Chapter 6

p16\textsuperscript{INK4a} immunostaining as an alternative to histology review for reliable grading of cervical intraepithelial lesions

M.G. Dijkstra
D.A.M. Heideman
S. C. de Roy
L. Rozendaal
J. Berkhof
K. van Krimpen
K. van Groningen
P.J.F. Snijders
C.J.L.M. Meijer
F.J. van Kemenade

Journal of Clinical Pathology 2010; 11: 972-977
ABSTRACT

Background:
Histomorphological grading of cervical intraepithelial neoplasia (CIN) is crucial for clinical management. CIN grading is however subjective and affected by substantial rates of discordance among pathologists, which may lead to overtreatment. To minimise this problem, histology review of CIN lesions by a consensus panel of pathologists is often used. Diffuse strong p16INK4a immunostaining has been proposed to aid the identification of true high-grade cervical lesions (i.e., CIN2/3).

Aim:
To assess the value of additional interpretation of p16INK4a immunostains for making a more reproducible diagnosis of CIN2/3 lesions.

Methods:
The authors used a series of 406 biopsies of cervical lesions, with known HPV status, stained for both H&E- and p16INK4a. First, in a randomly selected set of 49 biopsies, we examined the effect of additional interpretation of p16INK4a immunostained slides, on the agreement of CIN diagnosis among three pathologists. Second, the full series of samples was used to assess the accuracy of p16INK4a-supported lesion grading by a single pathologist, by evaluating the degree of diagnostic agreement with the consensus diagnosis of expert pathologists based on H&E-stained sections only.

Results:
The study shows that the inter-observer agreement between three pathologists for the routine H&E-based diagnosis, ranged from fair (weighted kappa 0.44 (95% CI: 0.19-0.64)) to moderate (weighted kappa 0.66 (95% CI: 0.47-0.79)). The concordance increased substantially for p16INK4a-supported grading (mean weighted kappa 0.80 (95% CI: 0.66-0.89)). Furthermore, an almost perfect agreement was found between the p16INK4a-supported diagnosis of a single pathologist and the consensus diagnosis of an expert pathology panel (kappa 0.88 (95% CI: 0.85-0.89)).

Conclusions:
These data demonstrate that additive use of p16INK4a immunohistochemistry significantly improves the accuracy of grading CIN lesions by a single pathologist, equalling an expert consensus diagnosis. Hence, we advocate the combined use of p16INK4a-stained slides and conventional H&E-sections in routine histopathology to improve accuracy of diagnosis.
INTRODUCTION

Cervical cancer is caused by a persistent infection with high-risk human papillomavirus (hrHPV) types. Cervical cancer develops via premalignant precursor lesions, referred to as cervical intraepithelial neoplasia (CIN), of which CIN3 is the most advanced precursor lesion. Accurate grading of CIN lesions is important for clinical management of patients, because CIN1 and CIN2/3 lesions are differently treated. Also, the outcome of cervical screening trials is dependent on accurate CIN lesion grading. However, histomorphological diagnosis of CIN is complicated by a variety of cellular changes associated with inflammation, pregnancy, and/or atrophy. These changes may mimic precancerous cervical lesions, thereby making cervical histology, i.e., the diagnostic interpretation of hematoxyline-eosine (H&E)-stained slides, subjective and prone to variability. This is reflected in poor inter-observer agreement between pathologists. In particular, the differential diagnosis between immature squamous metaplasia and CIN1/2, or between low-grade (CIN1) and high-grade (CIN2/3) lesions may be difficult. To overcome these problems, difficult lesions are usually adjudicated by more than one pathologist, and in case of clinical trials the diagnosis of CIN lesions is often reviewed by expert pathologists. Collectively, this emphasises the need for specific biomarkers to aid objective CIN lesion grading, and to identify true high-grade dysplasia of the cervix.

A promising candidate marker to identify high-grade CIN lesions is the cellular protein p16INK4a, as it is over-expressed in cervical cells transformed in response to the expression of the hrHPV E7 oncoprotein. Indeed, several studies have demonstrated that cellular p16INK4a immunoreactivity increases with CIN grade. Moreover, prospective follow-up studies suggest that diffuse p16INK4a positivity might aid the identification of dysplastic lesions at risk for neoplastic progression. Actually, it was demonstrated that, unlike p16INK4a-negative lesions, 44% of women with a p16INK4a-positive lesion which was classified as “not CIN 2/3” by consensus pathology, were diagnosed with CIN 2/3 at follow-up.

