A Cytological Scale for Cervical Carcinogenesis

Elizabeth Stern, Alan B. Forsythe, Lee Youkeles, and Wilfrid J. Dixon

School of Public Health [E. S. J] and Department of Biomathematics, School of Medicine [A. B. F., L. Y., W. J. D.], University of California, Los Angeles, California 90024

SUMMARY

We have postulated a gradient of changes in cervical carcinogenesis ranging from preliminary nonspecific cellular modifications and atypias, through dysplasia as a precursor state, with increasing degrees of severity, to the noninvasive and invasive stages of cancer. In this paper we describe construction and validation of a numerical scale in relation to the proposed order of progression. We present evidence in support of both the ordering of the categories, defined on the basis of cytological criteria, and the calibration and reliability of the scale. The scale provides a quantitative measure for use in studying natural progression in a defined population. The index of cytological progression can also be used as a biological measure in estimating adverse effects, to be correlated with exposure to the suspected causal factor.

INTRODUCTION

With the advent of the Papanicolaou (Pap) smear, a new concept of the pathogenesis of cancer was introduced when it was found that abnormal exfoliated cells could be correlated with histological changes consistent with neoplasia but confined to the mucosa. The noninvasive intraepithelial lesions were designated as dysplasia and in situ cancer. Implicit in this concept is the assumption that there is a preinvasive stage of cancer and that dysplasia is a precursor state.

The status of dysplasia as a cancer precursor and the relationship of preinvasive cancer to the invasive and clinical stages of cancer have not yet been fully established. The sequential relationships may be clarified by studies on the natural history of cancer of the cervix in which continuity of the changes over time can be confirmed by repeat testing of a population of women kept under observation. There is little likelihood of following women once in situ cancer is diagnosed, since intervention may occur even before this stage is reached. However, information on the dynamics of this process may be provided by prospective study of women with dysplasia.

The Pap smear is a test that can be conveniently repeated and is suitable for such studies. The results of this test have traditionally been reported in terms of the 5 classes introduced by Papanicolaou in 1954 (4): Class I, absence of atypical or abnormal cells; Class II, atypical cytology but no evidence of cancer; Class III, cytology suggestive of, but not conclusive for, cancer; Class IV, cytology strongly suggestive of cancer; Class V, cytology conclusive for cancer. This system of classification has been useful because, as Papanicolaou stated, "smears cannot always be judged as positive or negative." However, there was no specific class provided for dysplasia, although Papanicolaou had described the typical dyskaryotic cells that exfoliate from this lesion (4).

A related problem is the variability in diagnosis of the intraepithelial lesions. In this regard, a committee appointed to study reproducibility of histopathological diagnosis reported that pathologists show a much greater ability to assign a lesion to a position on a numbered scale reflecting a biological spectrum of progressively advancing disease than to agree on a diagnosis by a named category (3). The variability considered by the committee was in relation to histological findings, but the problem applies equally to the classification and interpretation of cytological findings.

In order to study the transition from a normal state to cancer, we have postulated a spectrum of change from completely negative, through a range of cell modulations, atypias, and degrees of dysplasia, to cancer. This progression can be represented as an ordered series of changes based on cytostatological observations in the laboratory and is consistent with our perception of the natural history of the disease (6).

The ordered categories suggested by the biological criteria make one hope that it is worthwhile to look for a scale. If these categories can be assigned numerical values with appropriate spacing, the usefulness of the cytological evaluation is improved, particularly in longitudinal studies. The advantages of scalar data over categorical data are especially evident in analyzing results of such studies. For example, concomitant variables can be considered more effectively when the outcome is scalar. We can use standard procedures, have more specific interpretation, and use smaller sample sizes.

In this paper we describe construction and validation of a numerical scale in relation to the proposed order of progression.

PROCEDURES AND FINDINGS

The data base utilized in the analysis is from a prospective study of the effects of a steroid contraceptive on the progress of dysplasia. The women with dysplasia are detected by Pap smear screening of subjects reporting for an initial visit to community family planning clinics. A systematic sample of the women with negative Pap smears is also being followed. Our aim was to develop a cytological scale to be used to provide a quantitative estimate of the degree of change noted in the Pap smear.
test at entry and at each follow-up visit in the longitudinal study. Women participating in the follow-up are tested at regular intervals. Slides are read without knowledge of clinical status or contraceptive method. Slides to be read, including slides from previous visits, are placed at random in slide folders without identification. Slides from previous visits are reread to ensure reliability and guard against possible long-term trends in screening and rating. Slides are first screened by a cytotechnologist and then reviewed and assigned to a category by the one pathologist. The results of blind replicate readings by a single observer are thus available.

