Observer variability in histopathological reporting of transitional cell carcinoma and epithelial dysplasia in bladders

A J Robertson, J Swanson Beck, R A Burnett, S R Howatson, F D Lee, A M Lessells, K M McLaren, S M Moss, J G Simpson, G D Smith, H B Tavadia, F Walker

Abstract

Sections from 90 urinary bladder biopsy specimens were examined by 11 consultant histopathologists with varying experience to determine the appropriateness of existing pathology terminology. Analysis with κ statistics showed fair to good agreement in the grading and staging of transitional cell carcinoma. There was also reasonable agreement in the diagnosis of high grade dysplasia in random biopsy specimens from the urothelium adjacent to the neoplasm, but very poor agreement for lesser degrees of dysplasia. It is concluded that the present classification of bladder carcinomata is reliable and that pathologists can determine stage with a high degree of reproducibility and grade with a fair degree of reproducibility.

Although flow cytometric assessment of nuclear DNA in urinary bladder transitional cell carcinoma has recently been shown to be of value as a prognostic indicator,[1] the need for expensive and technically demanding equipment confines most laboratories to the use of light microscopy. Tumour grade and stage are well recognised prognostic indicators,[2 3] and several different histological classification systems have been defined.[3-6] The World Health Organisation classification, which is now extensively used in this country, has four basic diagnostic criteria: pattern of growth, cell type, grade of malignancy and stage of the tumour. Previous studies have indicated that there may be observer variation in the application of these grading systems.[7-10] With the recognition that bladder carcinoma may result from a "field change phenomenon" clinicians are increasingly submitting random or selected biopsy specimens from sites other than the tumour itself. Assessment of these for urothelial dysplasia is important for determining subsequent clinical management. A previous study has indicated good reproducibility in the grading of dysplasia in such biopsy specimens.[11] It is important that clinicians and pathologists should be reassured of the reliability of histological reporting of bladder biopsy specimens.

Methods

Eleven consultant pathologists considered to be representative of Scottish histopathological practice as a whole contributed to the study. The members came from pathology laboratories in Aberdeen (n = 2), Dundee (n = 1), Edinburgh (n = 2), Airdrie (n = 1), Perth (n = 1), Stirling (n = 1) and Glasgow (n = 3) and varied in years of consultant experience (five to 25 years) and nature of substantive post (National Health Service staff n = 6: university staff n = 5). All members of the panel had undertaken both undergraduate and postgraduate training in Scotland.

The study comprised a series of 90 bladder biopsy specimens from patients presenting as new cases of bladder cancer to the urology unit at the Western General Hospital, Edinburgh. Each case had a series of four biopsy specimens taken: tumour, tumour base, urothelium near tumour and urothelium distant from tumour. The slides were screened by one of the group (AML) before inclusion in the study so that cases deemed to be technically unsatisfactory or which had inadequate material could be excluded. Two slides were circulated from each case—one from the putative neoplasm and one from a random biopsy specimen from elsewhere in the bladder. The study was coordinated from Dundee University by an independent organiser who allocated a confidential code to each participant. Three circulations, each of 30 cases, were studied consecutively, with a meeting of the group at the end of each circulation when slides about which there was substantial disagreement were discussed using a video camera and microscope projection system. A fourth circulation comprised a series of repeat cases from the previous batches 22 of which had caused difficulty when first seen, the remaining eight being randomly selected. The participants were aware that this circulation comprised repeat slides, but the identification codes had been changed.

The participants found that all the tumour biopsy specimens had abundant abnormal tissue. The random specimens taken for diagnosis of epithelial dysplasia were always small, usually about 3 mm in diameter. Desquamation of epithelium was present in some of these lesions but never affected more than 50% of the surface. The diagnosis was made on the residual basal cells in the area of desquamation and the epithelium adjacent to the affected zone.

Following review and discussion of the existing bladder pathology terminology a proforma was designed for completion after examination of each slide in the circulation (Appendix). The classification we used is based on that described by the WHO[5] and hence relates as closely as possible to standard diagnostic practice.