In this study, we investigated whether the conjunctive interpretation of p16INK4a immunostains and conventional H&E-stained slides, so-called p16INK4a-supported grading, could increase the inter-observer agreement of CIN diagnosis by histopathologists compared to diagnosis made on H&E-stained sections alone. Moreover, we investigated whether this p16INK4a-supported diagnosis may serve as a proxy for the consensus diagnosis of expert pathologists, based on sole H&E-stained sections.
MATERIALS AND METHODS

Study population
Formalin-fixed paraffin-embedded (FFPE)-samples of cervical lesions, collected in the period 1999 to 2008 and of which sufficient material was left for further analysis, were selected from the files of three Pathology Departments (VU University Medical Center Amsterdam, Kennemer Gasthuis Haarlem and Spaarne Ziekenhuis Hoofddorp, the Netherlands). The series consists of 406 biopsies, of which 338 samples were originally diagnosed (further referred to as original lab diagnosis) as CIN2/3 lesions (i.e., 118 CIN2 and 220 CIN3) and 68 samples as ≤CIN1 lesions (i.e., 2 normal cervical epithelium, 65 CIN1 and 1 squamous metaplasia (squM)). One section of each sample was H&E stained, and 1 adjacent section was used for monoclonal p16INK4a immunohistochemistry (IHC) as described below.

Ethical approval was waived since the material for this study was anonymized according to the regulation in the Netherlands 17.

Immunohistochemistry (IHC) for p16INK4a
FFPE-sections (4µm) were deparaffinised and stained with primary mouse antibody (p16INK4a Ab-4, Clone 16P04, also known as JC2, (Lab Vision Corporation, Neomarkers, Fremont, California, USA) in the automated IHC staining system Bond-max (Leica Microsystems GmbH, Wetzlar, Germany). Antigen retrieval was performed with epitope retrieval 2 (eR2) according to standard procedures. Negative controls were similarly processed except for omission of the primary antibody. Sections from naevus biopsies were used as positive controls.

p16INK4a immunoexpression in epithelial cells was evaluated according to combined criteria previously set forth by both Klaes et al. 11 and Shi et al. 18 Staining of either the cell cytoplasm or nucleus, or both, was counted as a positive result. Distribution of staining was scored as: (0) negative (i.e., < 5% of cells positive); (1) focal staining: either focal scattered positive cells, or small cell clusters (examples shown figure 1D and 1E); (2) basal staining (i.e., low intense, diffuse staining restricted to the lower 1/3 of the epithelium) (example shown in figure 1C); (3) diffuse p16INK4a positivity, continuous p16INK4a staining of more than 1/3 of the epithelium (example shown in figure 1B), and (4) diffuse full thickness staining (i.e., positive cells involve the full thickness of the epithelium; example shown in figure 1A). In case some samples showed more than one pattern of p16INK4a staining, than scoring was based on the highest category. Based on a pilot study (data not shown) and literature data 7, 10, 19-22, a dichotomous classification of p16INK4a expression was used to distinguish high-grade (CIN2/3) from low-grade lesions (CIN0/1). Diffuse p16INK4a immunopositivity of more than one third up to the full thickness of the epithelium (scores 3 and 4; fig. 1A and 1B) was considered to support CIN2/3 diagnosis, and negative, focal, or diffuse, low intense basal staining (scores 0, 1 and 2; fig 1C to 1F), was indicative for ≤CIN1 diagnosis.
Figure 1. p16<sub>INK4a</sub> immunostaining patterns in cervical biopsies

Representative examples of p16<sub>INK4a</sub> immunostaining patterns are shown. A) Diffuse full thickness immunopositivity, B) Diffuse staining of more than 1/3 of the epithelium, C) Basal staining (diffuse, low intense staining restricted to lower 1/3 of the epithelium) D) and E) Focal staining, and F) Immunonegative sample.

