Cervical squamous intraepithelial lesions grades 1 and 2: value of an alternative morphological classification and of p16 immunohistochemistry in the prediction of clinical outcome

INTRODUCTION

Cervical cancer is a neoplasia with a high incidence worldwide, with 528,000 new cases/year, 85% of which occur in developing countries, being responsible for 266,000 deaths\(^1\). In Brazil, for the year 2016, 16,540 new cases were estimated, with a rate of 15.85/100,000 women\(^2\).

Multiple nomenclature and criteria are still used for diagnostic interpretation of preinvasive squamous lesions in cervical biopsies, contributing to the great intraobserver and interobserver variability. The lowest indices of diagnostic agreement are observed in the categories of cervical intraepithelial neoplasia grade 2 (CIN 2) and grade 1 (CIN 1). At a study using samples originally interpreted by pathologists from Costa Rica, two independent reviewing pathologists of the National Cancer Institute agreed with only 13% and 31% of the CIN 2 diagnoses, and only 24% and 20% of CIN 1 diagnoses. There was no agreement between them for any cases diagnosed as CIN 2\(^3\). Additionally, at a more recent North-American study, agreement of independent reviewing gynecologic pathologists with the initial diagnosis made by pathologists of the state of New Mexico was 38.2% for CIN 1 and 38% for CIN 2\(^4\). At a study using the diagnostic categories benign, low-grade squamous intraepithelial lesion (LSIL) (CIN 1), and high-grade squamous...
histopathological cervical specimens(6, 7), being recommended
inter- and intraobserver agreement in the evaluation of
cytoplasmic p16 overexpression(6).

In the transforming infections by human papillomavirus
(HPV) of high oncogenic risk, there is overexpression of p16,
which can be detected by immunohistochemistry. This happens
due to oncoprotein E7-mediated E2F displacement from its
binding site on pRb. Increase in free E2F levels stimulates nuclear
pressure on cells – one cannot rule out a high grade lesion (ASC-H) –, atypical
squamous cells of uncertain significance (ASC-US)/HPV-16+ or
atypical glandular cells (AGC) or not otherwise specified (NOS)
(4th context)(9).

The use of adjunct p16 immunohistochemistry increases
inter- and intraobserver agreement in the evaluation of
histopathological cervical specimens(6, 7), being recommended
by the Lower Anogenital Squamous Terminology (LAST)
Consensus – that officially introduced complementary methods
to the anatomopathologic exam in the diagnostic flowchart of
cervical tissue samples – in the following contexts: 1st) samples
in which there is doubt between precancerous lesion and benign
mimics; 2nd) cases in which the pathologist is considering (or
wants to confirm) the diagnostic interpretation of CIN 2(8), and 3rd)
diagnostic disagreement between two professionals analyzing
the specimen, and whose differential diagnosis includes precancerous
lesion. The use of p16 in samples of non-neoplastic cervical tissue
or presenting CIN 1 is contraindicated – except in situations of
high risk for misdiagnosis of high-grade lesion (defined
by previous colpocytologic exam as HSIL, atypical squamous
cells – one cannot rule out a high grade lesion (ASC-H) –, atypical
squamous cells of uncertain significance (ASC-US)/HPV-16+ or
atypical glandular cells (AGC) or not otherwise specified (NOS)
(4th context)(9).

The application of p16, despite its high negative predictive value
(NPV)(8, 10), is considered limited in predicting evolution or stratifying
the risk of patients with CIN 1, probably not altering management
protocols currently recommended(11, 19). By contrast, the use of p16 in
samples with diagnosis of CIN 2 is associated with downgrading –
that is from -IN 2 to LSIL, when p16 is negative — in up to around one
third of the cases and, therefore, potentially reduces the frequency of
unnecessary excisional procedures in patients with lesions of lower
biological risk(14, 15). However, it is important to remember that some
cases of HSIL can exhibit negative or doubtful results for p16, that
the marker is not also 100% specific, and interpretation is made
difficult in small or scanty samples, tangential histological sections
or unoriented tissue(10).

More recently, taking into account the limitations of
morphological analysis and also immunohistochemical
complementary exam with p16, another group of researchers
proposed a pragmatic subdivision of histopathological diagnoses
of cervical biopsies into LSIL, HSIL, and questionable grade
squamous intraepithelial lesion (QSIL), suggesting that the
patients with lesions classified as QSIL should initially undergo a
more rigorous follow-up(17).

