Diagnostic distance of high grade prostatic intraepithelial neoplasia from normal prostate and adenocarcinoma

Rodolfo Montironi, Roberto Pomante, Paola Colanzi, Deborah Thompson, Peter W Hamilton, Peter H Bartels

Abstract

Objective—To develop a distance measure based methodology to support the morphological evaluation of high grade prostatic intraepithelial neoplasia (PIN), a direct precursor of prostate cancer.

Methods—Eight morphological and cellular features were analysed in 20 cases of high grade PIN found in radical prostatectomy specimens from patients with adenocarcinoma. The diagnostic distance was evaluated to measure the extent to which the feature outcomes of the individual high grade PIN cases differed from the expected outcome profile of normal prostate, low and high grade PIN, and cribriform and large acinar adenocarcinoma. The belief value for high grade PIN was evaluated with a Bayesian belief network (BBN).

Results—Complete separation existed between the cumulative absolute diagnostic distances of these 20 cases from the prototype feature outcomes of high grade PIN and normal prostate the values for which were \( \leq 3 \) (range 0 to 3) and \( \geq 9 \) (range 9 to 15), respectively. The distances from low grade PIN (range 3 to 9), cribriform adenocarcinoma (range 2 to 8), and large acinar adenocarcinoma (range 5 to 10) were intermediate and showed overlap in their distribution. When taking into consideration whether the severity of feature changes was increasing or decreasing in comparison with the category prototype outcomes, the cumulative directional diagnostic distances from high grade PIN ranged from \(-3\) to \(+3\). Positive distance values were seen relative to low grade PIN (range \(+3\) to \(+9\)) and relative to normal prostate (range \(+9\) to \(+15\)). Negative values were found relative to cribriform adenocarcinoma (range \(-8\) to \(+2\)). The distance values from large acinar adenocarcinoma ranged from \(-2\) to \(+4\) and partly overlapped with those from the high grade PIN category. A bivariate scattergram derived from both diagnostic distance measures showed excellent separation between the groups’ distances. BBN analysis confirmed the morphology based diagnosis.

The distance evaluation resulted in 18 cases whose belief value for high grade PIN ranged from 0.60 to 0.87. In the remaining two cases the results of the BBN analysis showed a belief value of 0.50 and 0.57 for low grade PIN and of 0.49 and 0.38 for high grade PIN, respectively.

Conclusions—Distance measure based methodology represents a useful diagnostic decision support tool for the accurate evaluation of high grade PIN.

Keywords: prostatic intraepithelial neoplasia; diagnostic distance; rank difference; Bayesian belief network; diagnostic decision support systems

Dysplastic epithelial changes and putative pre-malignant lesions of the prostate were recognised more than 50 years ago. Recently, McNeal and Bostwick described diagnostic criteria for intraductal dysplasia, now called prostatic intraepithelial neoplasia (PIN). PIN is a spectrum of morphological and cellular intraglandular changes, initially graded on a scale of 1–3, but later divided into low grade (equivalent to PIN 1) and high grade (equivalent to PIN 2 and PIN 3). The histological features of PIN are well defined, but practising pathologists may encounter difficulties in applying these criteria in the routine practice of histopathology. In view of the clinical significance of high grade PIN as a strong predictor of adenocarcinoma, diagnostic accuracy and reproducibility are important considerations in patient management.

Descriptive terms and words are used to define the criteria involved in the diagnosis and grading of individual PIN lesions. For instance, when examining a case, the presence of “profound nuclear enlargement” points to high grade PIN, whereas “no nuclear enlargement” is indicative of normal prostate. The severity of feature changes is evaluated subjectively by the pathologist. To avoid dependence of the analysis on the observer, reference images for each feature and its outcomes should be consulted for comparison. This is usually done with Bayesian belief networks (BBN) with computer graphic schematic representations, or a ranked series of typical images may be displayed to allow a comparison.

