Diagnostic Reproducibility of Pap Testing in Two Regions of Mexico: The Need for Quality Control Mechanisms

P. Alonso de Ruiz,¹ E. C. Lazcano Ponce,² R. Duarte Torres,¹ I. Ruíz Juárez,³ & I. Martínez Cortez⁴

To assess the reproducibility of diagnostic results obtained by examining Pap smears for cervical neoplasia, a study was conducted using a single group of 20 Pap smears, 3 negative and 17 from patients with varying degrees of neoplasia. These smears were examined by 14 volunteer readers (13 cytotechnologists and 1 cytopathologist) from the Mexican states of Oaxaca and Veracruz, and also by a highly experienced cytopathologist certified by the Mexican Board of Pathological Anatomy whose work provided a reference standard.

Individual variability, as assessed by the Kappa coefficient of concordance, showed considerable difference in the diagnostic results obtained by different readers—the degree of agreement depending on the type of cervical lesion involved and the number of specimens from patients with that type of lesion. There was little diagnostic agreement when the specimens were assessed for particular classes of cervical neoplasia—mild, moderate, or severe neoplasia, carcinoma in situ, or invasive cervical cancer. (The greatest concordance was found in diagnosing specimens from subjects with invasive cervical cancer.) However, when the diagnosis was assessed continuously, using Kappa weighted in accordance with the five possible diagnoses of cervical neoplasia, the apparent reproducibility of the diagnoses improved greatly, Kappa coefficients for the 14 readers ranging from 0.31 to 0.72.

In general, these data support the view that there is a need in Mexico and other parts of the Americas to establish quality control mechanisms monitoring cytologic diagnosis of cervical neoplasia, to standardize diagnostic nomenclature using a system such as the Bethesda System, to institute periodic certification, and to provide continuing training. As this suggests, it is necessary not only to evaluate but also to bring about organizational changes in order to expeditiously prevent or correct the problems that currently constrain achievement of efficient and effective cytologic diagnosis.

Quality control programs are important for maintaining diagnostic accuracy in gynecologic cytology (Pap testing), so as to ensure timely detection of cancer of the uterine cervix in the population. Quality control mechanisms for Pap testing include monitoring of guidelines and adequate reproducibility with regard to speci-

¹ Pathology Unit, National Autonomous University of Mexico, General Hospital, Ministry of Health, Mexico City, Mexico.
² Center for Research on Population Health, National Institute of Public Health, Cuernavaca, Morelos, Mexico. Reprint requests and other correspondence should be addressed to Dr. Eduardo César Lazcano Ponce, Centro de Investigaciones en Salud Poblacional, Instituto Nacional de Salud Pública, Avenida Universidad 655, Colonia Sta. María Ahuacatitlán, C.P. 62508, Cuernavaca, Morelos, Mexico. E-mail: elazcano@insp.mx.
³ Veracruz University, Veracruz, Veracruz, Mexico.
⁴ Dr. Aurelio Valdivieso Civil Hospital, Ministry of Health, Oaxaca, Oaxaca, Mexico.
men taking, fixing, staining, and reading. Such mechanisms can contribute to increasing the diagnostic benefits obtained from Pap testing by reducing the numbers of false negative and false positive readings, and can also reduce the cost of caring for patients with invasive cervical neoplasia (1–2).

As a central element in cancer prevention programs that has reduced death from cervicouterine cancer in many countries by more than 70% (3–7), the Pap test is considered one of the greatest diagnostic successes in medical practice. However, no such impact has been observed in many developing countries, including some in Latin America. Indeed, it has been estimated that shortcomings of programs for timely detection of cervicouterine cancer among the general public of developing countries account, in some cases, for over 60% of all deaths from cervical neoplasia (8). This is due, among other things, to the unreliability of results provided by centers reading gynecologic cytology specimens, low levels of coverage for at-risk women, and poor medical care quality (9).

The Pap test can truly contribute effectively to cervical cancer prevention only if a quality control program is established in each cytodiagnostic laboratory to ensure that Pap smears are obtained effectively and efficiently, and also to ensure diagnostic accuracy. Both of these essential activities, in turn, depend primarily upon the availability of specialized training, continuing education, and prior institutional certification.

The work reported here provides an assessment of Pap diagnostic reproducibility, using samples from 20 subjects with cervical neoplasia and other cervical conditions. These samples were examined by 14 readers participating in a program for timely detection of cervico-uterine cancer in the Mexican states of Veracruz and Oaxaca.

