Interobserver variation in the diagnosis and grading of dyskaryosis in cervical smears: Specialist cytopathologists compared with non-specialists


Abstract

Aims—To compare the assessment of dyskaryosis in cervical smears made by specialist consultant cytopathologists and consultant general histopathologists.

Methods—One hundred and ten cervical smears were circulated to 10 observers from five district general hospital histopathology departments and five major departments of cytopathology. Their responses were analysed by five consultant general histopathologists and five consultant specialist cytopathologists. In 54 of the 110 cases, the histology of a corresponding cervical biopsy specimen was compared with the smear assessments.

Results—Specialist cytopathologists were more consistent than non-specialists when diagnosing and grading dyskaryosis. They chose the higher grades of dyskaryosis more frequently than the non-specialists.

The cytopathologists recommended referral for colposcopy more frequently, but if they asked for a repeat smear, they wanted it done within three months more frequently than the histopathologists. The specialists were more frequently in agreement with the biopsy grade of intra-epithelial neoplasia than the non-specialists, whose smear diagnoses tended to underestimate the severity of the histopathological abnormality.

Conclusions—This study has shown major differences between specialist and non-specialist cytopathologists in the diagnosis and grading of cervical smears and in the recommended management of patients with abnormal smears. These differences may result in uneven clinical management of women with smear abnormalities. It is therefore important to explore possible strategies for standardising the reporting of cervical smears, such as centralisation of screening services, accreditation in cytopathology for non-specialist consultants, and the value of participation in external quality assessment schemes.

(J Clin Pathol 1994;47:515–518)

The increase in the number of cervical smears performed in England and Wales as a result of the cervical screening programme has meant that more consultant pathologists who can undertake cytopathology are needed. The Royal College of Pathologists' has expressed the view that additional consultant posts are necessary, but felt that the creation of whole time posts in cytopathology was not the appropriate solution. It was proposed, instead, that newly created consultant posts should be half time cytopathology and half time histopathology. Noticeably, most advertisements for consultant histopathology posts in district general hospitals now specify that the successful candidate must take part in the cytopathology service.

In university departments, however, and in some of the larger district general hospitals there is a separate subdepartment of cytopathology run by one or more consultant cytopathologists. These departments usually have a much greater cytopathology workload than the average district general hospital laboratory, although this may be due to larger numbers of non-gynaecological specimens rather than cervical smears. They may also function as training centres for pathologists and cytoscreeners.

Thus a cervical smear deemed abnormal by cytoscreeners will be reported either by a specialist consultant cytopathologist or a general consultant histopathologist who undertakes not only cytopathology but also surgical pathology and postmortem examinations. We have designed a study which compares, for the first time, as far as we are aware, cervical smear reporting by these two groups of consultant pathologists.

Methods

One hundred and ten cervical smears were circulated, in batches of 10, to the panel of observers. The smears had been received by one of the participants (JPO'S), after having been screened at both a primary and at a more senior level before being passed on as requiring a consultant opinion. Because the slides included the output of a number of busy colposcopy clinics, there were more abnormal smears than would have been expected from a routine screening population.

The panel of observers comprised five consultant histopathologists and five specialist consultant cytopathologists. The consultant histopathologists worked in district general hospitals and carried out duties in surgical pathology and necropsy work as well as cytopathology. The cytopathologists' duties were largely or completely confined to
cytopathology. The participants were asked to decide on the presence and grade of dyskaryosis and to specify their recommendations for further management. No clinical details were supplied.

In 54 of the 110 cases a cervical biopsy specimen had been taken at the same time as the cervical smear. These biopsy specimens were examined, without knowledge of the cervical smear findings, by an independent consultant histopathologist with a special interest in gynaecological pathology (SMI). They were graded as showing normal squamous epithelium, hyperplastic squamous epithelium, cervical intraepithelial neoplasia (CIN) grades I, II, III, or invasive carcinoma. The biopsy findings were compared with the observers' opinions on the corresponding smears.