However, the degree of nuclear enlargement in high grade PIN is comparable to that in invasive adenocarcinoma. Therefore, an individual case can be diagnosed as belonging to one of the categories only when a certain number of distinctive features has been evaluated—that is, a conclusion is based on all the feature changes that occur in a case and on
Table 1 Diagnostic features (or clues) and outcomes or severity of clue changes

<table>
<thead>
<tr>
<th>Gland pattern changes</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1: none</td>
<td>*(1)</td>
<td>*(2)</td>
<td>*(3)</td>
<td>*(3)</td>
</tr>
<tr>
<td>2: moderate</td>
<td>*(1)</td>
<td>*(3)</td>
<td>*(3)</td>
<td>*(3)</td>
</tr>
<tr>
<td>3: severe</td>
<td>*(1)</td>
<td>*(2)</td>
<td>*(3)</td>
<td>*(3)</td>
</tr>
</tbody>
</table>

Table 2 Sequences (profiles) of the clue outcomes typical for the five diagnostic categories

<table>
<thead>
<tr>
<th>Gland pattern changes</th>
<th>NP</th>
<th>PINlow</th>
<th>PINhigh</th>
<th>PACeri</th>
<th>PACacinar</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: none</td>
<td>*(1)</td>
<td>*(2)</td>
<td>*(3)</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>2: moderate</td>
<td></td>
<td></td>
<td>*</td>
<td>*(3)</td>
<td>*(3)</td>
</tr>
<tr>
<td>3: severe</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Material and methods
This study involved high grade PIN found in 20 radical prostatectomy specimens from patients with stage B prostatic adenocarcinoma. The age of the patients ranged from 60 to 70 years. None of the patients had received chemotherapy, hormonal or radiation treatment before surgery. The radical prostatectomy specimens were fixed for 24–48 hours in neutral buffered formalin (4%) and then cut into slices approximately 0.3 cm thick. Using haematoxylin and eosin stained sections, one of our team reviewed all the histological slides, controlled the quality of the material, and selected the slides for evaluation. PIN was selected in slides where prostatic adenocarcinoma was not present to avoid the risk of measuring the intraductal spreading of adenocarcinomatous cells.

The investigation also included 10 prostate biopsies with high grade PIN to show whether the diagnostic distance based methodology being developed in prostatectomy specimens was applicable to biopsy material.

DIAGNOSTIC FEATURES AND CATEGORIES
The following eight diagnostic features (or clues) were evaluated in haematoxylin and eosin stained sections:

1. gland pattern changes—this is related to the alteration in the normal prostate architecture as observed at low microscope magnification
2. gland cribriformity—acinar structures filled with epithelial cells forming multiple gland-like lumens
3. basal cell nuclear recognition—related to the degree of basal cell layer disruption
4. basal cell nuclear prominence
5. secretory cell cytoplasm appearance
6. secretory cell stratification
7. secretory cell cytoplasm appearance
8. secretory cell nuclear prominence.

These features and their outcomes, or severity of clue changes, were based on the description of the morphological spectrum of PIN given by Bostwick and are shown in Table 1. A ranked series of typical instances was available for comparison and reference. The features were analysed by the same observer in all the cases. The two features related to tissue architecture were evaluated at an objective magnification of 10:1. The remaining features were assessed under a 40:1 objective. For each case, two to three fields of view were evaluated.

Five cases selected at random from the 20 prostatectomy specimens were also evaluated by a second observer. Complete agreement with the first observer in the outcome assignment was observed with the following features: gland pattern changes, gland cribriformity, basal cell nuclei recognition, and secretory cell nuclear stratification. For the features basal cell nuclear prominence, secretory cell cytoplasm appearance, and secretory cell nuclear prominence, an identical outcome was assigned by both observers in four cases; in the fifth case, contiguous outcomes were selected. For the feature secretory cell nuclear enlargement, agreement was observed in three cases; in only one of the other two cases contiguous outcomes were assigned.

In addition, the following five diagnostic categories (or alternatives) were considered:

1. normal prostate (NP)
2. low grade prostatic intraepithelial neoplasia (PINlow)
3. high grade prostatic intraepithelial neoplasia (PINhigh)
4. prostatic adenocarcinoma with cribriform pattern (PACeri)
5. prostatic adenocarcinoma with large acinar pattern (PACacinar).

In NP the ducts and acini appear lined by a two cell-type epithelium—that is, the basal cell layer and the luminal (secretory) cell layer. PIN is characterised by the presence of cells...
that appear stratified and crowded. PIN of low grade shows cell features of mild or moderate dysplasia. PIN of high grade shows those of severe dysplasia or carcinoma; in particular, the cells display profoundly enlarged nuclei and large prominent nucleoli, similar to those seen in invasive adenocarcinoma. PACcri basically corresponds to Gleason primary grade 3C—that is, masses and cords of cribriform/papillar tumour with smooth rounded edges. Even though PACcri covers the primary grade spectrum from Gleason 1, 2, 3A to 3B, it basically corresponds to the grade 3A—that is, glands of moderate to large size. The expected profile of feature outcome for each diagnostic category was reported previously and is shown in table 2.