METHODS

General Procedures

The 20 Pap smears examined included three negative smears (one each from a subject having cervicitis with metaplasia, hyperkeratosis, and human papillomavirus infection) and 17 showing different clinical stages of cervical neoplasia (seven with moderate dysplasia, one with severe dysplasia, six with carcinoma in situ, and three with invasive cervical cancer). Eleven of the positive smears were also found positive for human papillomavirus infection. Each of the 14 readers, who included 13 cytotechnologists and one cytopathologist, described the morphology of cervical lesions indicated by their examination of these smears, using the methodology and nomenclature employed for external evaluation purposes by the National Reference Center for Pap Quality Control in Mexico.

Participation in the study was voluntary. Random identification cards were used so as to make the identity of the participating readers unknown. The conditions under which Pap smear diagnoses are generally made in Mexico were simulated, and each new reading was made independently of the prior reader or specimen, permitting the assumption that diagnostic error was randomly distributed.

A reference reading of the specimens was made by the chief of the cytopathology laboratory in the pathology unit of the National Autonomous University of Mexico, Mexico City General Hospital, who had 30 years of experience and was certified by the Council of Pathologic Anatomy of Mexico.

The Pap smears evaluated were obtained from the cytopathology laboratory at the Mexico City General Hospital. Thus, for patients with positive cases of cervical neoplasia, the histopathologic result was available. In addition, the quality of the specimens involved was considered appropriate for making a diagnosis.
The cytologic diagnoses made by the readers were set down on a form designed by the cytopathology laboratory of the Mexico City General Hospital to assess a person's ability to diagnose a cervical lesion. This form offered the following alternatives: (1) normal smears, (2) human papillomavirus infection, (3) mild dysplasia, (4) moderate dysplasia, (5) severe dysplasia, (6) carcinoma in situ, and (7) invasive cervical cancer.

The diagnostic classification and categorization used for evaluation purposes in this study matched those used for Pap test diagnosis at gynecologic cytology reading centers in Mexico and employed the reference diagnostic nomenclature for dysplasias used by the National Cytology Reference Center of Mexico. Graphic presentations were devised to illustrate the individual variability observed in diagnosing the various degrees of cervical neoplasia.

Statistical Analysis

The cytology readings obtained were analyzed statistically to determine their reproducibility, based on comparison against a standard and estimation of the Kappa intraclass coefficient (10–11) and also Kappa weighted using four categories: (1) normal or mild dysplasia, (2) moderate or severe dysplasia, (3) in situ cancer, and (4) invasive cervical cancer.

The Kappa concordance coefficient is used to compare the reproducibility of two or more measures (in this case readings), since for intraclass correlations (binary data) it represents the best coefficient of indirect agreement. This is distinct from situations in which cervical neoplasia is evaluated as an ongoing event, where weighted Kappa is used (12).

The Kappa concordance coefficient expresses the relationship between the results observed and the results expected from a particular mathematic model. Its value depends closely on the prevalence of cases in the study population. When there is a tendency toward misclassification error of two or more different specimens associated with the same individual reader, this can be ascribed mainly to two factors: (a) the two reading measures taken were imperfect or (b) one assessment was not performed independently of the other.

To interpret the results, the guidelines proposed by Landis and Koch (13) were used. According to these guidelines, a Kappa value of <0.0 concordance is very poor, 0.0–0.20 is slight, 0.21–0.40 is fair, 0.41–0.60 is moderate, 0.61–0.80 is strong, and 0.81–1.0 is almost perfect. The Statistical Package for the Social Sciences was employed to process the results (14).

RESULTS

Group Observation

Considerable variation was observed in the diagnosis of specific grades of cervical neoplasia. In this regard, the Kappa coefficients indicate that diagnostic reproducibility was closely related to the type of cervical lesion involved and the prevalence of that type of lesion among the 20 smears examined (Table 1). Kappa values for detection of human papillomavirus infection were fair (Kappa = 0.39) in the total sample, moderate (Kappa = 0.42) in Veracruz, and fair (Kappa = 0.34) in Oaxaca.

Despite the fact that seven specimens (35% of the total) indicated moderate dysplasia, diagnostic concordance here was only slight for the total sample (Kappa = 0.17). The Kappa value (0.02) relating to the lone smear (5% of the total) indicating severe dysplasia reflected a very low (null) diagnostic reading agreement associated with the low prevalence of this type of lesion in the study sample. The Kappa value relating to diagnosis of cancer in situ (Kappa = 0.14) was slight. The best diagnostic agreement was found for invasive cervical cancer, where the Kappa values
were fair in Veracruz (0.31) and moderate in Oaxaca (0.43).