Considering the persisting problems in the histopathological
diagnosis of CINs (especially CIN 2, a category of low diagnostic
reproducibility, variable/uncertain biology, which can generate
unnecessary interventions, costs and psychological
stress), our study aimed to assess the performance of different
systems of morphological classification and also of p16
immunohistochemistry in predicting an unfavorable outcome in
cases of CIN 1 and CIN 2.

**MATERIALS AND METHODS**

**Cases and clinical outcome**

This is a retrospective longitudinal study, which used initial
biopsies with histological diagnosis of CIN 1 or CIN 2 of patients
treated at Núcleo de Prevenção de Doenças Ginecológicas (NUPREV)
of Universidade Federal de São Paulo (UNIFESP) between January
2008 and March 2015. We evaluated sections obtained from paraffin
blocks with tissue microarrays (TMA) – previously prepared with
samples from NUPREV for use in studies about biomarkers in CINs –
and also sections obtained from conventional paraffin blocks: using
samples of TMA blocks previously constructed, 17 CIN 1 and 13 CIN
2 cases were included; using conventional paraffin blocks, 17 CIN 1
and 19 CIN 2 additional cases were included.

Cases were selected for this study based on suitability of
samples (that is, presence of the area of interest in the sections for
p16 immunohistochemical evaluation) and also on the presence
of clinical follow-up with anatomopathological examination. The
outcome was obtained through the active search of subsequent
diagnoses for all tested cases in the files of the Department of
Pathology of UNIFESP: “unfavorable clinical outcome” was
defined by subsequent CIN 3 or more severe histological diagnosis.
The following were excluded: patients with immunosuppression
of any type or intensity; women who were pregnant or breastfeeding
at diagnosis; patients clinically followed-up at another service and/or
lost to follow-up in this service; patients with favorable clinical
outcome, but with clinical follow-up shorter than six months;
patients with clinical outcome possibly unfavorable, but based
only on results of cytopathological tests with no histopathological
confirmation in this Department.
TMA construction

The technique was performed as described by Kononen et al. (1998)\(^{(18)}\), using a device of fixed base for withdrawal of 1-mm tissue cylinders. According to availability and size of material, one to three cylinders were obtained per paraffin block. From the new slides obtained after TMA confection, and when necessary, from the whole histological sections, we carried out a review of initial diagnosis by Richart’s classification with division of cases into CIN 1 and CIN 2, according to classical criteria previously established\(^{(19)}\), performed in consensus by two pathologists.

Morphological and immunohistochemical reclassification

According to sample availability, we conducted reclassification according to: a) alternative morphological evaluation as proposed by Herfs and Crum\(^{(17)}\), distributing cases into LSIL, QSIL and HSIL; and b) reclassification of CIN 2 cases with the use of p16 immunohistochemistry (see the following criteria), distributing the cases into LSIL and HSIL.

Immunohistochemical reactions and interpretation

Slides went through special immunohistochemical stains and polarization, received histological sections of 4 micrometers, obtained through a standardized conventional microtome, and underwent deparaffinization in incubator – temperature set at 57ºC for 60 min. After that, the automated protocol was applied according to the manufacturer’s guidelines – DAKO Autostainer 48 tm – with the steps described next: antigen retrieval in buffer solution PT Link (Dako), washing in buffer solution for 5 min and application of primary antibody (p16\(^{INK4a}\) clone G175-405, Zeta Corporation, in titers standardized by the service in 1/50); reaction amplification by the use of polymer EnVision Flex (Dako) for 20 min, washing in buffer solution for 5 min, color development of reaction in diaminobenzidine [(DAB); Sigma-Aldrich Chemical] Flex for 5 min and counterstaining with hematoxylin Flex for 10 min. Interpretation was made according to LAST\(^{(10)}\) recommendations into: negative – absence of immunoexpression or focal and/or discontinuous immunoexpression; and positive – diffuse block-positivity (nuclear, with or without associated cytoplasmic positivity) affecting at least the lower third of the epithelium (Figure). The cervical sample with CIN 3 was used as positive external control. The typical endocervical columnar epithelium was used as external negative control; and plasmocytes, used as positive internal control.