### DIAGNOSTIC DISTANCE MEASURES

Each outcome of a diagnostic clue can be represented by a single number. For example, in the diagnostic clue “gland pattern changes”, the outcomes can be none (1), moderate (2) or severe (3) (fig 1). Considered in this way, one can calculate the difference between the clue outcome of a given case from a defined prototype for a given diagnostic category. For example, if one observes gland pattern changes “moderate” (2) in a given high grade PIN case and the normal prostate prototype is represented by gland pattern changes “none” (1) then the difference is 1. This difference remains a positive value regardless of the direction of shift and has been termed the absolute rank difference. This value can be calculated over the full set of diagnostic clues and summed to give a cumulative absolute diagnostic distance.

A second distance measure was computed to account for the direction of change in clue values. This was termed the directional rank difference and in the above example would result in a value of –1. As before this can be calculated for each diagnostic clue and summed over the whole case to give a cumulative directional diagnostic distance. In this distance measure, the equal occurrence of positive and negative differences within a given string can result in a cumulative distance of 0. This needs to be
Figure 2. High grade PIN case 1. Cumulative absolute diagnostic distance from the five diagnostic categories. The smallest distance is from the high grade PIN category. The direction of feature outcome changes is not considered.

taken into consideration when using this distance measure.

BAYESIAN BELIEF NETWORK
A BBN was used to evaluate the level of certainty (or uncertainty) in high grade PIN diagnosis. A shallow network was used with eight first level descendant nodes for the diagnostic clues, each independently linked by a conditional probability matrix to a root node containing the five diagnostic alternatives. The features, their outcomes, and the categories are exactly the same as in the diagnostic distance evaluation and correspond to those previously used to develop the BBN applied in this study. A detailed description of the algorithmic mechanism of network initialisation and belief propagation in BBN has been given by Bartels et al., Bibbo et al., and Hamilton et al. and their theoretical application in quantitative histopathology has been extensively researched by Bartels et al. For each case the time needed to evaluate the features and to calculate the diagnostic distance and the belief value was approximately 25 minutes.

Results
The basic steps in the rank difference and diagnostic distance calculations are best shown with the following two examples.

EXAMPLE 1
High grade PIN case 1 shows the following clue outcomes: gland pattern changes, moderate (2); gland cribriformity, absent (1); basal cell nuclear recognition, recognisable (1); basal cell nuclear prominence, moderate (2); secretory cell nuclear enlargement, great (3); secretory cell stratification, moderate (3); secretory cell cytoplasm appearance, granular (3); and secretory cell nuclear prominence, some (2). Therefore, case 1 is represented by the following outcome sequence or profile: (2,1,2,3,3,2,3,2). As shown in table 2, the expected clue outcome profile of low grade PIN is (2,1,1,2,2,2,2,2). This sequence represents: gland pattern changes, moderate; gland cribriformity, absent; basal cell nuclear recognition, recognisable; basal cell nuclear prominence, moderate; secretory cell nuclear enlargement, moderate; secretory cell stratification, mild; secretory cell cytoplasm appearance, moderately clear; and secretory cell nuclear prominence, some. The comparison between the two sequences is reported in table 3 together with the rank differences and the diagnostic distances, calculated considering the direction of changes (fourth column).

EXAMPLE 2
The sequence of the clue outcomes of large acinar adenocarcinoma is (3,1,3,4,3,1,1,3). The comparison with the sequence of the high grade PIN case 1 is reported in table 4.

Therefore, the cumulative diagnostic distances of high grade PIN case 1 from low grade PIN (example 1) and large acinar adenocarcinoma (example 2) are 3 and 10, respectively. When the direction of severity of changes is considered the distances are +3 and −2, respectively. Figures 2 and 3 show the diagnostic distance accumulation graphically. The order of the features is arbitrary and kept constant so that intermediate results could be compared.