Steps toward formulating a scale include: Step 1, considering only 2 possible states, noncancer or cancer; Step 2, placing dysplasia between these; Step 3, identifying additional cytological categories within these 3 states; Step 4, placing an ordering on the categories; Step 5, assigning numerical values to each of the ordered categories.

The 1st 2 steps are well accepted. At Step 3, additional categories have been recommended for dysplasia (mild, moderate, and severe) and for cancer [in situ, in situ with minimal stromal invasion, microcarcinoma, and invasive cancer (5)]. For this study we have added categories at the negative or low end of the biological spectrum (metaplasia, atypia, borderline dysplasia). The main focus of this report is on Steps 4 and 5; that is, verifying the ordering of categories and estimating the spacing or distance between them.

Ordering of Cytological Categories. Perhaps the most important single criterion in the development of a scale is to determine whether the original cytohistological categories are indeed ordered in terms of the underlying pathology. This proposition was tested by the criterion that if the ordering is correct the transition over a short period of time should be more frequent between categories close to each other than between distant categories. The relative frequency of moving from 1 category to another within a 6-month period of follow-up is illustrated in the transition matrix (Table 1).

<table>
<thead>
<tr>
<th>6 mo.</th>
<th>Negative</th>
<th>Metaplasia</th>
<th>Atypia</th>
<th>Borderline dysplasia</th>
<th>Minimal dysplasia</th>
<th>Slight</th>
<th>Moderate</th>
<th>Severe</th>
<th>Borderline cancer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 mo.</td>
<td>62</td>
<td>32</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>49</td>
<td>27</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>15</td>
<td>63</td>
<td>15</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>28</td>
<td>37</td>
<td>27</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>23</td>
<td>40</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>13</td>
<td>38</td>
<td>25</td>
<td>13</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1 shows the transition pattern when tests of 312 women at the 6-month follow-up are compared with their tests 6 months later at the 12-month follow-up. Categories in increasing order of severity are listed on the left side and across the top. The left side represents status at 6 months; the top is status 6 months later. For example, among the 67 cases with borderline dysplasia at 6 months, 37% were still there at 12 months, while 28% were 1 category lower at atypia, and 27% were 1 category higher at minimal dysplasia, but only 2% had skipped a category back to metaplasia, and only 3% had skipped a category in the other direction to slight dysplasia. Thus there is a higher probability of transition from a given score to an adjacent score than to a more distant score. This closeness of ordering may be measured by ranking the categories of Table 1 and correlating the rank at 6 months with the rank at 12 months. This was calculated to be 0.78. These findings support the assumption that the cytohistological categories are in order of progressive severity.

Conversely, if we reordered the categories, there was, in general, a higher probability of transition from a given score to a more distant score than to an adjacent score. We also found that no reordering of the categories produced a higher rank correlation than the proposed order.

Spacing of Cytological Categories. The next step was to determine a scale for the categories. A useful spacing of the categories would meet the assumptions of approximate Gaussian (normal) distribution and homogeneous variance across the range of values. Scales that are compatible with these 2 criteria often can be considered additive in the sense that a change of, e.g., 5 units has the same meaning across the range of measurements. Such scales are easier to work with and easier to interpret. It has been shown that homogeneity of variances is much more important than the precise shape of the distribution (1).

The overall range of the categories was from 0 (negative) to 100 (invasive cancer), with the score for minimal dysplasia set...
at about midway between negative and in situ cancer. The frequency distribution of these categories was determined for the sample of approximately 300 women with normal smear tests and 300 women with dysplasia in the follow-up study. The distribution for the overall community family planning clinic population given in Table 2 was estimated on the basis that the follow-up study population was made up of a 1 in 20 sampling of normals as a comparison for the 5% of dysplasia prevalent in those clinics.

Initially, the scores assigned to the categories, defined on the basis of cytological criteria, had been equally spaced. A description of our cytological criteria and accompanying photomicrographs have been published (6). The empirical scale values, reflecting the cytopathologist's concept of the spacing between categories, are given in Table 2, Column A.

We then took advantage of the fact that most distributions look Gaussian (normal) if one looks only at the center of the distribution. The spacing was altered to fit a Gaussian distribution modified to allow for the possibility that women with dysplasia or cancer are less likely to be represented in our population than in a previously unscreened population. Many of these women would have been referred for treatment and thus would not have been included among the 30% who had been given a Pap test prior to their initial visit to the family planning program. Using "normal-probability paper" (2), the approximation of a normal distribution produced scale values as indicated in Table 2, Column B. This application of the normal distribution modified to allow for the possibility that women with dysplasia or cancer are less likely to be represented in our population than in a previously unscreened population.