Diffuse p16<sub>INK4a</sub> positivity of more than one third of the epithelium (figures A and B), was considered to support CIN2/3 diagnosis, and staining restricted to the lower 1/3 of the epithelium or no staining (figures C, D, E and F), was indicative of ≤CIN1 diagnosis.
Study design

A random selection of 49 samples (40 CIN2/3 and 9 ≤CIN1, by original lab diagnosis) was used to evaluate the value of additional interpretation of p16\(^{INK4a}\)-stained slides, to improve the inter-observer agreement between pathologists. Three pathologists (FvK, KvG, and CvK), from three different laboratories, with experience in cervical pathology were invited to join this study. First, each pathologist independently reviewed every case and rendered a diagnosis using H&E sections only. Subsequently, at least one month later, a second diagnosis was made based on a separate review using both H&E- and p16\(^{INK4a}\)-stained sections (p16\(^{INK4a}\) -supported diagnosis). Prior to the evaluation of the p16\(^{INK4a}\) immunostained slides, all investigators received photographed examples of each category of p16\(^{INK4a}\) immunoexpression, and instructions on how to perform categorization. The pathologists were instructed to use the interpretation of p16\(^{INK4a}\) staining patterns as additional, complementary information to either confirm or revise the preliminary diagnosis established on the H&E sections. To distinguish high-grade (CIN2/3) from low-grade lesions (CIN0/1), they were advised to use the dichotomous classification described above. The data were used to compare the agreement in diagnosis between pathologists using sole H&E-stained sections to the agreement after additional interpretation of p16\(^{INK4a}\) immunostains.

The total cohort of 406 biopsy samples was used to study the accuracy of the p16\(^{INK4a}\) -supported diagnosis of a single pathologist, by comparison to the ‘gold standard’, i.e., a consensus diagnosis. For the consensus diagnosis, an expert pathologist (FvK) reviewed all 406 samples by evaluation of H&E sections, blinded to the original lab diagnosis and HPV status. Lesions with altered grading, compared to the original laboratory diagnosis, were adjudicated with a second expert pathologist (LR); i.e., 2 out of 3 equals consensus (majority diagnosis). The histological diagnosis of CIN1, CIN2, and CIN3 was made according to the criteria as described in the AFIP atlas of gynaecological tumors 23. HrHPV DNA test results were used to evaluate this expert consensus diagnosis.

To establish a p16\(^{INK4a}\) -supported diagnosis for all 406 samples, one of the expert pathologists (FvK) reassessed the H&E-stained sections in conjunction with matched p16\(^{INK4a}\)-stained slides. For this review the cases were renumbered, and the pathologist was blinded to the original diagnosis, to his previous H&E-based diagnosis and to the hrHPV status.

HPV detection

Detection of hrHPV on DNA extracts from FFPE-sections of cervical biopsies was performed by GP S+/6+- PCR enzyme immunoassay (PCR-EIA), using a cocktail of 14 hrHPV types (i.e., HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66 and 68) essentially as described before 22. Additional type-specific hrHPV E7 PCR, was applied to GP S+/6+- PCR-EIA-negative biopsies to exclude false-negativity due to potential viral integration 24. Beta-globin PCR was performed on each DNA extract as a quality control.
Data and statistical analysis

Kappa statistics were used to assess the inter-observer agreement on CIN diagnosis between pathologists. Weighted kappa values take into account that some disagreements are more serious than others. In a weighted analysis a disagreement between diagnoses of negative and CIN3 is computed differently than disagreement between diagnoses of negative and CIN1. Quadratic weighted kappa was calculated within each pair of observers using 4 diagnostic categories: CIN0, CIN1, CIN2, and CIN3. Unweighted kappa values were calculated for dichotomized categories, based on whether surgical treatment was indicated (cut-off CIN2). McNemar testing was used to compare the concordance between observers, for CIN diagnosis based on H&E-stained slides, with the concordance based on conjunctive interpretation of p16INK4a-stained slides.

The accuracy of the p16INK4a-supported diagnosis of a single pathologist was assessed by calculating the concordance with the expert consensus diagnosis (i.e. the gold standard), using kappa statistics.

All statistical analyses were performed using SPSS11.5-software. CIs were calculated. P-values of 0.05 or less were considered statistically significant.

RESULTS

p16INK4a-supported grading increases inter-observer agreement on diagnosis of CIN lesions

CIN lesion grading, based on 49 H&E-stained cervical slides, showed a substantial amount of inter-observer variation between 3 pathologists. Most pronounced disagreement among observers was between CIN1 and CIN2 lesions, and between CIN2 and CIN3 lesions. The additional interpretation of p16INK4a-stained slides reduced the inter-observer variation, and improved the rate of agreement in lesion grading among pathologists substantially (Table 1).