Observer variability was examined for three

Department of Pathology, Ninewells Hospital and Medical School, Dundee
A J Robertson
J Swanson Beck
R A Burnett
S R Howatson
F D Lee
A M Lessells
K M McLaren
S M Moss
J G Simpson
G D Smith
H B Tavadia
F Walker

Correspondence to: Professor J Swanson Beck, Department of Pathology, Ninewells Hospital and Medical School, PO Box 120, Dundee DD1 9SY, Scotland.

Accepted for publication 10 August 1989

main items: degree of differentiation, stage of the neoplasm, and grade of urothelial dysplasia in the random biopsy specimen. In the histopathological assessment of differentiation the category "papilloma" was never used by the pathologists, so only the three categories of well, moderately, and poorly differentiated transitional cell carcinoma were included. For assessment of stage a five-point scale was used for batch 1 with "invading muscle" being regarded as more serious than "invasion with no muscle present". After discussion by the group of the interpretation of these categories statistical analyses were also carried out with "possible invasion" grouped together with "within basement membrane", and "invasion with no muscle present" grouped with "definite invasion". For the subsequent circulations a further subdivision of "possible invasion" into two categories ("low" and "high" probability of invasion) was introduced. Analyses were then carried out using the full six-point scale, and a three-point scale with "possible invasion—low probability" being combined with the "non-invasive" and "possible invasion—high probability" with the "definite invasion" category.

The analysis of urothelial dysplasia used three categories: none, low grade, and high grade. In the final circulation of slides the low grade category was subdivided into two with inclusion of a borderline lesion to see if this assisted classification; but the use of the additional category was very limited, and the results are presented only for the combined low grade category.

The method of analysis used has been described in detail elsewhere.[12 13] In brief, κ statistics are a measure of overall agreement without requiring assumptions concerning the "correct" diagnosis and including a correction for the amount of agreement which could be expected by chance alone.[14] The overall value of κ for more than two categories is defined as the weighted average of the values for the individual categories.[15] The value of κ can range from $-1{\cdot}0$ to $+1{\cdot}0$—a value of 0 indicates chance agreement alone, while a value of $+1{\cdot}0$ indicates perfect agreement. (A negative value would indicate systematic disagreement among observers.) It is generally accepted that a value of 0·75 or above reflects excellent agreement, 0·4–0·75 fair to good agreement, and values of less than 0·4 poor

Transitional cell urinary bladder neoplasms

(a)

Slide No: *Tumour slide* ☐

QUALITY: Satisfactory = 1
Unsatisfactory = 2 ☐

SPECIFY ADDITIONAL DIFFERENTIATION:
Squamous cell = 1
Adenocarcinoma = 2
Other (please specify) = 3
None = 4 ☐

IS NEOPLASM: Papillary = 1
Papillary/Solid = 2
Solid = 3 ☐

DEGREE OF DIFFERENTIATION: Well differentiated = 1
Moderately ,, = 2
Poorly ,, = 3
Papilloma = 4 ☐

STAGE:
Within basement membrane = 1
Possible submucosal invasion = 2
Definite ,, ,, = 3
Invading muscle = 4
Submucosal invasion, but no muscle present = 5 ☐

Random Biopsy Slide

(b)

QUALITY: Satisfactory = 1
Unsatisfactory = 2 ☐

NORMAL: Yes = 1
No = 2 ☐

EVIDENCE OF DYSPLASIA: Low grade = 1
High grade = 2
None = 4 ☐

?OTHER LESION:

Hyperplasia	(Present = 1)	☐
Inflammation	(Absent = 2)	☐
Squamous metaplasia		☐
Cystitis cystica		☐
Glandular metaplasia		☐

PATHOLOGIST CODE: ☐

DATE: ☐☐☐☐☐☐

Table 1 Kappa statistics for batch 1 slides

Degree of differentiation:				
Well 0·22	Moderate 0·03	Poor 0·38		Overall 0·19
Stage:				
Within basement membrane		0·40	0·54	
Possible invasion		0·10		
Definite invasion		0·34	0·45	
Invasion: no muscle present		0·27		
Invading muscle		0·62	0·62	
Overall		0·32	0·51	
Dysplasia:				
Low grade 0·08	High grade 0·34	None 0·35		Overall 0·25

agreement.[16] As in previous studies a "majority" diagnosis was calculated for each slide for each of the classifications.

