

Molecular and clinical support for a four-tiered grading system for bladder cancer based on the WHO 1973 and 2004 classifications

Bas WG van Rhijn^{1,2,3,4}, Mireia Musquera¹, Liyang Liu³, André N Vis^{2,4}, Tahlita CM Zuiverloon^{2,5}, Geert JLH van Leenders⁵, Wim J Kirkels², Ellen C Zwarthoff⁵, Egbert R Boevé⁴, Adriaan C Jöbsis⁶, Bharati Bapat³, Michael AS Jewett¹, Alexandre R Zlotta^{1,7} and Theo H van der Kwast^{5,8}

¹Division of Urology, Department of Surgical Oncology, University Health Network, Princess Margaret Hospital, Toronto, ON, Canada; ²Department of Urology, Erasmus MC, Rotterdam, The Netherlands; ³Department of Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada; ⁴Department of Urology, Sint Franciscus Hospital, Rotterdam, The Netherlands; ⁵Department of Pathology, Josephine Nefkens Institute, Erasmus MC, Rotterdam, The Netherlands; ⁶Department of Pathology, Sint Franciscus Hospital, Rotterdam, The Netherlands; ⁷Department of Urology, Mount Sinai Hospital, University of Toronto, Toronto, ON, Canada and ⁸Department of Surgical Pathology, University Health Network, Toronto, ON, Canada

Currently, the use of two classification systems for bladder cancer grade is advocated in clinical guidelines because the WHO2004 classification has not been sufficiently validated with biological markers and follow-up. The slides of 325 primary non-muscle invasive bladder cancers from three hospitals were reviewed by one uropathologist in two separate sessions for the WHO1973 (G1, G2 and G3) and 2004 (papillary urothelial neoplasm of low malignant potential (LMP), low-grade (LG) and high-grade (HG)) classifications. *FGFR3* status was examined with PCR-SNaPshot analysis. Expression of Ki-67, P53 and P27 was analyzed by immunohistochemistry. Clinical recurrence and progression were determined. We performed validation and cross-validation of the two systems for grade with molecular markers and clinical outcome. Multivariable analyses were done to predict prognosis and pT1 bladder cancer. Grade review resulted in 88 G1, 149 G2 and 88 G3 lesions (WHO1973) and 79 LMP, 101 LG and 145 HG lesions (WHO2004). Molecular validation of both grading systems showed that *FGFR3* mutations were associated with lower grades whereas altered expression (Ki-67, P53 and P27) was found in higher grades. Clinical validation showed that the two classification systems were both significant predictors for progression but not for recurrence. Cross-validation of both WHO systems showed a significant stepwise increase in biological (molecular markers) and clinical (progression) potential along the line: G1—LG—G2—HG—G3. The LMP and G1 categories had a similar clinical and molecular profile. On the basis of molecular biology and multivariable clinical data, our results support a four-tiered grading system using the 1973 and 2004 WHO classifications with one low-grade (LMP/LG/G1) category that includes LMP, two intermediate grade (LG/G2 and HG/G2) categories and one high-grade (HG/G3) category.

Modern Pathology (2015) 28, 695–705; doi:10.1038/modpathol.2014.154; published online 28 November 2014

In the western world, bladder cancer is the 4th most common cancer in men and the 10th in women.¹ Most (70–80%) bladder cancers are non-muscle

invasive at first diagnosis (pTa, pT1, carcinoma *in situ*).^{2–5} Generally, the prognosis is good although 30–80% will recur and 1–45% will progress within 5 years.^{3–7} Stage and grade are the most important variables for progression as opposed to size and multiplicity, which are stronger predictors of recurrence.^{5–7} Historically, both three-tiered (1966-Bergkvist, 1973-WHO, 1998-WHO/ISUP and 2004-WHO) and four-tiered (1982-Modified Bergkvist and 1999-WHO) grading systems for bladder cancer have been described.^{5,8}

Correspondence: Dr BWG van Rhijn, MD, PhD, FEBU, Department of Surgical Oncology (Urology), Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

E-mail: basvanrhijn@hotmail.com

Received 12 September 2013; revised 5 October 2014; accepted 6 October 2014; published online 28 November 2014

Point mutations in the *FGFR3* gene are well documented in inherited skeletal anomalies associated with dwarfism, such as achondroplasia and thanatophoric dysplasia.⁹ An oncogenic role has been proposed for mutant *FGFR3* in bladder cancer.¹⁰ Surprisingly, *FGFR3* mutations were related to favorable bladder cancer with 84% of pTaG1 tumors having a mutation as compared with 7% of \geq pT2G3 lesions.¹¹ In non-muscle invasive bladder cancer, *FGFR3* mutations were associated with a low risk for progression^{12–15} and genetically stable disease.^{16,17} On the other hand, increased expression of the proliferation marker Ki-67, nuclear overexpression of P53 and decreased expression of the cell-cycle inhibitor P27 have been linked to higher grade and progressive non-muscle invasive bladder cancer.^{12,14,18} These four molecular markers are well-established representatives of the two major pathways in bladder cancer carcinogenesis.¹²