The agreement on CIN diagnosis between pathologists based on H&E sections only ranged from fair (weighted kappa 0.44) to moderate agreement (weighted kappa 0.66), according to standards of Landis et al. 25; with a group (mean) weighted kappa value of 0.54 (95% CI: 0.38-0.69). While, the addition of a p16INK4a-stained slide for grading, improved the concordance between all observers considerably, with kappa values ranging from 0.79 to 0.82 (weighted group kappa 0.80 (95% CI: 0.66-0.89)).

Of interest, additional evaluation of p16INK4a expression patterns reduced the number of cases in which diagnosis of CIN2/3 was made for all pathologists (Table 2); 36 high-grade CIN cases were downgraded to ≤CIN1, i.e., 15 cases of atypical squamous metaplasia and 21 cases of CIN1. When inter-observer agreement was evaluated with respect to clinically relevant categories (i.e., ≤CIN1 and CIN2/3), a significantly increased agreement for 2 out of 3 pairs of pathologists, i.e., pathologist 1 versus pathologist 2 (McNemar, p=0.01), and pathologist 2 versus pathologist 3.
(McNemar, p=0.02) was found. A significant increase in group kappa value from 0.44 (95% CI: 0.27-0.60) for the H&E-based diagnosis, to 0.76 (95% CI: 0.64-0.84) for the p16INK4a-supported diagnosis was observed.

Table 1. Kappa values for agreement between pairs of pathologists before and after additional interpretation of p16INK4a-stained sections

<table>
<thead>
<tr>
<th>PA 1 versus PA 2</th>
<th>H&amp;E-based diagnosis</th>
<th>95% CI</th>
<th>p16-supported diagnosis</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.44</td>
<td>0.19-0.64</td>
<td>0.82</td>
<td>0.52-0.92</td>
<td></td>
</tr>
<tr>
<td>0.66</td>
<td>0.47-0.79</td>
<td>0.80</td>
<td>0.67-0.88</td>
<td></td>
</tr>
<tr>
<td>0.53</td>
<td>0.29-0.70</td>
<td>0.79</td>
<td>0.54-0.91</td>
<td></td>
</tr>
</tbody>
</table>

Group (mean) kappa

<table>
<thead>
<tr>
<th>PA 1 versus PA 2</th>
<th>H&amp;E-based diagnosis</th>
<th>95% CI</th>
<th>p16-supported diagnosis</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.54</td>
<td>0.38-0.69</td>
<td>0.80</td>
<td>0.66-0.89</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Frequency of CIN 2/3 diagnosis before and after additional use of p16INK4a stains

<table>
<thead>
<tr>
<th>H&amp;E-based CIN 2/3 diagnosis</th>
<th>p16-supported CIN 2/3 diagnosis</th>
<th>CIN 2/3 cases downgraded</th>
<th>Cases upgraded to CIN 2/3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA 1</td>
<td>30</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>PA 2</td>
<td>30</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>PA 3</td>
<td>36</td>
<td>21</td>
<td>15</td>
</tr>
</tbody>
</table>

p16INK4a-supported diagnosis as an alternative to expert consensus review

The accuracy of the p16INK4a-supported diagnosis of a single pathologist was assessed by evaluating the degree of diagnostic agreement with the (expert) consensus diagnosis. A consensus diagnosis for all 406 lesions was established by review of the original lab diagnosis by an expert pathologist, as described above. This resulted in 70 specimens (21%) that were reclassified in relation to the original lab diagnosis, of which 29 samples were upgraded and 41 samples were downgraded; including 8 samples that were re-diagnosed from ≤CIN1 to CIN2/3 lesions and 11 samples vice versa. The hrHPV DNA status of the samples was used to evaluate the consensus diagnosis and, as shown in Figure 2, the test-results seem to support the consensus diagnosis.

For comparison, the results of dichotomized p16INK4a-supported diagnosis of a single pathologist for all 406 lesions are shown in Table 3.
Figure 2. hrHPV status and p16INK4a expression in relation to the consensus diagnosis.

Results of both high-risk HPV DNA testing and immunohistochemical analysis of the p16INK4a protein expression (dichotomized into 'positive' (scores 3/4) and 'negative' (scores 0/1/2), on 406 biopsy samples in relation to the histology grade according to the consensus diagnosis.