Results

The analyses for batch 1 are presented separately from those of batches 2–3 as it was felt that the level of agreement was most likely to change in the light of the considerable discussions following the initial circulation, and the stage classification was changed slightly at this point. Results for batch 4 are also presented separately because this was a recirculation of difficult cases. The reporting of this batch is also compared with that of the first circulation of the same slides incorporated in batches 1–3.

Table 1 shows the κ statistics for batch 1 for agreement on the degree of differentiation. Overall agreement was poor (κ = 0·19), being particularly low for the moderate category. When five separate categories are considered, overall agreement on stage of the neoplasm was poor (κ = 0·32), being highest, however, for the clinically important categories of within basement membrane and invading muscle (table 1). When the categories were grouped into three distinct possibilities only, the overall agreement improved (κ = 0·51). Agreement on epithelial dysplasia in the random biopsy specimens was poor (κ = 0·25), particularly for the category of low grade.

Table 2 gives the equivalent statistics for batches 2 and 3 combined. The agreement on both degree of differentiation and stage increased considerably compared with batch 1; this applies despite the further subdivisions of stage with the addition of one extra category. When stage was again grouped into three categories, the overall agreement increased to κ = 0·69 ("possible invasion +" grouped with "within basement membrane" and "possible invasion −"), or to κ = 0·66 ("possible invasion +" grouped with "definite invasion" and "invasion with no muscle present"). Agreement on invading muscle was then excellent. Agreement on urothelial dysplasia in the random biopsy specimens however,

Table 2 Kappa statistics for batches 2 and 3 slides

Degree of differentiation:				
Well 0·47	Moderate 0·28	Poor 0·60		Overall 0·44
Stage:				
Within basement membrane		0·61	0·67	0·70
Possible invasion (low probability)		0·06		
Possible invasion (high probability)		0·05		
Definite invasion		0·44	0·52	0·53
Invasion: no muscle present		0·27		
Invading muscle		0·86	0·86	0·86
Overall		0·51	0·66	0·69
Dysplasia:				
Low grade 0·06	High grade 0·41	None 0·31		Overall 0·23

Table 3 Kappa statistics for batch 4 slides

Degree of differentiation:				
Well 0·25	Moderate 0·09	Poor 0·36		Overall 0·22
Stage:				
Within basement membrane		0·39	0·41	0·46
Possible invasion (low probability)		0·02		
Possible invasion (high probability)		0·00		
Definite invasion		0·24	0·36	0·40
Invasion: no muscle present		0·34		
Invading muscle		0·79	0·79	0·79
Overall		0·26	0·41	0·45
Dysplasia:				
Low grade 0·05	High grade 0·44	None 0·24		Overall 0·21

remained poor, again most noticeably for low grade.

Table 3 presents the κ statistics for batch 4 for the three classifications. As might be expected, because this batch consisted of a recirculation including "difficult" slides, the κ statistics were generally lower than those for batches 2–3. Comparison with table 4, which presents the equivalent statistics for the same slides on first reading (as part of batches 1–3), however, shows that agreement on both stage and differentiation improved with a significant increase in κ ($p < 0·05$ for differentiation and stage in six categories; $p < 0·001$ for stage in three categories). There was a small but not significant increase in the κ value for urothelial dysplasia.

When the observations of individual pathologists on the two successive readings of each slide were compared (table 5), repeatability was good for stage (κ = 0·56), fair for differentiation (κ = 0·39), and poor for dysplasia (κ = 0·27).

Analysis of agreement with the majority diagnosis added little to the results. For the stage and dysplasia classifications, the majority diagnosis tended to be weighted towards one end of the scale (within basement membrane, no dysplasia), so that the deviation of individual pathologists from this consensus was usually toward a more severe diagnosis. In the recirculated slides the majority diagnosis was the same on stage for 80% of cases (24/30), on differentiation for 77% (23/30), and on urothelial dysplasia for 73% (22/30). In all three areas the trend was towards a lower category of diagnosis in those cases that changed.

