by the INDRE cytotechnologist and pathologist to review each slide under the microscope.

An additional deficiency of the PAPNET system is frequent underestimation of infectious processes, such as human papillomavirus infection, *Candida* infection, and trichomoniasis. It should be pointed out that if state laboratory technicians experienced difficulties in properly diagnosing associated lesions, as suggested by the low degree of agreement between their diagnoses and those of INDRE, the problem would be even more pronounced if interpretations were made using the automated system.

Overall, the results of this study demonstrate that use of PAPNET involves

greater difficulties than conventional microscopic examination of cervical smears—despite the fact that PAPNET was used only to distinguish normal cases from suspected cases that should be subjected to conventional microscopy, thus eliminating a step that tends to overburden cytotechnologists and pathologists. It also seems clear that a way needs to be found to detect and review the false negative results generated by PAPNET before granting approval to a technology that, so far, has failed to demonstrate its superiority to conventional microscopy.

REFERENCES

1. Tolles WE, Bostrom RC. Automated screening of cytological smears for cancer: the instrumentation. *Ann N Y Acad Sci* 1956;63: 1211–1218.

2. Specification of automated cytodiagnostic system proposed by the International Academy of Cytology. *Ann Quant Cytol* 1984;6:146.

3. Bogdanich W. The Pap test misses much cervical cancer through lab's error. *Wall Street Journal.* 2 November 1987.

4. Tucker J, Stenkvist B. Whatever happened to cervical cytology automation? *Ann Cell Pathol* 1990;2:259–266.

5. Koss L, Eunice L. Evaluation of the Papnet cytologic screening system for quality control of cervical smears. *Am J Clin Pathol* 1994;101:220–229.

6. Marino Conde E. Observaciones y mediciones. In: Moreno AL, Cano VF, García RH, eds. *Epidemiología clínica* 2nd ed. Mexico City: Interamericana McGraw-Hill; 1994:69–98.