The clue outcome profiles for the 20 high grade PINs are shown in table 5. In nine of 20 cases the profile is identical and represented by the sequence (2,1,1,2,3,3,3,3). The remaining 11 deviate slightly in the outcome severity of some cell features. For instance, the secretory

<table>
<thead>
<tr>
<th>Case No</th>
<th>Gland changes</th>
<th>Gland cribriformity</th>
<th>Basal cell recognition</th>
<th>Basal cell prominence</th>
<th>Nuclear enlargement</th>
<th>Nuclear stratification</th>
<th>Cytoplasm appearance</th>
<th>Nuclear prominence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Diagnostic distance evaluation of prostatic intraepithelial neoplasia

Figure 3  Cumulative directional diagnostic distance in high grade PIN case 1. The direction of feature outcome changes is considered. Positive values are observed in the comparison with low grade PIN and normal prostate. Negative values are observed from high grade PIN, and from large acinar and cribriform adenocarcinoma.

Figure 4  When the direction of feature outcome change is not considered, the cumulative absolute diagnostic distances of the 20 cases of high grade PIN from the five diagnostic categories range from 0 to 15. A complete separation exists between the distances from high grade PIN and normal prostate.

cell nucleolar prominence is some (2) instead of frequent (3). Another feature that contributes to the deviation is the secretory cell stratification that is either mild (2) or pronounced (4) instead of moderate (3). These outcome variations from the most frequent profile derive from the non-homogeneity of high grade PIN morphology.

When the direction of feature outcome change is not considered, the absolute rank difference takes a value from 0 to 3. A rank difference of zero is frequently observed in the evaluation of the distance from the high grade PIN category. A rank difference of 1 is frequent from low grade PIN and cribriform adenocarcinoma and, to some extent, from large acinar adenocarcinoma. Rank differences of 2 and 3—the latter rarely present—are mainly from normal prostate and, to a minor degree, from large acinar adenocarcinoma.

The cumulative absolute diagnostic distances of the 20 cases of high grade PIN from the five diagnostic categories range from 0 to 15. A complete separation exists between the distances from high grade PIN and from normal prostate, the values for which are ≤ 3 (range 0 to 3; mode 0) and ≥ 9 (range 9 to 15; mode 12), respectively. The distances from low grade PIN (range 3 to 9; mode 6), cribriform adenocarcinoma (range 2 to 8; mode 5) and from large acinar adenocarcinoma (range 5 to 10; mode 7) occupy a position intermediate between those from high grade PIN and normal prostate and show distribution overlap (fig 4).

When the direction of feature outcome change is considered, the directional rank difference ranges from a negative value, through 0, to a positive value. A negative rank difference value is frequently observed in the evaluation of the distance from the cribriform and, to some extent, large acinar adenocarcinoma categories. A 0 rank difference is frequent from high grade PIN and, to a minor degree, from cribriform and large acinar adenocarcinoma. A positive rank difference is mainly observed from normal prostate and low grade PIN.

The cumulative directional diagnostic distances of the 20 cases of high grade PIN from the five diagnostic categories range from −8 to +15. Compared with the prototype clue outcomes of PINhigh, the diagnostic distances range from −3 to +3, most of them being 0 (mode 0). The highest positive values are in comparison with normal prostate (range +9 to +15; mode +12) and negative values are observed with cribriform adenocarcinoma (range −8 to +2; mode −5). Distance values intermediate between those from high grade PIN and normal prostate are observed in the comparison with low grade PIN (range +3 to +9; mode +6). The distance values from large acinar adenocarcinoma range from −2 to +4 (mode +1) and partly overlap with those from high grade PIN (fig 5).

Figure 6 shows the bivariate scatterplot of the diagnostic distances of the 20 cases of high grade PIN from the five diagnostic categories when considering both measures of rank differences. Separation of distances between the diagnostic categories is achieved.

The cumulative absolute and directional diagnostic distance values derived from the biopsy material fall within the category ranges observed in the prostatectomy specimens. However, there are minor differences in the category mode (data not shown) because the analysis of secretory cell nucleolar prominence and cytoplasm appearance is not always straightforward in biopsies as it is in the
PIN forms a morphological continuum from normal prostate to invasive adenocarcinoma. When a prostate specimen is analysed, the pathologist has to identify its exact location along this spectrum. Basically, this is done by comparing morphologically the case under investigation with known points along the spectrum that represent diagnostic categories or grades. The case is then allocated to the category to which it is closest or most similar. In practice, a pathologist looks through the microscope and the lesion is immediately diagnosed into one of the categories. Subsequently, features are mentioned that led to the conclusion. This is called synthetic or inductive diagnosis making, an important part of which is probably subconscious. This type of diagnosis making has the advantage of speed, but factors may play such a role that a classification shift in continuous lesions, including PIN, may result and a diagnosis given that is not strictly correct. Allam et al listed some of these factors in a recent paper on observer variability in the diagnosis of high grade PIN and adenocarcinoma, taking into consideration the level of interest in prostate pathology, the subjective application of diagnostic criteria, and the influence of peers and clinical colleagues.