DATA ANALYSIS
The degree of agreement between observers was expressed by the κ type statistics. The simple κ statistic, introduced by Cohen,2 takes account of the fact that part of an observed agreement between observers is due to chance. kappa is the observed agreement, corrected for chance agreement, divided by the maximum possible agreement, also corrected for chance agreement. That is to say, it expresses numerically the excess of the observed agreement over the chance agreement. For the classifications where there were more than two choices open to the observer, such as the degree of dyskaryosis, the weighted κ3 was calculated, to allow for the fact that some disagreements are more serious than others, and to give greater numerical weight to such disagreements. A

weighted κ value of more than 0·5 was arbitrarily taken as representing reasonable interobserver agreement.

To compare the behaviour of the groups in their assessment of cervical smears and in their recommendations for management, the decisions of each group (five individuals) for each smear (110 smears) — 550 observations, for each of a number of parameters — were assembled. To analyse whether the observed differences were consistent across individual smears, each smear was given two scores: what proportion of the five histopathologists regarded it as showing a particular feature; and what proportion of the five cytopathologists did so. A paired r test and corresponding 95% confidence interval were used to assess the degree to which one group of observers reported the feature more than the other group. Estimated mean differences may differ slightly from the difference between the two proportions when some of the 550 observations are classed as inadequate and therefore excluded.

For the 54 smears for which there was an accompanying biopsy specimen, tables comparing the biopsy histology and the smear assessments for dyskaryosis were constructed, excluding readings classed as inadequate. Each patient was characterised by the proportion of cytopathologists who assigned a more severe grade than the biopsy specimen minus the proportion who assigned a less severe grade, representing the tendency of cytopathologists to over- rather than undercall in that case, and a corresponding difference of proportions for the histopathologists.

Results
To assess the degree of agreement among the specialist cytopathologists, a symmetrical agreement matrix (table 1) was formed in which each observation from each cytopathologist was compared with the corresponding observations made by the other four observers:

yielding 5 (observers) × 4 (comparisons) × 110 (smears) = 2200 paired comparisons in all. A similar table (table 2) was constructed for the general histopathologists. Weighted κ statistics derived from these tables show that specialist cytopathologists achieved considerably greater agreement than the non-specialists who showed little more agreement than would be expected by chance (table 3).

Similar matrices were constructed for management recommendations and are available from J P O'Sullivan. These also show a much greater degree of consistency among specialist

Table 1 Variation among cytopathologists for 110 smears classified as dyskaryosis

<table>
<thead>
<tr>
<th></th>
<th>Inadequate</th>
<th>Negative</th>
<th>Blline</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
<th>? Invasive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate</td>
<td>14</td>
<td>24</td>
<td>10</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>Negative</td>
<td>24</td>
<td>84</td>
<td>40</td>
<td>6</td>
<td>6</td>
<td>20</td>
<td>12</td>
<td>192</td>
</tr>
<tr>
<td>Blline</td>
<td>10</td>
<td>40</td>
<td>76</td>
<td>16</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>232</td>
</tr>
<tr>
<td>Mild</td>
<td>7</td>
<td>6</td>
<td>32</td>
<td>155</td>
<td>20</td>
<td>3</td>
<td>4</td>
<td>496</td>
</tr>
<tr>
<td>Moderate</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>18</td>
<td>7</td>
<td>524</td>
</tr>
<tr>
<td>Severe</td>
<td>4</td>
<td>20</td>
<td>4</td>
<td>20</td>
<td>91</td>
<td>290</td>
<td>67</td>
<td>496</td>
</tr>
<tr>
<td>? Invasive</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>67</td>
<td>104</td>
<td>196</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>192</td>
<td>232</td>
<td>496</td>
<td>524</td>
<td>496</td>
<td>196</td>
<td>2200</td>
</tr>
</tbody>
</table>

Table 2 Variation among histopathologists for 110 smears classified as dyskaryosis