Table 6 shows the relation between stage and degree of differentiation for batches 2–3: there was a clear association between poor differentiation of the neoplasm and increased certainty of invasion (rank correlation coefficient = 0·62). There was a weaker relation between stage and dysplasia (table 7, rank correlation coefficient = 0·23), with a slight tendency towards an increased certainty of invasion in

Table 4 Kappa statistics: previous circulation of batch 4 slides

Degree of differentiation:				
Well 0·20	Moderate 0·08	Poor 0·23		Overall 0·16
Stage:				
Within basement membrane		0·33	0·34	0·32
Possible invasion (low probability)		0·02		
Possible invasion (high probability)		0·06		
Definite invasion		0·15	0·32	0·30
Invasion: no muscle present		0·28		
Invading muscle		0·63	0·63	0·63
Overall		0·22	0·35	0·34
Dysplasia:				
Low grade 0·06	High grade 0·25	None 0·26		Overall 0·18

Table 5 Kappa statistics for repeat readings

Differentiation:			
Well differentiated 0·38	Moderate 0·29	Poor 0·55	Overall 0·39
Stage:			
Within basement membrane/possible invasion 0·55	Definite invasion 0·54	Invading muscle 0·77	Overall 0·56
Dysplasia:			
Low 0·15	High 0·40	None 0·34	Overall 0·27

Table 6 Correlation (No(%)) of stage and differentiation for batches 2 and 3 slides

	Differentiation:		
Stage	*Well differentiated*	*Moderately differentiated*	*Poorly differentiated*
Within basement membrane	160 (54)	120 (43)	8 (3)
Possible invasion (low probability)	8 (16)	40 (81)	1 (2)
Possible invasion (high probability)	5 (15)	18 (55)	10 (30)
Definite invasion	9 (8)	60 (50)	51 (43)
Invasion: no muscle present	3 (7)	18 (42)	22 (51)
Invading muscle	0	20 (17)	98 (84)
Total	185 (28)	285 (43)	190 (29)

cases with high grade dysplasia. Table 8 shows the relation between the degree of differentiation and presence of urothelial dysplasia, with a slight association between good differentiation of the neoplasm and absence of dysplasia (correlation coefficient = 0·26).

Discussion

Current treatment of bladder carcinoma depends on histological assessment of grade and stage, although other methods including DNA and RNA histograms, cytogenetic studies, presence of carcinoembryonic antigen, and detection of blood group substances from urothelium have been described. Tumour grade and stage, however, remain the most important variables. Two groups have previously studied the reproducibility of grading bladder carcinomas. Ooms and colleagues examined 57 neoplasms and found that there was considerable inconsistency within and among cases: in almost 50% the neoplasm was graded differently on a subsequent occasion by the same pathologist.[9] It was concluded that in an appreciable proportion of cases there would have been implications for management. Busch *et al* described the reproducibility of grading by one pathologist on three occasions and showed an overall consistency of 80%.[7]

As clinical decisions on malignancy are based on a histopathological diagnosis it is important that appropriate histopathological classifications and the reliability of their application in practice should be critically re-examined. External quality assurance programmes have recently been started in histopathology[17]: these involve circulation of a small batch of slides of varied tissues of origin and diagnosis to a group of pathologists with subsequent comparison of the reported diagnosis. Such programmes allow a pathologist to have some feedback about his or her diagnostic ability when compared with that of peers. In contrast to these programmes, the present study attempted to evaluate the pathologist's ability to recognise histopathological features that influence the subjective process of diagnosis and to provide information about the reliability of terminology. This allows clinicians to determine how much confidence can be placed on a diagnosis. We believe that the need for such information is becoming urgent because histopathological diagnoses directly influence patient management and must be both accurate and cost effective.

It was reassuring to find in this study that histopathologists are reasonably reliable in their ability to distinguish the various grades and stages of transitional cell carcinomas. The κ value was consistently better for stage than for the grade of neoplasm, and stage is the most important variable in clinical management. Our findings seem to compare favourably with those of Ooms *et al*[9] on the grading of bladder tumours by a number of pathologists who reported a high level of inconsistency, although the exact criteria used for measuring this inconsistency were not clear. Inconsistency in the same person, measured by the correlation coefficients of two readings, was also reported as high, with about 50% of cases being graded differently on the second reading. Direct comparison with the present study is difficult because the categories used for grading were not the same. Ooms *et al* went on to promote morphometric grading[10] as being more accurate than standard light microscopic assessment, but it seems unlikely that such methods would ever be universally adopted.