The resulting kappa value for agreement with the consensus diagnosis is 0.88 (95% CI: 0.85-0.89), representing a very high concordance between these two diagnoses. As such, the p16INK4a-based diagnosis of a single pathologist seems a good alternative to expert histology review. Diffuse p16INK4a immunopositivity was better associated with (consensus) high-grade CIN (kappa 0.88 (95% CI 0.85-0.89)) than was the presence of high-risk HPV (kappa 0.69 (95% CI: 0.63-0.75)). Furthermore, addition of hrHPV DNA test results to p16INK4a-based dichotomous lesion grading (i.e., positive in both p16INK4a IHC and PCR-EIA assays equals high-grade, and negative in one or both assays equals low-grade) did not further improve the degree of agreement with the consensus diagnosis (kappa of 0.84 (95% CI: 0.78-0.91)) in our series.

Table 3. Concordance between the p16INK4a-supported diagnosis of a single pathologist and the consensus diagnosis of an expert pathology panel

<table>
<thead>
<tr>
<th>Consensus diagnosis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CIN0/1</td>
</tr>
<tr>
<td>p16-supported diagnosis</td>
<td>67</td>
</tr>
<tr>
<td>Total</td>
<td>71</td>
</tr>
</tbody>
</table>

DISCUSSION

This study shows that the inter-observer agreement in grading of cervical lesions between 3 pathologists, from 3 different laboratories, based solely on H&E-stained slides ranged from fair (weighted kappa 0.44 (95% CI: 0.19-0.64)) to moderate (weighted kappa 0.66 (95% CI: 0.47-0.79)). In particular the distinction between CIN1 and CIN2 lesions was subject of discussion. This is important, since clinical management of women with CIN1 (i.e., surveillance), and CIN2/3 lesions (i.e., surgical treatment) is different.
The additional interpretation of p16\textsuperscript{INK4a} immunostains resulted in a significant improvement in the overall concordance of CIN lesion grading, which will result in higher uniformity of patient management. These results are in agreement with data from two other studies that demonstrated an improved diagnostic accuracy for diagnosing CIN lesions with the use of p16\textsuperscript{INK4a} IHC\textsuperscript{26,27}. Furthermore, the combined interpretation of H&E- and p16\textsuperscript{INK4a} stains reduced the number of high-grade CIN diagnosis. The interpretation of consecutive p16\textsuperscript{INK4a} immunostains, apparently, helps to reassure pathologists in grading aggressive-appearing low-grade lesions as truly low-grade, and avoids upgrading of these lesions to be quite on the safe side. In addition, in other studies\textsuperscript{27,28} the additional use of p16\textsuperscript{INK4a} immunohistochemistry was reported to reduce the number of missed high-grade cases which also increases diagnostic accuracy. Moreover, the pathologists reported that, with the aid of an additional p16\textsuperscript{INK4a}-stained section, lesion grading was much easier and faster. These results underline the clinical implications of addition of p16\textsuperscript{INK4a} IHC for CIN diagnosis, in terms of efficiency and to reduce overtreatment; the additional use of p16\textsuperscript{INK4a} IHC provides greater accuracy of CIN grading with less variability, and thus could help to avoid unnecessary diagnostic and surgical procedures associated with pregnancy related morbidity and psychological distress. Although a cost-effectiveness analysis has not been performed, our first impression is that cost reduction owing to less surgical procedures would outweigh the costs associated with the implementation of the use of p16\textsuperscript{INK4a} staining into the standard diagnostic procedure. Our results furthermore showed an almost perfect agreement between the p16\textsuperscript{INK4a}-supported diagnosis of a single pathologist, and the consensus diagnosis of an expert pathology panel. Therefore, we concluded that the p16\textsuperscript{INK4a}-supported diagnosis was at least as accurate as this ‘gold standard diagnosis’, and may serve as a proxy for the consensus diagnosis.

In our series, diffuse p16\textsuperscript{INK4a}-expression was better associated with high-grade CIN (kappa 0.88 (95% CI: 0.85-0.89)), than was the presence of high-risk HPV DNA (kappa 0.69 (95% CI: 0.63-0.75)). As such, HPV testing had no further additive value to p16\textsuperscript{INK4a}-supported lesion grading.