<table>
<thead>
<tr>
<th></th>
<th>Inadequate</th>
<th>Negative</th>
<th>Blline</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
<th>? Invasive</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inadequate</td>
<td>20</td>
<td>37</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td>1</td>
<td>88</td>
</tr>
<tr>
<td>Negative</td>
<td>37</td>
<td>84</td>
<td>42</td>
<td>61</td>
<td>16</td>
<td>41</td>
<td>3</td>
<td>284</td>
</tr>
<tr>
<td>Blline</td>
<td>7</td>
<td>42</td>
<td>22</td>
<td>85</td>
<td>20</td>
<td>20</td>
<td>4</td>
<td>200</td>
</tr>
<tr>
<td>Mild</td>
<td>10</td>
<td>61</td>
<td>85</td>
<td>430</td>
<td>137</td>
<td>84</td>
<td>17</td>
<td>824</td>
</tr>
<tr>
<td>Moderate</td>
<td>3</td>
<td>16</td>
<td>20</td>
<td>137</td>
<td>60</td>
<td>26</td>
<td>16</td>
<td>304</td>
</tr>
<tr>
<td>Severe</td>
<td>10</td>
<td>41</td>
<td>20</td>
<td>84</td>
<td>52</td>
<td>168</td>
<td>29</td>
<td>404</td>
</tr>
<tr>
<td>? Invasive</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>17</td>
<td>16</td>
<td>29</td>
<td>26</td>
<td>96</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>284</td>
<td>200</td>
<td>824</td>
<td>304</td>
<td>404</td>
<td>96</td>
<td>2200</td>
</tr>
</tbody>
</table>

Table 3 Internal consistency of the groups

<table>
<thead>
<tr>
<th>Cytopathologists</th>
<th>Histopathologists</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion agreement</td>
</tr>
<tr>
<td></td>
<td>Observed</td>
</tr>
<tr>
<td>Grade of dyskaryosis</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>0·95</td>
</tr>
<tr>
<td></td>
<td>0·96</td>
</tr>
</tbody>
</table>
cytopathologists than non-specialists (table 3).

Table 4 examines the frequency with which the two groups of pathologists assigned the smears into the six diagnostic categories (normal, borderline, mild, moderate, severe and invasive). The non-specialists diagnosed normality and mild dyskaryosis significantly more frequently than did the specialist cytopathologists. The latter, however, had a significantly greater index of suspicion for invasive carcinoma.

When the management recommendations of the two groups were compared (table 5), the cytopathologists suggested referral for colposcopy significantly more often than the histopathologists. They also requested a repeat smear within three months—that is, early recall—significantly more often than the histopathologists.

Table 6 examines the degree of agreement with the biopsy findings achieved by the two groups of observers. Although the 54 cervical biopsy specimens were of different types (12 colposcopic biopsies, 38 loop excisions, and four knife conisations) they were not subdivided for comparison purposes. The specialist cytopathologists agreed with the biopsy findings substantially more often than the non-specialists. When they differed from the biopsy diagnosis the non-specialists underestimated the severity of the biopsy abnormality significantly more often than the specialists.

Discussion

Our results show that there were major differences in the interpretation of cervical smear abnormalities between the two groups of pathologists, which suggests that general histopathologists who practise cytopathology need further training in it. Specialist cytopathologists were more consistent among themselves than non-specialists when diagnosing dyskaryosis and more frequently agreed with the biopsy diagnosis. The non-specialists diagnosed mild dyskaryosis in preference to the other diagnostic categories and frequently underestimated the severity of the biopsy diagnosis. In line with their tendency to diagnose higher grades of abnormality the specialists advised colposcopy, other gynaecological referral, and "repeat smear within three months" more often than the non-specialists who were very inconsistent in their overall management recommendations. In terms both of internal consistency and concordance with the biopsy diagnosis, therefore, the non-specialist cytopathologists performed less well than the specialists.