Table 7 Correlation (No(%)) of stage and dysplasia for batches 2 and 3 slides

	Dysplasia		
Stage	*None*	*Low grade*	*High grade*
Within basement membrane	201 (68)	81 (27)	15 (5)
Possible invasion (low probability)	26 (53)	23 (47)	
Possible invasion (high probability)	17 (52)	13 (39)	3 (9)
Definite invasion	46 (38)	43 (36)	31 (26)
Invasion: no muscle present	17 (40)	15 (35)	11 (26)
Invading muscle	54 (46)	38 (32)	26 (22)
Total	361 (55)	213 (32)	86 (13)

Table 8 Correlation of differentiation and dysplasia for batches 2 and 3 slides

	Grade of dysplasia			
Differentiation	*None*	*Low*	*High*	*Total*
Good	129 (70)	54 (29)	2 (1)	185 (100)
Moderate	154 (54)	110 (39)	20 (7)	284 (100)
Poor	76 (40)	49 (26)	61 (32)	186 (100)

Our findings on the correlation between stage and degree of differentiation in bladder carcinoma support those of other groups, including Kern[2] and Pagano *et al.*[18] The latter study also showed that dysplasia in random bladder biopsy specimens was found more frequently in high stage and grade neoplasms. Our study also shows an association between high grade urothelial dysplasia and advanced stage and poor differentiation of the neoplasm. It has previously been shown that patients with bladder carcinoma and associated epithelial dysplasia in a random biopsy specimen are at an increased risk of developing new tumours and also show an increased risk of progression. It is therefore important that the assessment of dysplasia should be accurate. Murphy and Soloway indicated that the major problem in histological assessment of these biopsy specimens involved mild/moderate lesions,[19] a suggestion strongly supported by our own study. Considerable difficulty was experienced in identification of low grade epithelial dysplasia. We believe that the following criteria should be used when assessing these biopsy specimens: in low grade lesions there is some loss of cell polarisation with mild variation in nuclear size and nuclear crowding and irregularity; in high grade lesions there is pronounced change in, or complete loss of, cell polarisation with abnormal superficial cells, considerable nuclear pleomorphism, and hyperchromatism. Desquamation of the superficial cells is also often a feature of the high grade lesions. In a recent study of primary dysplasia, of the eight cases diagnosed as mild, only one progressed.[20] Hence the difficulties associated with the assessment of mild dysplasia and its distinction from conditions such as regenerative atypia are probably not of major clinical importance and it is reassuring that the clinically important high grade dysplasia is more readily identifiable. Before our study only two pathologists in the group had had experience of assessing dysplasia in routine practice. When they were excluded from the statistical analysis there was no deterioration in the overall level of agreement. The difficulty with this assessment seems to be not one of the experience of the pathologist, but the unreliability of the histopathological criteria.

When the results for assessment of dysplasia in the bladder tumours were compared with our previous findings in assessment of cervical dysplasia,[13] we found that overall agreement in dysplasia in random biopsy specimens was worse for the bladder than for the cervix among the same group of pathologists.

In contrast to the study by Nagy *et al*,[11] who described continuing improvement in agreement about assessment of dysplasia over several circulations of a small set of photographs of lesions, we could find no evidence of a significant improvement in the ability to grade accurately this feature in our series of slides. It was of considerable interest, nevertheless, to note the considerable improvement in the κ values of the other features studied between the first and subsequent circulations. Although this may partly be attributable to the group adjusting to use of the proforma for reporting, we believe that a major contribution to this improvement was most likely the educational benefit of group discussion of the slides with use of a video projection system. This therefore seems to indicate a need for improved education and postgraduate pathology training facilities at all levels of expertise. With the increasing introduction of medical audit studies of this type are becoming more important.