Currently, the simultaneous use of two classification systems for grade is advocated by the American Urological Association³ and European Association of Urology⁴ guidelines because the WHO2004 classification¹⁹ has not been sufficiently validated against the WHO1973 system²⁰ with biological markers and follow-up. The main reasons to propose a new classification were the lack of clear definitions for the three WHO1973 grades and the high percentage of bladder cancer classified as G2 (the default diagnosis). The WHO 2004 classification system was proposed in 1998 by the WHO/International Society of Urological Pathology and adopted as a three-tiered (papillary urothelial neoplasms of low malignant potential (LMP), low-grade (LG) and high-grade (HG) urothelial cancer) grading system by the WHO in 2004.^{19,21} First, the WHO2004 system has been based on the detailed histological criteria. Second, patients with LMP would have lower recurrence percentages and patients with LG non-muscle invasive bladder cancer would have close to no progression compared with HG cases. Consequently, the new classification led to a grade shift precluding a one-to-one translation between both grading systems: WHO1973 G1 tumors became WHO2004 LMP or LG, G2 became LG or HG and G3 are all HG (Figure 1).^{5,19–25}

In the present study, we hypothesized that the biological potential and clinical course (recurrence and progression) across the WHO1973 and 2004 systems would increase as follows: LMP < G1 < LG < G2 < HG < G3 (Figure 1). To test this hypothesis, we performed clinical and molecular validation for both WHO classifications. Second, we performed clinical and molecular cross-validations between the two WHO classifications to explore differences between LMP, LMP/G1, G1/LG, LG/G2, G2/HG and HG/G3. This led to the proposal of a four-tiered classification system for grade.



Figure 1 Comparison of the two (WHO 1973 and 2004) classification systems for grade in bladder cancer, which are both currently advocated by clinical guidelines. We hypothesized that the biological and clinical potential across the WHO1973 and 2004 systems would increase along the grades LMP, G1, LG, G2, HG and G3 as shown by the dashed arrows in the figure. Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; G1, grade 1; etc.; WHO, World Health Organization.

Materials and methods

Patients

The study group consisted of 325 primary (first diagnosis) non-muscle invasive bladder cancers from two hospitals in Rotterdam, The Netherlands (Sint Franciscus Hospital: $N=117$ and Erasmus MC: $N=121$) and one hospital in Toronto, Canada (University Health Network, Princess Margaret Hospital: $N=87$) in the period 1986–2006. None of them had a hereditary skeletal disorder documented. Mean age at diagnosis was 66.4 years (standard deviation: 11.8 years). All patients underwent a macroscopically complete trans-urethral resection with muscle in the specimen. Random biopsies, a standard re-resection and a single instillation of chemotherapy after resection were not routinely done in the study era. The surveillance of the patients consisted of urethro-cystoscopy and cytology every 3–4 months in the first 2 years and at a lower frequency, every 6–12 months, if no recurrence was detected during follow-up. Upper urinary tract imaging was done every 2 years in high-risk cases or if indicated by clinical suspicion. Recurrence was defined as the development of histologically confirmed urothelial cancer in follow-up. Progression was defined as recurrent disease in the muscularis propria (\geq pT2) and/or metastasis. Patients were followed until their last clinical visit, progression or death.

Pathology Review

We used newly cut or archived (4 μ m), hematoxylin and eosin (H&E) stained histological sections from archival tumor tissue, which was fixed in 10% buffered formaldehyde, dehydrated and embedded in paraffin. Histological grade was reviewed by one uro-pathologist (TvdK) for the WHO1973 and 2004 classifications systems. This was done in two separate rounds. The pathologist was unaware of molecular and/or clinical data.

Immuno-Histochemistry

All cases were routinely processed in two laboratories (Erasmus MC, Rotterdam and Mount Sinai Hospital, Toronto). The monoclonal antibodies against Ki-67 (clone MIB-1), P53 (clone DO-7) and P27 (clone 1B4 (Rotterdam) or clone 57 (Toronto)) were used. Known positive and negative controls were included in each run. The slides were independently assessed by BvR/AV (Rotterdam) and by BvR/TvdK (Toronto). The cutoff levels were 25% (Ki-67), 10% (P53) and 50% (P27).^{12,15} We used immuno-histochemical staining of whole slides for the Rotterdam cases and tissue microarray technology was used in Toronto.