The tendency of the non-specialists to diagnose mild dyskaryosis in preference to other diagnostic categories may well be a major cause of the disparity between mildly abnormal cervical cytology and histology noted by other investigators. In contrast, the specialist cytopathologists showed a preference for the higher grades of abnormality and, in particular, raised the possibility of invasive carcinoma far more frequently than did the histopathologists. The latter finding may simply reflect the underlying philosophy of screening which aims to detect smear abnormalities and to have them treated without delay; a suspicion of invasiveness in this context is likely to expedite an appointment for colposcopic evaluation and biopsy of the cervix. The non-specialist cytopathologists, who are primarily histopathologists, would be biased against raising a suspicion of invasive malignancy because such a diagnosis in biopsy material is frequently a precursor to major surgery.

Table 4  Degree of dyskaryosis recorded by cytopathologists and histopathologists

<table>
<thead>
<tr>
<th></th>
<th>Cytopathologists (534 readings)</th>
<th>Histopathologists (528 readings)</th>
<th>Difference</th>
<th>95% confidence interval</th>
<th>t test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>48 (9.0%)</td>
<td>71 (13.4%)</td>
<td>- 4.9%</td>
<td>- 7.5% to + 2.3%</td>
<td>- 3.80</td>
<td>0.0002</td>
</tr>
<tr>
<td>Borderline</td>
<td>58 (10.9%)</td>
<td>50 (9.5%)</td>
<td>+ 1.5%</td>
<td>- 1.7% to + 4.8%</td>
<td>+ 0.93</td>
<td>0.36</td>
</tr>
<tr>
<td>Mild</td>
<td>134 (23.2%)</td>
<td>206 (39.0%)</td>
<td>- 15.1%</td>
<td>- 20.1% to - 10.1%</td>
<td>- 9.55</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Moderate</td>
<td>131 (24.5%)</td>
<td>76 (14.4%)</td>
<td>+ 10.0%</td>
<td>+ 4.8% to + 15.1%</td>
<td>+ 3.85</td>
<td>0.0002</td>
</tr>
<tr>
<td>Severe</td>
<td>124 (23.2%)</td>
<td>101 (19.1%)</td>
<td>+ 3.9%</td>
<td>- 1.2% to + 9.0%</td>
<td>+ 1.51</td>
<td>0.13</td>
</tr>
<tr>
<td>Invasive</td>
<td>49 (9.2%)</td>
<td>24 (4.5%)</td>
<td>+ 4.6%</td>
<td>+ 1.5% to + 7.8%</td>
<td>+ 2.92</td>
<td>0.0042</td>
</tr>
<tr>
<td>(Severe + invasive)</td>
<td>173 (32.4%)</td>
<td>125 (23.7%)</td>
<td>+ 8.5%</td>
<td>+ 3.9% to + 13.2%</td>
<td>+ 3.65</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Table 5  Management recommendations made by cytopathologists and histopathologists

<table>
<thead>
<tr>
<th></th>
<th>Cytopathologists (570 readings)</th>
<th>Histopathologists (550 readings)</th>
<th>Difference</th>
<th>95% confidence interval</th>
<th>t test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refer colposcopy or other investigation</td>
<td>288 (52.4%)</td>
<td>246 (44.7%)</td>
<td>+ 7.6%</td>
<td>+ 3.0% to + 12.3%</td>
<td>+ 3.25</td>
<td>&lt; 0.0016</td>
</tr>
<tr>
<td>Early recall</td>
<td>112 (20.4%)</td>
<td>47 (8.5%)</td>
<td>+ 11.8%</td>
<td>+ 7.1% to + 16.6%</td>
<td>+ 4.93</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