If this spacing were to be considered correct, we would expect uniformity in the average standard deviation of replicate readings along the entire scale. However, the new spacing resulted in a larger standard deviation of replicate readings at the low end of the scale when compared to the high end, so that the criterion of homogeneous variability throughout the scale was not fulfilled. Trial and error of several spacings at the lower end of the scale finally produced score values at the low end of the spectrum with a replicate reading variance very similar to that in the dysplasia range (see Table 2, Column C, for this final set of scores). This scale, in which the error is constant along the negative and dysplasia range, is also consistent with the assumption of an approximate Gaussian distribution. More work is needed on score values at and beyond the in situ level because of the dearth of cases in this population.

Reproducibility of Readings. Because of the policy of rescreening previous slides along with the new slide obtained at the current visit, replicate readings for each slide were available. An analysis was performed on 1200 readings of about 400 slides for which there were at least 3 replicates. This resulted in a pooled standard deviation of 4.3 units of the cytological scale for replicate readings of slides in the negative range and a pooled standard deviation of 5.1 units in the dysplasia range (see Table 3).

The reliability coefficient (7) takes into account the variation both within and between smears and has a maximum value of 1.0. The reliability coefficient for 3 replicate readings is 0.89 in the negative range and 0.87 in the dysplasia range. For 1 reading, the reliability coefficient is 0.74 in the negative range and 0.69 in the dysplasia range. The high reliability coefficients show that the reproducibility obtained in reading negative and dysplasia smears is such that it enables us to distinguish between the various categories in both the negative and dysplasia ranges.

The finding that categories close together on the scale are chosen more often in replicate reading of the same slide than in reading different slides is also consistent with the proposition that the present ordering of categories and scale spacing essentially reflects the proposed concept of a biological continuum of abnormality. If categories far apart...
Table 3
Reproducibility analysis of papanicolaou readings by analysis of variance

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares</th>
<th>Degrees of freedom</th>
<th>Mean square</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative range (0-27)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between slides</td>
<td>34,451</td>
<td>195</td>
<td>176.67</td>
<td>9.44*</td>
</tr>
<tr>
<td>Within slides</td>
<td>7,334</td>
<td>392</td>
<td>18.71</td>
<td></td>
</tr>
<tr>
<td>Repeat readings S.D. = √18.71 = 4.3 units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability coefficient: 3 readings: 0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 reading: 0.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysplasia range (28-65)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Between slides</td>
<td>40,081</td>
<td>206</td>
<td>194.57</td>
<td>7.54*</td>
</tr>
<tr>
<td>Within slides</td>
<td>10,681</td>
<td>414</td>
<td>25.80</td>
<td></td>
</tr>
<tr>
<td>Repeat readings S.D. = √25.80 = 5.1 units</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability coefficient: 3 readings: 0.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 reading: 0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < 0.01.

Evidence in support of the ordering and spacing of the categories does not constitute proof of the biological relationships, but the scale does provide a quantitative measure for use in studying the natural progression in a defined population. We can test the hypotheses that the ordered changes in the negative range precede dysplasia and that dysplasia is a cancer precursor. In prospective studies on carcinogenesis in which latency is prolonged and incidence low, it may be necessary to measure progression not in terms of differential rates, but as changes in degree of severity, that is, changes in average score. In particular we can follow the course of changes in individual patients by plotting their numerical scores over time. In a process such as dysplasia, in which there is a high probability of reversal toward the negative range, it is important that these cases are not lost by being referred back into the general population of "normal" women. By use of the cytological scale, we can keep track of the scores of such patients, measuring the extent of reversal and watching for recurrence in terms of an increase in score.

Finally, the index of cytological progression can be used as a biological measure in estimating adverse effects, to be correlated with exposure to the suspected causal factor. We have, in fact, developed the scale for such a study of adverse effects of steroid contraceptives on cervical dysplasia.

ACKNOWLEDGMENTS

We acknowledge with thanks the cooperation of Dr. Carl F. Coffelt and the staff of the County of Los Angeles Community Health Services. We obtained assistance from the Health Sciences Computing Facility, UCLA, sponsored by an NIH Research Resources Grant RR-3.

REFERENCES

A Cytological Scale for Cervical Carcinogenesis

Elizabeth Stern, Alan B. Forsythe, Lee Youkeles, et al.

Cancer Res 1974;34:2358-2361.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/34/9/2358

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.