The alternative in diagnosis making is the so-called analytical process. The pathologist evaluates the microscope image(s) of an individual case to derive a simplified image that is represented by characteristic features and outcomes. In the next step, the simplified image is compared with typical images in memory to make a diagnosis. The analytical process is far better than the synthetic one because the pathologist has to evaluate a series of features and their degree of changes, and to compare them with prototypes present in the pathologist’s memory. In the current study we followed the analytical approach. The classic works by Bostwick and coworkers were very helpful to identify the features and their changes that have to be analysed each time prostate material is evaluated morphologically. They correspond to the features and outcomes listed in table 1.

The analytical approach to PIN diagnosis and grading might be vulnerable to errors that can be related to the stage of visual image interpretation or classification. The visual image interpretation process is purely subjective because it relies on the templates in mind. To make it less subjective a series of microscope photographs that represent the individual characteristic features and their outcomes were made available to the observers evaluating the cases included in this study.

When considering the direction of feature outcome changes, nine cases with cumulative directional diagnostic distance from the high grade PIN category of 0 have a belief value greater than 0.70 whereas the other cases with mostly negative distances show belief values ranging from 0.50 to 0.70. The two cases with a belief value lower than 0.50 have a distance value of \(-2\).

**Discussion**

Figure 5 When the direction of feature outcome change is considered, the cumulative directional diagnostic distances of the 20 cases of high-grade PIN from the free diagnostic categories range from \(-8\) to \(+15\). In comparison with the prototype clue outcomes of PIN\(_{\text{high}}\), the diagnostic distances range from \(-3\) to \(+3\). The distance values from large acinar adenocarcinoma partly overlap with those from high grade PIN.

**Figure 6** Bivariate scatterplot of the diagnostic distances of the 20 cases of high grade PIN from the five diagnostic categories when considering both measures of rank differences. Separation of distances between the diagnostic categories is achieved.
They were forced to analyse the individual items reported in Table 1 (and graphically displayed in the series of photographs), and to select the outcomes that best matched those of the microscope image and that formed the feature outcome profile of the case. This approach was tried by one of our group (RM) in a recent morphometry related training course held in Sendai, Japan. The observer agreement in the outcome assignment was in excess of 90% even among participants with a low level of interest and experience in prostatic pathology.

The profiles of feature outcomes representing the different diagnostic categories and grades can be determined in two ways. It may be derived from the personal experience of an expert. For instance, according to experience, the nucleolar prominence is considered infrequent in normal prostate, some in low grade PIN, and frequent in high grade PIN and adenocarcinoma. This type of information is based on the fact that the expert has examined a large number of prostate pathology cases and has acquired full knowledge about the association between the degree of nucleolar prominence and the diagnostic categories. Another way to define the range of feature profiles is to analyse the frequency of the feature outcomes for each category and to select the most frequent outcome for a given diagnostic category. For instance, the nucleolar prominence in normal prostate is infrequent in 98% of the cases and some in 2%. From this frequency count, the information derived is that in normal prostate nucleolar prominence is infrequent. In the present study the outcome profiles were based on the experience of one of our group (RM). In a preliminary study involving some of the clue outcomes typical for the diagnostic categories listed in Table 2, concordance between the outcome selected as representative of a certain category and that derived from the observation frequency was observed.

A further issue concerns the comparison of the profile of the case under investigation with that of the diagnostic categories. This is done by calculating the (rank) difference between outcomes as shown in a previous paper. The diagnostic clues, related to the morphological evaluation of PIN following treatment with finasteride, were expressed in linguistic terms and the outcome options had been ordered according to the degree of severity. The sum of all ranked differences of the full string of diagnostic clues formed a metric measure called the cumulative diagnostic distance.