Table 6  Comparison with the biopsy findings

<table>
<thead>
<tr>
<th></th>
<th>Cytopathologists (566 readings)</th>
<th>Histopathologists (567 readings)</th>
<th>Difference</th>
<th>95% confidence interval</th>
<th>t test</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above biopsy diagnosis</td>
<td>82 (30.8%)</td>
<td>61 (22.8%)</td>
<td>+ 21.0%</td>
<td>+ 15.1% to + 26.9%</td>
<td>+ 23.9%</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Complete agreement</td>
<td>107 (40.2%)</td>
<td>87 (32.6%)</td>
<td>+ 7.6%</td>
<td>+ 3.0% to + 11.2%</td>
<td>+ 3.25</td>
<td>&lt; 0.0016</td>
</tr>
<tr>
<td>Below biopsy diagnosis</td>
<td>119 (44.6%)</td>
<td>77 (28.9%)</td>
<td>- 21.7%</td>
<td>- 37.5% to - 5.9%</td>
<td>+ 13.5%</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Tendency to over- rather than undercall</td>
<td>119 (44.6%)</td>
<td>77 (28.9%)</td>
<td>- 21.7%</td>
<td>- 37.5% to - 5.9%</td>
<td>+ 13.5%</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>t test</td>
<td>0.025</td>
<td>0.0081</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p value</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
This study may be criticised for attempting
to correlate biopsy histology with smear
abnormalities. We would maintain that these
two different tests, although dependent on
different variables, should reflect the nature of
the underlying lesion and thus show a good
correlation. That the biopsy interpretation
was carried out by only one pathologist may
be regarded as inconsistent with the philoso-
phy of a study on interobserver variation.
However, the observer in question not only
had a special interest in cervical pathology
but, having conducted two previous studies
on observer variation in the reporting of cervi-
cal biopsy specimens, was aware of the
problems of subjectivity in this area.

The British Society of Clinical Cytology
has recognised that there is a widespread lack
of uniformity in cervical smear reporting and
has produced guidelines on grading dys-
karyosis as well as advice on management
recommendations. These guidelines may in
part be responsible for the consistency achieved
by the specialist cytopathologists. However, the performance of the non-specialists, who report a substantial proportion, if not the majority, of abnormal cervical smears in the United Kingdom, suggests that standardisation of cervical smear reporting is far
from optimal.

Centralisation of cervical smear reporting
in specialist laboratories is the single course of
action which might simultaneously improve the
quality of reporting and its consistency. However, this option would be expensive and
would require major reorganisation of existing
services. Furthermore, it would be impractical because it would result in a sudden
demand for specialist consultants in cyto-
pathology which would be out of keeping with
the number of trainees in the speciality. The
second possible option, favoured by six of the
10 panel members, that cytopathology should
be carried out by specialist cytopathologists
appointed to district general hospitals, would
be unworkable for similar reasons. In the cur-
rent economic climate, it might be more prac-
ticable and less disruptive to introduce an
accreditation scheme in cytopathology for
non-specialist consultants which could be
combined with the regional external quality
assessment schemes currently in operation.

1 Burston J, Hudson E, Williams ED. Consultant
histopathologists with responsibility for cytopathology.
2 Cohen J. A coefficient of agreement for nominal scales.
Educational and Psychological Measurement. 1960;20:
37-46.
3 Cohen J. Weighted kappa: nominal scale agreement
with provision for scaled disagreement or partial credit.
4 Soutter WP, Wisdom S, Brough AK, Monaghan JM.
Should patients with mild atypia in a cervical smear be
93:70-4.
5 Campion MJ, McCance DJ, Cuzick J, Singer A.
Progressive potential of mild cervical atypia. Prospective
cytological, colposcopic and virological study. Lancet
6 Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans
DMD, Gradwell E, et al. Observer variation in
histopathological diagnosis and grading of cervical
7 Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans
DMD, Gradwell E, et al. Reporting cervical intraepithelial
neoplasia (CIN): Intra- and interpathologist variation
and factors associated with disagreement. Histopathology
8 Evans DMD, Hudson EA, Brown CL, Boddington MM,
Hughes HE, MacKenzie EF, et al. Terminology in
gynaecological cytopathology: report of the working
council of the British Society for Clinical Cytology. J Clin
Pathol 1986;39:933-44.
9 Evans DMD, Hudson EA, Brown CL, Boddington MM,
Hughes HE, MacKenzie EF. Management of women
with abnormal cervical smears: supplement to terminol-
ogy in gynaecological cytopathology. J Clin Pathol
1987;40:530-1.