Statistical analysis

The categorical variables were described as graphs and tables. The only discrete quantitative variable (age in completed years) was described using mean and standard deviation. If applicable, Fisher’s exact test or its extension was conducted for evaluation of significance probability. Value of \(p < 0.05\) was considered statistically significant.

RESULTS

Sample description

When the inclusion and exclusion criteria were applied, 66 patients were included, of which 34 presented initial diagnoses of CIN 1, and 32, of CIN 2, with mean age of 32 ± 9 years. After a minimum follow-up period, we observed nine (13.6%) cases in which there was subsequent histological diagnosis of lesion of higher histological severity (CIN 3 or more severe): two occurred in patients with biopsies initially classified as CIN 1; the other seven, in patients with biopsies initially classified as CIN 2 (Table 1). We must highlight that seven of the nine patients with an unfavorable outcome had presented a concomitant cytology of HSIL; one, previous recent cytology of HSIL; and one, previous concomitant cytology of ASC-H. Besides, all later histopathology diagnoses of more severe histology [eight cases with CIN 3 and one case with squamous-cell carcinoma (SCC) stage IB1] occurred in the first six months of follow-up, indicating that the most severe lesions were probably already present, but were not sampled in the biopsies used in this study.
TABLE 1 – General data on the cases (n, age group, and outcome)*

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Age (mean ± SD)</th>
<th>Unfavorable outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN 1</td>
<td>34</td>
<td>32.46 ± 8.47</td>
<td>2</td>
</tr>
<tr>
<td>CIN 2</td>
<td>32</td>
<td>32.06 ± 8.82</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>32.27 ± 8.58</td>
<td>9</td>
</tr>
</tbody>
</table>

CIN 1: cervical intraepithelial neoplasia grade 1; CIN 2: cervical intraepithelial neoplasia grade 2; SD: standard deviation.

*See text of “Results” (in the section “Sample description”) for more details regarding cases that presented unfavorable outcome.

TABLE 2 – General frequency of p16 immunoexpression in Richart’s and the alternative classifications, and specific frequency of p16 immunoexpression (correlating both different classifications)

<table>
<thead>
<tr>
<th>General frequency of p16 immunoexpression</th>
<th>Positive p16</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Richart’s classification (n = 66)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIN 1</td>
<td>5/34</td>
<td></td>
</tr>
<tr>
<td>CIN 2</td>
<td>21/32</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26/66</td>
<td></td>
</tr>
<tr>
<td>Alternative classification (n = 62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSIL</td>
<td>4/33</td>
<td></td>
</tr>
<tr>
<td>QSIL</td>
<td>10/18</td>
<td></td>
</tr>
<tr>
<td>HSIL</td>
<td>9/11</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23/62</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specific frequency of p16 immunoexpression</th>
<th>LSIL</th>
<th>QSIL</th>
<th>HSIL</th>
<th>Total</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIN 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P16 Negative</td>
<td>28</td>
<td>90.3%</td>
<td>-</td>
<td>-</td>
<td>28</td>
</tr>
<tr>
<td>Positive</td>
<td>3</td>
<td>9.7%</td>
<td>2</td>
<td>100%</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>100%</td>
<td>2</td>
<td>100%</td>
<td>33</td>
</tr>
<tr>
<td>CIN 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P16 Negative</td>
<td>1</td>
<td>50%</td>
<td>8</td>
<td>50%</td>
<td>2</td>
</tr>
<tr>
<td>Positive</td>
<td>1</td>
<td>50%</td>
<td>8</td>
<td>50%</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>100%</td>
<td>16</td>
<td>100%</td>
<td>11</td>
</tr>
</tbody>
</table>

CIN 1: cervical intraepithelial neoplasia grade 1; CIN 2: cervical intraepithelial neoplasia grade 2; LSIL: low-grade squamous intraepithelial lesion; HSIL: high-grade squamous intraepithelial lesion; QSI: questionable grade squamous intraepithelial lesion.

* Fisher’s exact test or its extension.

P16 immunohistochemistry

Positivity for p16 was observed in 26 (39.4%) of the cases – five (14.7%) of the 34 CIN 1 and 21 (65.6%) of the 32 CIN 2. Comparison of the marker frequency distribution in the alternative classification demonstrated increasing positivity indices in the three proposed tiers. In the cases initially classified as CIN 1, positive p16 was more frequent in the patients reclassified as QSIL than in those reclassified as LSIL (p = 0.019) (Table 2). The p16 marker alone – even independently of any classification – obtained statistical significance in outcome prediction, besides expressive values of sensitivity and NPV for this determination (Table 3).

