Virchows Arch A (1987) 411:479–484

Virchows Archiv A

Factors influencing the subjective grading of bladder cancer

Cecile Colpaert, Gerda Goovaerts, and Norbert Buyssens
Department of Pathology, University Hospital Antwerp, University of Antwerp, U.I.A.
Wilrijkstraat 10, B-2520 Edegem-Antwerpen, Belgium

Summary. In order to study factors which influence the subjective grading of bladder cancers the hypothesis of an ideal examination system was tested. The subjective results were compared with morphometry on the nuclear size of non selected cells. The results show that with precise criteria, with examination of the total section surface field by field and with recording of a grading per field a better interobserver consistency can be obtained. Morphometry demonstrates the heterogeneity and the Gaussian distribution of nuclear sizes in normal and diseased urothelium. It shows that subjective grading is influenced by a small number of larger nuclei and that even in grade 3 tumours 50% of the nuclei have sizes within normal values.

Key words: Bladder cancer – Grading – Morphometry

Introduction

When a panel of grading experts examines a set of sections they produce a classification without specifying how each member of the panel arrived at his conclusion. There must be inevitably differences in the time spent, in the conscious or unconscious priority given to established criteria, in the interpretation of this criteria (how slight is slight?), in the magnification used for overall grading and in the selection of microscopic fields or cells on which the final opinion was based. This results in divergent opinions and in interobserver consistencies which are sometimes less than 50%.

The difficulties in grading bladder tumours have been extensively studied by Ooms et al. (1983a, b). These authors introduced morphometry of nuclei in subjectively selected areas to improve the interobserver grading results and demonstrated that these were superior to histological grading alone (Ooms et al. 1985), but they did not address the questions raised above.

An ideal situation would be that where with strict criteria each pathologist would examine the total surface of the section with the same magnification and during the same time. This implies that:

1. the criteria should be sufficiently precise, avoiding individual interpretation; 2. the observer examines systematically all the microscopic fields of the section, that he grades each field separately and that he records his grading per field. The final grading is the sum of all separate gradings. Because he has to commit himself on each field he has to check intra-observer consistency continuously; 3. he records the time he needs for each section and compares it with that for other sections.

This system is similar to an analytic geographic air survey with built-in self controls.

The aim of this study was to demonstrate whether this system could improve grading accuracy. The sections were divided by indian ink lines into small nearly equal rectangular fields which allowed standardization of the surface, exact retrieval and easy interobserver comparison. Extensive morphometry was done on *non* selected cells in order to disclose objective features which could explain subjective grading results.

Material and methods

The material consisted of endoscopic resections of unselected papillary bladder tumours fixed in AFA or Bouin fixative, processed in Paraplast, cut in sections of 5 µm and stained with H and E or with Haemalun Safran Phloxin. All the sections were examined by the same 3 observers: 1 senior pathologist, (PR), 1 junior pathologist (DG) and 1 medical student (ST) who also performed all the morphometrical studies.

Offprint requests to: N. Buyssens

In order to standardize the microscopical fields which had to be examined the coverslip was marked with lines in permanent ink dividing the surface in 7×7 rectangles (Fig. 1). The surface of one rectangle corresponded roughly to a field of a $10\times$ objective and a $10\times$ ocular in a standard Zeiss microscope: the grading was done with a $10\times$ objective. The observer had to grade each field separately: his findings were recorded on a form with 7×7 boxes which were numbered from 1 to 49 (Fig. 2). The starting point marked by a cross on the section and on the form was the left upper corner and the readings were done from left to right and from top to bottom. The grading values of all the fields which contained epithelium were added and divided by the number of these fields: the result was a grading index which was 1.00 minimum if all the fields were grade I and 3.00 maximum if all the fields were grade III. Intra- or interobserver consistency was arbitrarely accepted if the difference was ± 0.15 or less.

Morphometry was done on the same sections. They were examined with an oil immersion objective $100\times$, a drawing tube and a MOP Video Plan Zeiss Kontron. The nuclear parameters studied were: area, perimeter and maximal diameter. In this study only the area will be reported, because it corresponds to the nuclear size as judged by the observer.

Three groups of cases were studied:

1. A pilot group of 7 cases. The grading criteria were those of the WHO classification (Mostofi et al. 1973). The cases were studied twice with an interval of one month. For morphometry normal urothelium in endoscopic resections of prostate was used as a control: in each of five different cases 200 adjacent nuclei in a random field were measured, thus totalling 1000 nuclei. In order to have an idea of the distribution of nuclear areas in a random field of a random choosen tumour all the nuclei of this field, totalling 1477 nuclei, were measured.