Individual Observation

As can be seen in Figure 1, the results obtained by the 14 different readers in identifying human papillomavirus infection varied considerably. One reader had null Kappa agreement, one had slight agreement, five had moderate agreement, six had strong agreement, and one had very strong agreement. There was also considerable interobserver variation regarding detection of various types of cervical neoplasia. Regarding other diagnoses, the number of readers with null (Kappa <0.0) intraclass concordance was six (42.9%) in regard to moderate dysplasia, 13 (92.9%) in regard to severe dysplasia, four (28.6%) in regard to cancer in situ, and one (7.1%) in regard to invasive cervical cancer (Figure 2). The best intraclass concordance occurred with respect to diagnosis of invasive cervical cancer, where five readers had Kappa values exceeding 0.50.

Figure 1. Intraclass Kappa values for diagnosis of human papillomavirus infection by the 14 readers examining the 20 study specimens.
Figure 2. Intraclass Kappa values for diagnosis of moderate dysplasia, severe dysplasia, *in situ* carcinoma, and invasive cervical cancer by the 14 readers examining the 20 study specimens.
Kappa Weighted When Cervical Cancer Was Assessed Continuously

As the data in Table 1 and Figures 1 and 2 indicate, the level of diagnostic reproducibility was poor when the cervical cytopathology diagnoses were evaluated on an intraclass basis. However, when the diagnosis was assessed continuously, using Kappa weighted in accordance with the four possible diagnoses of cervical neoplasia (1 = normal smears or mild dysplasia, 2 = moderate or severe dysplasia, 3 = cancer in situ, and 4 = invasive cervical cancer), the reproducibility of cervical neoplasia diagnosis improved greatly. As Figure 3 shows, in this latter case the concordance coefficients for the various readers fell between 0.31 (fair) and 0.72 (strong). This finding provides empirical evidence of the incorrectness of using intraclass concordance as an external quality control measure for a continuous event such as cervical neoplasia and points up the Regional need to propose diagnostic classifications based on current knowledge of the natural history of cervical cancer.

DISCUSSION

General Considerations

The Pap test is a diagnostic procedure subject to error whose evaluation depends on adopted criteria. Traditionally, the diagnostic accuracy of Pap test diagnosis is measured by comparing it to a “gold standard,” which in pathologic anatomy is the histologic diagnosis. Deficiencies in the method produce false negative or false positive results. Diagnostic accuracy is usually expressed in terms of sensitivity and specificity, and may be determined when the histologic reference for the cases is available and there exists a preponderance of negative cases. These measures of diagnostic accuracy show in terms of probability whether the procedure has correctly identified cervical disease or health status (11). When it is not possible to obtain a cervical biopsy, the reproducibility of diagnoses by multiple readers can be used as the evaluation criterion, as was done in the study reported here.

From the public health perspective, quality control and measurement of the index

Figure 3. Weighted Kappa values for diagnosis of cervical neoplasia by the 14 readers examining the 20 study specimens, treating the observed cervical neoplasias as ongoing events and using the following evaluation categories: (1) normal smears or mild dysplasia, (2) moderate or severe dysplasia, (3) carcinoma in situ, and (4) invasive cervical cancer.
of error are basic elements in timely cervical cancer detection programs serving the general population. However, the incidence of false negative results depends on the quality of the cytodiagnostic laboratory, which arises not merely from internal and external Pap diagnosis evaluation but also from such fundamental things as the presence of experienced and accredited staff members, available materials and resources, ease of operation, and accessibility (15–18).

In other words, in order to ensure diagnostic accuracy in the reading of Pap test results, it is necessary to have a reliable cytopathology laboratory. Overall, there must be an effective interaction between the quality of the procedure by which the specimen is taken and subsequently prepared and the reading of the Pap test by qualified personnel, all of this leading to a high level of diagnostic accuracy. A schematic version of this interaction is shown in Figure 4.

In Mexico and elsewhere in the Americas, early cervical cancer detection is impeded by significant problems associated with supervision of cytodiagnostic laboratories, among them the lack of implementation of Pap test quality control measures and deficient accreditation of specialized staff members.

As noted by Donabedian (19), quality assessment is not possible without criteria and standards—criteria being elements that allow us to make a judgement and standards being more specific quantitative guidelines that make it possible to determine both magnitude and frequency—such as the number of cytologic specimens that should be examined by a cytologist in

![Figure 4. A schematic diagram of factors influencing the quality of results obtained by a cytopathology laboratory.](image-url)