With respect to the implementation of routine p16\textsuperscript{INK4a}-staining in gynaecopathology, we agree with Tsoumpou et al.\textsuperscript{29}, that for the interpretation of p16\textsuperscript{INK4a} immunoreactivity and the reliable use of p16\textsuperscript{INK4a} IHC for cervical lesion grading, standardized protocols including the use of validated antibodies, and a standardized scoring system with photographed examples of the different categories, are important. Such a standardized, quality-controlled reagent set and immunostaining protocol has already been validated for use in cervical cytology\textsuperscript{30}. For histopathology, our findings support the view of Dray and colleagues\textsuperscript{31}, who propose that a diffuse, positive, parabasal staining pattern is suggestive of a transforming hrHPV infection and accompanied high-grade CIN lesion, whereas p16\textsuperscript{INK4a} immunoreactivity restricted to the lower part of the epithelium (one third), focal scattered staining, or absence of staining is indicative of ≤CIN1 diagnosis.

In our study, a few CIN lesions demonstrated notable findings, causing discrepancies between the consensus diagnosis and the p16\textsuperscript{INK4a}-supported diagnosis (Figure 2). First, 11 of 335 (3%) consensus high-grade CIN lesions (i.e., 9 CIN2, 2 CIN3) were considered truly negative for p16\textsuperscript{INK4a}. 

In our series, diffuse p16\textsuperscript{INK4a}-expression was better associated with high-grade CIN (kappa 0.88 (95% CI: 0.85-0.89)), than was the presence of high-risk HPV DNA (kappa 0.69 (95% CI: 0.63-0.75)). As such, HPV testing had no further additive value to p16\textsuperscript{INK4a}-supported lesion grading.
Six of these samples tested hrHPV positive; 2 samples were HPV16 positive, 1 sample HPV16+42 positive, 1 sample HPV31 positive, 1 sample HPV58 positive, and 1 sample HPV51+52+6 positive. These results may either indicate that a minority of high-grade CIN lesions could be missed if diffuse p16\textsuperscript{INK4a} -staining is used as leading criterion for identification, or it may suggest overclassification by consensus histopathology. Second, 4 out of 71 (6%) low-grade lesions showed diffuse p16\textsuperscript{INK4a} positivity, three of which also tested hrHPV positive. These lesions might have been under classified by consensus histopathology, or represent lesions with a progression risk to CIN2/3, as suggested by Negri and co-authors \cite{14}. Unfortunately, we do not have data on the clinical follow-up of these patients. The potential of p16\textsuperscript{INK4a} immunostaining as a prognostic marker will require further studies.

There were some limitations to the study design. Although statistically sufficient in number, the sample size of the subset study (n=49) was small, and despite random selection from the total cohort, contained a high proportion of samples in which the consensus diagnosis had been difficult to establish. This implicates that, for these samples, the inter-observer agreement on H&E-based CIN diagnosis, will probably also be low and therefore a bias in favour of the p16\textsuperscript{INK4a} -supported diagnosis might have occurred.

In summary, our data showed that the conjunctive interpretation of p16\textsuperscript{INK4a} immunostains, significantly improved the accuracy of interpreting and grading cervical lesions on biopsy samples. Moreover, the p16-supported diagnosis of a single pathologist was as accurate as the consensus diagnosis of an expert pathology panel, and therefore may be used as a proxy to the expert consensus diagnosis. Taking into account the speed of the diagnostic process, and the relative ease of cervical lesion grading with additional interpretation of p16\textsuperscript{INK4a} immunostains, we advocate the combined use of H&E-stained and p16\textsuperscript{INK4a}-stained sections in routine histopathology, to improve accuracy of diagnosis at an acceptable cost when used in large populations. Finally, based on data of our study, it seems that the p16\textsuperscript{INK4a} -supported diagnosis of a single pathologist might well be used as an alternative to histology review, increasing cost-effectiveness and saving time.

**TAKE-HOME MESSAGES**

- The combined interpretation of H&E- and p16\textsuperscript{INK4a} immunostains significantly improves the accuracy of interpreting and grading cervical lesions on biopsy samples.
- The accuracy of CIN lesion grading by a single pathologist with the additional use of p16\textsuperscript{INK4a} stains, is comparable to the consensus diagnosis of an expert pathology panel.
- p16\textsuperscript{INK4a} staining of CIN lesions should be implemented in routine daily practise.
REFERENCE LIST