2. From the pilot study is was concluded that the method of multiple field examination was feasable, and therefore a group of 52 new unselected cases we studied. Because we experienced difficulties in judging nuclear size in the pilot study we made a preliminary morphometric study in order to quantify size differences better. We used the material of the pilot study: 1 case of grade I, 2 cases of grade II and 2 cases of grade III. In a random field 100 adjacent nuclei were measured. The results showed that in grade I the largest nuclei were 3 times larger than the smallest one, in grade II they were 5 times larger and in grade III 10 times and more: these figures were used in the criteria and listed as nuclear size variability.

Because the criteria of the WHO classification appeared to be rather limited the criteria were adapted for greater precision as follows:

Grade I: presence of epithelial papillae or strands, nuclei lying in parallel lines and at a regular distance from each other, mitoses in basal layers, nuclear shape oval, nuclear size variability 1–3.

Grade II: presence of coalescent epithelial papillae or sheets, nuclei lying in whorls and at an irregular distance, mitosis in the middle layers, nuclear shape oval, nuclear size variability 1–5.

Grade III: presence of epithelial sheets, nuclei irregularly distributed, mitoses in the superficial layers, nuclear shape irregular, nuclear size variability 1–10.

It was arbitrarily decided that if in a field even a small zone (less than 10%) of the surface showed a higher grade than the rest of the field, the whole field was assigned to this higher grade. No morphometry was done in this group.

3. Since the findings in group 2 did not allow a clear separation of grades a third group of 20 new cases was examined after adaptation and modification of the criteria. Because estimation of nuclear variability as done in group 2 appeared to be difficult, it was arbitrarily decided to take the diameter of a RBC as comparative standard.

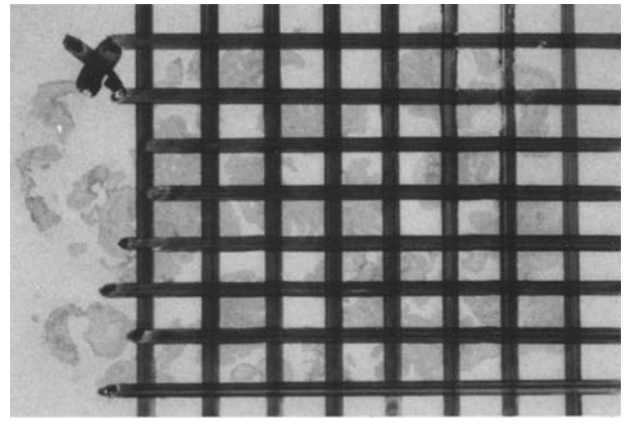

Fig. 1. Coverslip divided in 7×7 fields

1	2	3 1	4 2	5 2	6	7
8 1	9 1	10 2	11 2	12 2	13	14
15	16 1	17 2	18	19 2	20 2	21
22	23 1	24 2	25	26 1	27 2	28 1
29	30 2	31 3	32 2	33 2	34 1	35 1
36	37	38 2	39 1	40 1	41	42
43	44	45 2	46 2	47 2	48 1	49

Fig. 2. Example of form: 31 fields contain epithelium, the total score is 50, the grading index $50:31 = 1{,}93$

Grade I: free epithelial papillae, stratification of nuclei vertical in basal layers and horizontal in upper layers, nuclei not larger than $2\times$ the diameter of a RBC, nuclear shape oval, chromatin fine, mitoses in basal layer, nuclear cytoplasmic ratio <1.

Grade II: coalescent epithelial papillae, stratification of nuclei oblique and formation of whorls, nuclei larger than $2\times$ the

diameter of an RBC, nuclear shape rounded, chromatin coarse, mitosis up to the middle layer, nuclear cytoplasmic ratio ± 1.

Grade III: epithelial sheets, no stratification of nuclei, nuclei larger than 2× the diameter of a RBC, nuclear shape irregular, chromatin in blocs, mitoses in all levels, nuclear cytoplasmic ratio >1.