<table>
<thead>
<tr>
<th>FACTORS INFLUENCING SPECIMEN QUALITY:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Specimen procurement technique, cell representativeness of:</td>
</tr>
<tr>
<td>- Endocervical cells (from transformation zone of cervical canal)</td>
</tr>
<tr>
<td>- Epidermoid metaplasia</td>
</tr>
<tr>
<td>- Mucus</td>
</tr>
<tr>
<td>2) Recording of patient information:</td>
</tr>
<tr>
<td>- Socioeconomic characteristics</td>
</tr>
<tr>
<td>- Gynecological history</td>
</tr>
<tr>
<td>- Use of contraceptive methods</td>
</tr>
<tr>
<td>- Date of last menstruation</td>
</tr>
<tr>
<td>- Signs and symptoms at the time sample was taken</td>
</tr>
<tr>
<td>3) Specimen preparation technique:</td>
</tr>
<tr>
<td>- Distribution of the sample on the slide</td>
</tr>
<tr>
<td>- Fixation</td>
</tr>
<tr>
<td>- Staining</td>
</tr>
<tr>
<td>4) Transport to the cytopathology laboratory</td>
</tr>
<tr>
<td>5) Monitoring of productivity criteria</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FACTORS INFLUENCING DIAGNOSTIC QUALITY:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Laboratory personnel</td>
</tr>
<tr>
<td>2) Appropriate training:</td>
</tr>
<tr>
<td>- Institutional diagnostic certification (for cytotechnologists and cytopathologists)</td>
</tr>
<tr>
<td>- Continuing education</td>
</tr>
<tr>
<td>3) Prior experience</td>
</tr>
<tr>
<td>4) Random diagnostic error</td>
</tr>
<tr>
<td>5) Observer variability:</td>
</tr>
<tr>
<td>- Similar diagnostic criteria</td>
</tr>
<tr>
<td>6) Interlaboratory variability:</td>
</tr>
<tr>
<td>- Similar nomenclature</td>
</tr>
<tr>
<td>- Similar recording form</td>
</tr>
<tr>
<td>7) Diagnostic concordance:</td>
</tr>
<tr>
<td>- Cytohistologic case correlation</td>
</tr>
<tr>
<td>8) Diagnostic monitoring by the cytopathologist</td>
</tr>
</tbody>
</table>

Computerized Information System

QUALITY CONTROL

DIAGNOSTIC ACCURACY
- Final diagnosis
- Integrated reporting system

---

336 Bulletin of PAHO 30(4), 1996
a given time and the percentage of false negatives to be expected from a cyto-diagnostic center.

The aim of monitoring Pap test activities in a cytopathology laboratory is to ensure the diagnostic accuracy of Pap testing based on continuing review of the effectiveness and efficiency of the reading centers. Such review is accomplished largely through the following three actions: (1) evaluation of the cytopathology laboratory's quality by measuring its performance against appropriate criteria and standards; (2) localization of deficient functions; and (3) development of proposals for corrective action with regard to resources, procedures, functions, duties, education, and incentives. These steps should theoretically be carried out in a series of continuing cycles mirroring the continuing activities that must be implemented by cancer prevention programs in Latin America.

Improvement of Pap Testing

The Mexican research effort reported here demonstrated considerable diagnostic variation in Pap smear examinations directed at identifying morphologic transformations of the cervical epithelium and poor reproducibility of results. In addition, it indicated a need for a more efficient diagnostic nomenclature—such as the Bethesda System proposal, which has not yet been implemented in the Region—as well as for continuing training in cervical cytopathology.

The principal limitation of the proposed nomenclature is that, despite an increase in diagnostic concordance, in 30% of the low-grade lesions the diagnosis is not confirmed by cervical biopsy. Indeed, between 10% and 20% of these low-grade lesions appear to coexist with high-grade lesions (20), so the prevalence of poorly defined atypical cells can be very great.

For this reason, faced with the need to standardize cervical cytopathology reporting systems, we require a classification system consistent with our existing knowledge of the natural history of cervical neoplasia. Accordingly, the authors recommend a simplified classification system of five categories (negative, low-grade intraepithelial squamous lesions, high-grade intraepithelial squamous lesions, carcinoma, and atypical cells of indeterminate significance), adapted from the Bethesda System.

At present there is little information available in Latin America about the quality of Pap test diagnosis in mass detection programs. Factors reducing the reliability of cytologic diagnosis in the region include lack of government regulation of the practice of cytopathology; use of various different diagnostic nomenclatures; lack of registration, certification, and periodic recertification of Pap test reading centers, cytotechnologists, and pathologists; lack of continuing training for the latter personnel; lack of quality control mechanisms; and lack of computerized information registers. The implementation of strategies designed to deal with these problems is necessary in order to improve the efficiency and effectiveness of this medical intervention and thus achieve a generally positive effect upon the health of women in the region (21, 22).

Acknowledgment. We are grateful for the kind and objective services of the group of readers who participated in this study.

REFERENCES


Manuscript received 6 June 1996. Accepted for publication (following revision) in Spanish in the *Boletín de la Oficina Sanitaria Panamericana* and in English in the *Bulletin of the Pan American Health Organization* on 19 August 1996.