In conclusion, this study has shown that the present classification of bladder carcinomas is reliable and that pathologists can determine stage with a high degree of reproducibility and grade with a fair degree of reproducibility. Agreement on the degree of epithelial dysplasia in random biopsy specimens is fair for high grade lesions but very poor for low grade lesions. We found that there is undoubted benefit from a process of continuing education and group discussion of such lesions. The findings indicate greater consistency than was found in previous studies of breast and cervical pathology[12 13] and should be reassuring to clinicians managing bladder carcinoma.

We are grateful to Mrs Joyce Duncan for secretarial assistance. This study is funded by the Scottish Home and Health Department Grant K/RES/D3/18.

1 Blomjous CEM, Schipper NW, Baak JPA, Van Galen EM, De Voogt HJ, Meyer CJLM. Retrospective study of prognostic importance of DNA flow cytometry of urinary bladder carcinoma. *J Clin Pathol* 1988;**41**:21–5.
2 Kern WH. The grade and pathological stage of bladder cancer. *Cancer* 1984;**53**:1185–9.
3 Bergkvist A, Ljungqvist A, Moberger G. Classification of bladder tumours based on the cellular pattern. *Acta Chir Scand* 1965;**130**:371–8.
4 Koss LG. Tumours of the urinary bladder. In: *Atlas of Tumor Pathology*. 2nd series, fascicle 11. Washington, DC: Armed Forces Institute of Pathology, 1975.
5 Mostofi FK, Sobin LH, Torloni H. Histological typing of urinary bladder tumours. In: *International histological classification of tumours No 10*. Geneva: World Health Organisation, 1973.
6 Pugh RCB. The pathology of cancer of the bladder. *Cancer* 1973;**32**:1267–74.
7 Busch C, Engberg A, Norlen BJ, Stenkvist B. Malignancy grading of epithelial bladder tumours. Reproducibility of grading and comparison between forceps biopsy, aspiration biopsy and exfoliative cytology. *Scand J Urol Nephrol* 1977;**11**:143–8.
8 Ooms ECM, Kurver PHJ, Veldhuizen RW, Alons CL, Boon ME. Morphometric grading of bladder tumours in comparison with histologic grading by pathologists. *Hum Pathol* 1983;**14**:144–50.
9 Ooms ECM, Anderson WAD, Alons CL, Boon ME, Veldhuizen RW. Analysis of the performance of pathologists in the grading of bladder tumours. *Hum Pathol* 1983;**14**:140–3.
10 Ooms ECM, Blok APR, Veldhuizen RW. The reproducibility of a quantitative grading system of bladder tumours. *Histopathol* 1985;**9**:501–9.
11 Nagy GK, Frable WJ, Murphy WM. Classification of premalignant urothelial abnormalities. A Delphi study of the National Bladder Cancer Collaborative Group A. *Pathol Annu* 1982;**17**:219–33.
12 Beck J Swanson and Members of the MRC Breast Tumour Pathology Panel. Observer variability in reporting of breast lesions. *J Clin Pathol* 1985;**38**:1358–65.
13 Robertson AJ, Anderson JM, Swanson Beck J, *et al*. Observer variability in the histopathological reporting of cervical biopsy specimens. *J Clin Pathol* 1989;**42**:231–8.
14 Cohen JA. A coefficient of agreement for nominal scales. *Educational Psychological Measurements* 1960;**20**:37–46.
15 Fleis JL. *Statistical methods for rates and proportions*. 2nd ed. New York: Wiley & Sons, 1981.
16 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
17 Lee FD, Burnett RA. Quality assurance in histopathology. *J Pathol* 1987;**152**:247–51.
18 Pagano F, Garbeglio A, Milani C, Bassi P, Pegoraro V. Prognosis of bladder cancer. 1. Risk factors in superficial transitional cell carcinoma. *Eur Urol* 1987;**13**:145–9.
19 Murphy WM, Soloway MS. Developing carcinoma (dysplasia) of the urinary bladder. *Pathol Annu* 1982;**17**:197–219.
20 Zuk RJ, Rogers HS, Martin JE, Baithun SI. Clinicopathological importance of primary dysplasia of the bladder. *J Clin Pathol* 1988;**41**:1277–80.