Classification according to LAST consensus, alternative morphological classification and correlation with outcome

After p16 immunohistochemistry, following recommendations of LAST Consensus, among 32 CIN 2 cases, 11 (34.4%) would have gone downgrading, that is, would be reclassified as LSIL (Table 2). Comparison of group outcomes according to Richart’s classifications and after conduction of adjunct p16 immunohistochemistry in CIN 2 cases (according to recommendations of the LAST Consensus) did not show any statistically significant difference. Conversely, according to criteria of the alternative system proposed by Herfs and Crum, we would reclassify 33 (53.2%) of 62 cases as LSIL, 18 (29%) as QSIL and 11 (17.8%) as HSIL. The comparison among these three last subgroups regarding outcome demonstrated a statistically significant difference (Table 4).

TABLE 3 – Correlation between p16 immunoexpression and outcome

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Favorable</th>
<th>Unfavorable</th>
<th>Total</th>
<th>p-value</th>
<th>S</th>
<th>Sp</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>p16 Negative</td>
<td>39</td>
<td>68.4%</td>
<td>1</td>
<td>11.1%</td>
<td>40</td>
<td>60.6%</td>
<td>0.002*</td>
<td>0.889</td>
</tr>
<tr>
<td>Positive</td>
<td>18</td>
<td>31.6%</td>
<td>8</td>
<td>88.9%</td>
<td>26</td>
<td>39.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>100%</td>
<td>9</td>
<td>100%</td>
<td>66</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

S: sensitivity; Sp: specificity; PPV: positive predictive value; NPV: negative predictive value.

* Fisher’s exact test or its extension.

TABLE 4 – Outcome of cases according to the different classifications

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Favorable</th>
<th>Unfavorable</th>
<th>Total</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richart’s classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIN 1</td>
<td>32</td>
<td>94.1%</td>
<td>2</td>
<td>5.9%</td>
</tr>
<tr>
<td>CIN 2</td>
<td>25</td>
<td>78.1%</td>
<td>7</td>
<td>21.9%</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>86.4%</td>
<td>9</td>
<td>13.6%</td>
</tr>
<tr>
<td>LAST Consensus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSIL</td>
<td>41</td>
<td>93.2%</td>
<td>3</td>
<td>6.8%</td>
</tr>
<tr>
<td>HSIL</td>
<td>16</td>
<td>72.7%</td>
<td>6</td>
<td>27.3%</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>86.4%</td>
<td>9</td>
<td>13.6%</td>
</tr>
<tr>
<td>Alternative classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSIL</td>
<td>32</td>
<td>97%</td>
<td>1</td>
<td>3%</td>
</tr>
<tr>
<td>QSIL</td>
<td>15</td>
<td>83.3%</td>
<td>3</td>
<td>16.7%</td>
</tr>
<tr>
<td>HSIL</td>
<td>6</td>
<td>54.5%</td>
<td>5</td>
<td>45.5%</td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>85.5%</td>
<td>9</td>
<td>14.5%</td>
</tr>
</tbody>
</table>

CIN 1: cervical intraepithelial neoplasia grade 1; CIN 2: cervical intraepithelial neoplasia grade 2; LAST: Lower Anogenital Squamous Terminology; LSIL: low-grade squamous intraepithelial lesion; HSIL: high-grade squamous intraepithelial lesion; QSI: questionable grade squamous intraepithelial lesion.

* Fisher’s exact test or its extension.
DISCUSSION AND CONCLUSION

It is necessary to give a name to everything that is discovered in order to make communication and understanding easy. However, for this to happen, criteria that define that name must be widely known and also applied in a uniform and objective manner — a process that can take decades. Richart (1967) proposed a nomenclature that still remains in use in gynecology and pathology clinical practice in many centers, even after it has been the subject of criticism from its own author, despite its limitations and the publication of LAST, which introduced new terminology (revised) as gold standard for classification of non-invasive cervical squamous lesions. That is, CIN still plays an important role in everyday clinical communication due to its propagation and sedimentation in the scientific setting; many times the role and limitations — even more important — of biopsy as diagnostic test, endowed with sensitivity, specificity, predictive values, diagnostic reproducibility, among other characteristics, remaining in second plan. Once medical science always searches for better diagnostic tests, we consider the act of finding new forms of classifying familiar lesions necessary and healthy for progress.