In the present study the rank differences were calculated in two ways, one considering only the difference in the clue expression and the other the difference as well as the direction of changes—whether there was an increase or a decrease in the assessment of severity in relation to the prototype clue expression in the five diagnostic categories. The possibility of incorporating the direction of changes in the rank difference was not considered in the previous study on the effect of finasteride. However, it was seen that, for some of the outcome comparisons, the PIN cases evaluated after treatment had a lower degree of change compared with the prototypes of untreated PIN categories. Therefore, it was felt important to explore the possibility of expressing the existence of a difference and, at the same time, of recording the direction of changes in the outcome evaluation.

The advantages and disadvantages of the calculation of the rank differences in the two ways were seen when the rank differences were summed to obtain the diagnostic distance of the individual cases from the five diagnostic categories. We first evaluated the diagnostic distances when the direction of changes was not considered. It was observed that the distribution of the distances from the different categories was such that an excellent separation of high grade PIN from all the other categories was shown. However, overlap between the categories other than high grade PIN was present. Therefore, we went on to explore whether the consideration of the increasing or decreasing degree of changes could have helped obtain a better separation among the distances. In some way the second approach was better than the first because it allowed us to spread the cases along a range of values far wider than the first and to order the diagnostic distances and the related categories according to their degree of malignancy progression. However, there was some overlap between high grade PIN and large acinar adenocarcinoma distances. The problem was solved when the cases were represented in a bivariate scattergram derived from both ways to measure the diagnostic distances.

The diagnostic distance evaluation applied here was developed by Bartels et al to deal with features expressed in linguistic terms. When the methodology was applied to the present study, it was done to identify the category prototype that matched the PIN case under investigation and therefore the category or grade to which it belonged. The way the diagnostic distance measure was applied corresponds to the measure of similarities adopted by Hamilton et al in a paper on expert system support in the diagnosis of breast fine needle aspiration biopsy. In their experience, the distance measure represented a helpful evidence for the diagnostic classification of each case.

Our diagnostic distance based methodology has some similarity to that adopted in the breast cancer grading originally devised by Bloom and Richardson and more recently investigated. Eller-Vahlin et al, in a series of studies, three features were considered and for each features three different outcomes analysed. For instance, nuclear pleomorphism (one of the features) had outcomes of small regular uniform cells, moderate increase in size and variability, and profound variation. Therefore, in our study and that by Bloom and Richardson, features of linguistic, descriptive type are used and the severity of changes is identified and ranked. The diversity in the approach is that while we calculated the rank difference and the final diagnostic distance, Bloom and Richardson used a score ranging from 1 to 3. In particular, 1 is given if the case being examined shows no nuclear pleomorphism (regular
uniform cells), whereas 3 is given if it is similar to profound variation. To obtain the overall tumour grade the scores from each of the three features are added together.

In the medical literature there are papers which deal with the problems of managing uncertainty in histopathology. Several approaches have been proposed to consider the dependencies between elements and combinations of such elements in a reasoning sequence. One of the most effective is the use of Bayesian belief networks, because the prior knowledge of the expert can be used to identify which dependencies are relevant. In a previous paper we described the development of a BBN to deal with the problem of uncertainty in the diagnosis and grading of untreated PIN. When this BBN was applied to this study, the results were such that the morphology based diagnosis and diagnostic distance results were confirmed by high belief values in the majority of cases. The diagnostic distance based methodology and the BBN, even though used for different purposes, have a common ground in that both rely on the same feature and outcome evaluation for the assignment of the cases into one of the categories. For the two diagnostic decision support systems to work effectively each diagnostic alternative has to be supported by at least one unique clue pattern, once all evidence nodes are polled. For the network to be robust, it is helpful if different evidence nodes provide unique clue patterns for a given diagnostic alternative, so that the redundancy provides some safety. In fact, only after combining several clues did the evidence become strong and lead to an accumulation of belief in the diagnostic alternatives. The need for redundancy was also observed by Bibbo et al in their belief networks for grading prostate adenocarcinoma. In their experience the network worked in a robust fashion when 13 diagnostic clues, eight based on tissue architectural features and five based on nuclear features, were employed.

In conclusion, the distance measure based methodology represents a useful diagnostic decision support tool for the accurate evaluation of high grade PIN.

This research was supported in part by a grant from the National Cancer Institute, R 35 CA 53877 (PHB). The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

The content of this paper was discussed at the Training Course on Stereology, 3-D Structural Analysis and Semi-automated Morphometry, Sendai, Japan, October 1996.