Since fields which were completely homogenous for one grade were few, it was arbitrarily decided that if 70% or more of the surface of a field belonged to one grade, the field was assigned to this grade. If it was less the field was graded as mixed grade 1/grade 2, grade 1/grade 3 or grade 2/grade 3. Morphometry was done by measuring 295 nuclei on: a) one random field graded homogenous I, II or III by the 3 observers, in total 885 nuclei (3×295), b) on one random field in each case, in total 5900 nuclei (20×295). The field number 3 was used for this purpose. The nuclei were measured starting in the left upper corner, going from left to right and from top to bottom. Statistical analysis of morphometric values was performed by the *t*-test. Intra- and interobserver consistency have not been statistically analysed because of the small number and the values must be regarded as indicative.

Results

For the pilot study of 7 cases the intra- and the interobserver consistency are listed in Table 1. Except for the first confrontation between observers DG and ST the results were better than 60%.

The results of the nuclear areas in the normal are shown in Table 2 and an example of a mid value in Fig. 3: more than 90% of the nuclei were within a range between 20 and 40 μm^2.

The distribution of all the nuclei in 1 random field of the tumour showed a Gaussian distribution and more than 60% were within the normal range (Fig. 4). There were many more smaller nuclei than in the normal. The largest nuclei were 12× the size of the smallest.

In group 2 (52 cases) the grading indexes of all the cases examined by the three observers were compared (Fig. 5): the graph shows the distribution of the $52 \times 3 = 156$ observations. For simplicity only 1 decimal was used. The curve is continuous and does not allow any division in groups of grading indices. The interobserver consistency (Table 3) was slightly better than 50%.

In group 3 (20 cases) the grading indices of the 3 observers are shown in the graph of Fig. 6. The curve is continuous with a high median peak. The interobserver consistency (Table 4) was at least 80%.

The morphometric values for the homogenous grade I, II and III showed a Gaussian distribution. In grade I the mean was 32.29 μm^2, SD 6.83 and nearly 90% of the nuclei were within normal values (Fig. 7). In grade II the mean was 34.61 μm^2, SD 9.11 which was only slightly higher than for grade I, but significantly different which was also re-

Table 1. Interobserver consistency in the pilot group at first observation and one month later. Intraobserver consistency

Interobserver consistency				Intraobserver consistency	
Initial		Repeated			
Pr. –DG.	5/7	Pr. –DG.	5/7	Pr.	6/7
Pr. –St.	5/7	Pr. –St.	4/7	DG.	6/7
DG.–St.	1/7	DG.–St.	6/7	St.	5/7

Table 2. Areas of 200 nuclei in each of 5 normal cases

Area μm^2	Number				
	1	2	3	4	5
Minimum	5	8	12	12	8
Maximum	71	44	59	51	83
X	27	21	27	33	38
SD	10	7	7	8	10

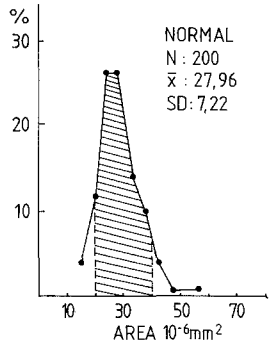

Fig. 3. Example of nuclear area of 200 nuclei in normal urothelium. The dotted lines here and in the following graphs show the range of 90% of the values

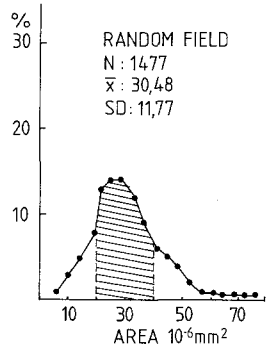

Fig. 4. Areas of the 1477 nuclei in a random field of a random bladder cancer

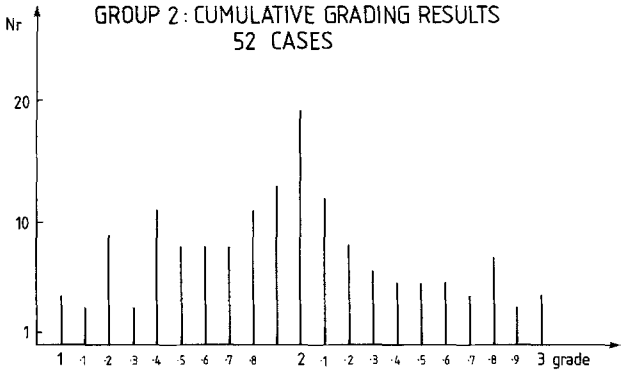

Fig. 5. Distribution of the grading indices of 52 cases by three observers, totalling 156 observations