In order to conduct this study, the fundamental classifications for comparison of any new proposal were included: the most widespread and the departure point of the study, Richart’s original (three grades: CIN 1, CIN 2, and CIN 3), and the currently adopted by the World Health Organization (WHO) — that of the LAST Consensus (two grades: LSIL and HSIL). We also tried to evaluate an alternative system that permits a more practical approach: the model proposed by Herfs and Crum, of pure morphological character, without the stringency of the criteria proposed by Richart and admitting the existence of a subgroup of lesions in which one cannot determine the grade, the QSIL (besides LSIL and HSIL).

Reclassifications impacted only the group of cases more subject to diagnostic variability. The cases belonging to group CIN 1, in most situations, remained in analogous categories or categories of equivalent meaning; the same cannot be affirmed for group CIN 2: a significant proportion of the cases would be reclassified in categories of potentially divergent management in the other two analyzed classification systems. For example, in our case analysis, only 65% of the CIN 2 would be classified as HSIL according to LAST recommendations, and 40% would be kept in the same category if we applied the alternative system proposed by Herfs and Crum.

The -IN terminology, regardless of site, is prone to subjectivity of criteria application by many diverse reasons, such as intra- and interobserver variability, limitations intrinsic to the material received for analysis — the latter with potential selection bias of this and several studies on the theme — and also the basic question of these diagnostic terms being applied to lesions with variable morphologic spectra — and possibly different clinical behaviours. The presented data confirm this already evidenced limitation of Richart’s classification.

Although the present study has not demonstrated statistically significant difference between groups determined by LAST Consensus and outcome, we can point there is a non-confirmed tendency possibly because of lack of statistical power ($p = 0.051$).

The search for information able to determine the presence of a disease or risk of a disease to progress/regress has been the recent aim of numerous studies in the varied fields of medicine. We search what we could call ideal biomarker: cheap, reproducible, with high predictive values, high sensitivity and specificity. In the present study, with the use of a biomarker in an attempt to predict outcome in cases of non-invasive squamous cervical lesions, we obtained results that contribute with scientific literature about the theme.

p16 is the most studied marker in the context of non-invasive HPV-related squamous lesions and the only currently officially indicated for routine use as an ancillary method in evaluation of these lesions. The frequencies observed for p16 immunostaining agree with descriptions available in the scientific literature. Data from medical literature widely demonstrate the value of dichotomic classification of non-invasive cervical squamous lesions attached to the routine application of p16.

This work corroborates the previously described information, demonstrating the power of p16 of eventually predicting in which cases there is higher risk of association with diagnosis of more severe, concomitant or future lesion, although with low specificity. It was also possible to note the high NPV of this marker, as also already demonstrated in the literature.

Due to the small dimensions of residual samples in paraffin blocks after obtainment of sections for p16 immunohistochemistry and also to non-availability of liquid based cytology, it was not possible to perform additional methods such as, for instance, HPV genotyping. However, we must highlight that although the cases of this work did not present better detailing regarding HPV types, the authors believe this case study is the closest to everyday situation in which, frequently, there are no detailed clinical reports, HPV test results, or even previous or concomitant cytological slides for review.

The introduction of a terminology that includes the term “questionable” in one of its categories can meet great resistance from pathologists and gynecologists: QSIL does
not present any definitive criterion for the assisting physician to use it, just uncertainty. This allows such terminology to potentially become a subterfuge of technical quality and also a confounding factor for the gynecologist that will provide the definitive treatment. In spite of the above, in this study, the alternative system was the only that obtained statistical significance ($p = 0.002$), including the highest rate of cases classified as HSIL displaying unfavorable outcome (45.5%). Based on the results of this study, some considerations are relevant: the lines of the used alternative system have been just recently described and not well studied, requiring further studies with more cases to determine their actual applicability; the absence of rigid criteria that define QSIL, despite open to criticism, could on the other hand allow a more complete or comprehensive evaluation of the sample, which should include correlation with results of other inter-related exams (12); the system is of simple application, and at an initial evaluation, without ethical obstacles for the proposition of its author that cases classified as QSIL be reassessed in six months, taking into consideration the potentially benign natural history of CIN 2 (23).

**References**