Table 3. Interobserver consistency in 52 cases

52 cases Interobserver consistency	
Pr. –DG.	26/52
Pr. –St.	30/52
DG.–St.	29/52

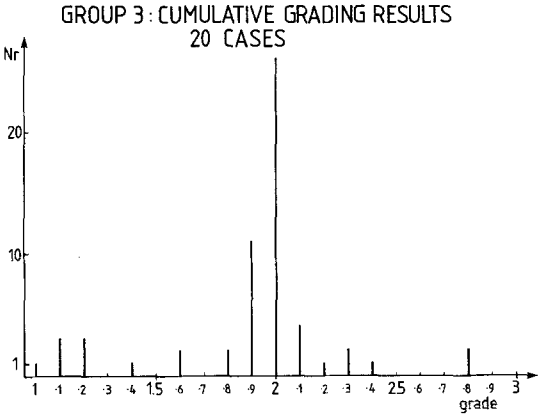

Fig. 6. Distribution of the grading indices of 20 cases by three observers, totalling 60 observations

Table 4. Interobserver consistency in 20 cases

20 cases Interobserver consistency	
Pr. –DG.	16/20
Pr. –St.	18/20
DG.–St.	16/20

flected in the higher SD and the wider spreading due to a moderate increase in larger nuclei; still 70% of the nuclei were within normal values (Fig. 8). The mean of grade III was 48.09 μm^2, SD 24.15, the spreading was wide and the largest

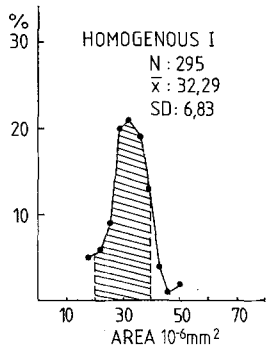

Fig. 7. Areas of 295 nuclei in Grade 1

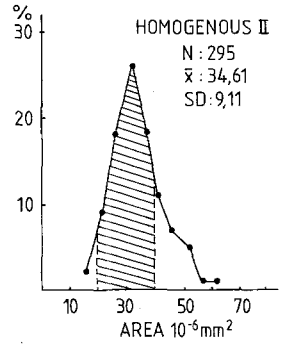

Fig. 8. Areas of 295 nuclei in Grade 2

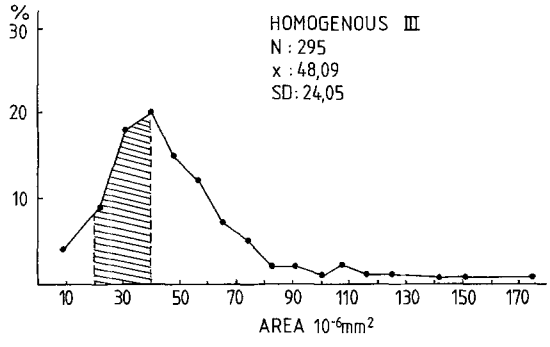

Fig. 9. Areas of 295 nuclei in Grade 3

nucleus had an area which was 18 times larger than the smallest, but 50% of the population was within normal values (Fig. 9).

The morphometric results of the random fields in each of the 20 cases, totalling 5900 nuclei, showed a wide spreading of means between 18 μm, SD 7 and 48 μm^2, SD 24. However some means were clearly distinct and examples of a low (Fig. 10), an intermediate (Fig. 11) and a high mean (Fig. 12) are shown.

The curves were similar to the previous group. The expected correspondance between the mor-

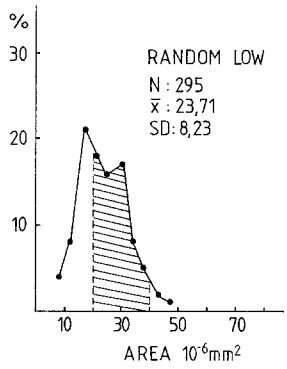

Fig. 10. Areas of 295 nuclei in a random field with low values

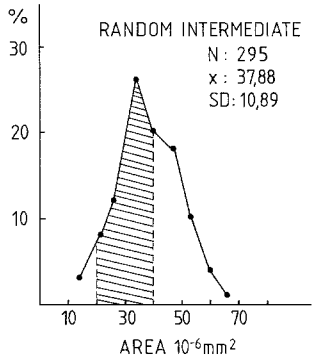

Fig. 11. Areas of 295 nuclei in a random field with intermediate values

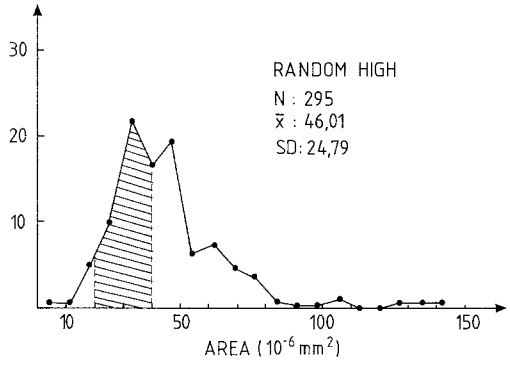

Fig. 12. Areas of 295 nuclei in a random field with high values

phometry values and the grading could not be strictly established.

Other findings were that the lines on the coverslip occupied $\pm 50\%$ of the surface, the surface of epithelium examined was ± 40 mm^2 (15–83) per case, the number of epithelium containing fields was ± 37 (25–44) per case, and that the time needed for the examination of one case was 20–30 min.

Discussion

This study demonstrates that with sufficiently precise criteria and systematic examination of the whole section more reproducible grading results can be obtained. By moving the microscope stage field by field the observer can reproduce the rectangular indian ink fields we used for study. A time of 20–30 min is excessive for daily practice but before this study was undertaken the senior author has often used (and still uses) this field by field examination system and did it in less than 10 min. This practice was in fact the incentive for this study.

Our findings demonstrate the heterogeneity of urothelial carcinomas since fields which were homogeneous for one grade were few among more than 3000 fields studied. It also became clear that in biological systems which are expressed as a Gaussian curve, criteria must be preset to allow distinction into different classes: for this reason we had to change the criterion for the surface of the more abnormal tissue. In group 2 we accepted a minimal surface of higher abnormality as sufficient for a higher grade and the result was that we arrived at a disproportionately high number of grade 2 tumours. By putting the cut off line at 70% or more for the majority of the cells per field in group 3 we arrived at a clearer distinction between the grades.

During this study the well known fact that criteria in grading systems are very general was confirmed: we all know that the lack of precision is done on purpose in order not to complicate matters too much. Our study shows that more precise criteria can be used without difficulty. The use of a computed grading index gives a good idea of the heterogeneity among fields and hence of the tumour. Since a continuum of values between 1 and 3 exists, the observers can determine for themselves the limits of the different grades and arrive at a general agreement. The morphometric data demonstrate the heterogeneity of the nuclei within each field, also in normal urothelium. Pathologists know this feature very well but they certainly do not know the magnitude of it: which pathologist would believe that in a grade 3 bladder carcinoma more than 50% of the nuclei are of normal size? You need cumbersome morphometry to prove it! Morphometry also shows that higher grading is related to a wider spreading of the values and to relatively few larger nuclei: this demonstrates that pathologists judge "selectively" and have little consideration for the silent majority of normal sized nuclei. This is not necessarily a shortcoming

because it is very likely that the more abnormal cells are the more malignant. Our morphometric studies demonstrate the absence of distinct nuclear size classes in normal and diseased urothelium, which means that when the human eye would be replaced by automatic size registration devices we still would have to draw conventional limits for grading purposes. Finally, if we ever want to develop expert systems for tumour grading we will need objective corrections for our subjective judgment.

Acknowledgement. The authors wish to thank Denise Van de Venne for typing the manuscript.

References

Mostofi FKM, Sobin LH, Torloni H (1973) Histological Typing of Urinary Bladder Tumors. International Histological Classification of Tumors, 10, Geneva, World Health Organisation

Ooms ECM, Anderson WAD, Alons ACL, Boon ME, Veldhuizen RW (1983a) Analysis of the Performance of Pathologists in the Grading of Bladder Tumors. Hum Pathol 14:140–143

Ooms ECM, Kurver PHJ, Veldhuizen RW, Alons CL, Boon ME (1983b) Morphometric Grading of Bladder Tumors in Comparison with Histologic Grading by Pathologists. Hum Pathol 14:144–150

Ooms ECM, Block APR, Veldhuizen RW (1985) The reproducibility of quantitative grading system of bladder tumors. Histopathology 9:501–509

Accepted June 5, 1987