FGFR3 Mutation Analysis

H&E slides served as templates for the manual micro-dissection procedure on the tissue-block. In case of multifocality, the papillary lesion with the highest grade/stage was taken. The largest tumor was taken if grade/stage was the same for multiple cancers. DNA was extracted using the DNeasy Tissue kit (Qiagen, Hilden, Germany). The *FGFR3* mutation analysis was performed in two institutes (Erasmus MC, Rotterdam and Mount Sinai Hospital, Toronto) with the PCR-SNaPshot method.¹⁵ In brief, three regions (exons: 7/10/15) were simultaneously amplified by PCR. After removal of excess primers and deoxynucleotides, specific SNaPshot primers able to detect 11 known oncogenic *FGFR3* mutations were annealed to the PCR products, separated by capillary electrophoresis and analyzed in an automatic sequencer.¹⁵ The codon numbering was according to the *FGFR3b* isoform expressed in epithelium.¹¹

Statistics

The SPSS computer software version 21.0 was used for data documentation and analysis. Two-sided Fisher's exact test was used to evaluate differences between molecular markers and the two WHO classification systems (molecular validation) and between categories of the two WHO systems (molecular cross-validation). The clinical outcome (recurrence and progression) in relation to tumor grade and the four molecular markers was analyzed using Kaplan–Meier statistics (clinical validation). Kaplan–Meier statistics were also applied to test differences in clinical outcome for a given grade category in one WHO system with two categories from the other WHO system (clinical cross-validation). Multivariable Cox regression analyses (backward stepwise method with *P* to enter <0.05 and *P* to remove >0.10) were performed to find independent prognostic variables among multiplicity, tumor size, carcinoma *in situ*, grade (WHO 1973, WHO 2004 and combined four-tiered grading system) and stage. A binary logistic regression model (backward

Table 1 Patient, tumor and molecular characteristics

	No. (%)
<i>Gender</i>	
Female	71 (22)
Male	254 (78)
<i>No. of tumors</i>	
Solitary	219 (67)
Multiple	106 (33)
<i>Tumor size (cm)</i>	
≤3	184 (57)
>3	141 (43)
<i>Grade (WHO 1973)</i>	
G1	88 (27)
G2	149 (46)
G3	88 (27)
<i>Grade (WHO 2004)</i>	
LMP	79 (24)
LG	101 (31)
HG	145 (45)
<i>Stage</i>	
pTa	171 (53)
pT1	154 (47)
<i>Associated CIS</i>	
No	263 (81)
Yes	62 (19)
<i>Instillation type</i>	
None	59 (18)
Chemotherapy	41 (13)
BCG	168 (52)
BCG + Chemotherapy	57 (17)
<i>No. of instillations</i>	
0	59 (18)
4–6	60 (19)
7–12	56 (17)
13–18	53 (16)
>18	97 (30)
<i>FGFR3 gene status</i>	
Wild type	138 (42)
Mutant	187 (58)
<i>Ki-67 expression</i>	
Normal	201 (62)
Altered	124 (38)
<i>P53 expression</i>	
Normal	228 (70)
Altered	97 (30)
<i>P27 expression</i>	
Normal	255 (79)
Altered	70 (21)
Total	325 (100)

Abbreviations: no., number; CIS, carcinoma *in Situ*; BCG, Bacillus Calmette-Guérin; LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; *FGFR3*, fibroblast growth factor receptor 3 gene.

stepwise method with P to enter <0.05 and P to remove >0.10) was constructed to predict stage (pTa or pT1) with the variables mentioned above. Multivariable testing was done with and without the molecular markers (*FGFR3*, Ki-67, P53 and P27) and by including each of the three classification systems for grade combined and separately. Statistical significance was assumed if $P < 0.05$.

Results

The patient, tumor and molecular characteristics of the studied population are shown in Table 1. Figure 2 contains examples of the *FGFR3* mutation analysis and altered expression of Ki-67, P53 and P27. Table 2 shows the head-on comparison between the WHO 1973 and 2004 classification systems for grade and stage. Data sets over time at 5-year intervals (1986–1990, 1991–1995, 1996–2000 and 2001–2006) did not show any differences related to pathological and clinical outcome (data not shown). As expected from Figure 1, the WHO 2004 system leads to grading into higher categories. This result was more pronounced in less differentiated non-muscle invasive bladder cancer.

Clinical Validation: WHO 1973 and 2004 and Molecular Markers

We determined recurrence and progression as prognostic end points. Recurrence was found in 219 (67%) patients and 63 (19%) patients progressed. Median and mean follow-up for recurrence were 2.1 and 3.4 years (interquartile range; 0.7–5.2 years), respectively. Median and mean follow-up for progression were 6.8 and 7.3 years (interquartile range; 3.8–9.3 years), respectively. Figure 3a and b shows that both WHO systems significantly predicted progression providing clinical validation. Conversely, neither system was able to predict recurrence (WHO 1973; $P = 0.441$ /WHO 2004; $P = 0.552$). Figure 4a–d contains the Kaplan–Meier curves providing clinical validation of the *FGFR3* gene status and Ki-67, P53 and P27 expression as prognostic markers for progression. The P -values were highly significant ($P < 0.001$). *FGFR3* mutations were related to favorable disease whereas altered expression of Ki-67, P53 and P27 was associated with worse outcome in NMI-BC. Of the four molecular markers, only Ki-67 ($P = 0.025$) and P53 ($P = 0.031$) expression significantly predicted recurrence.

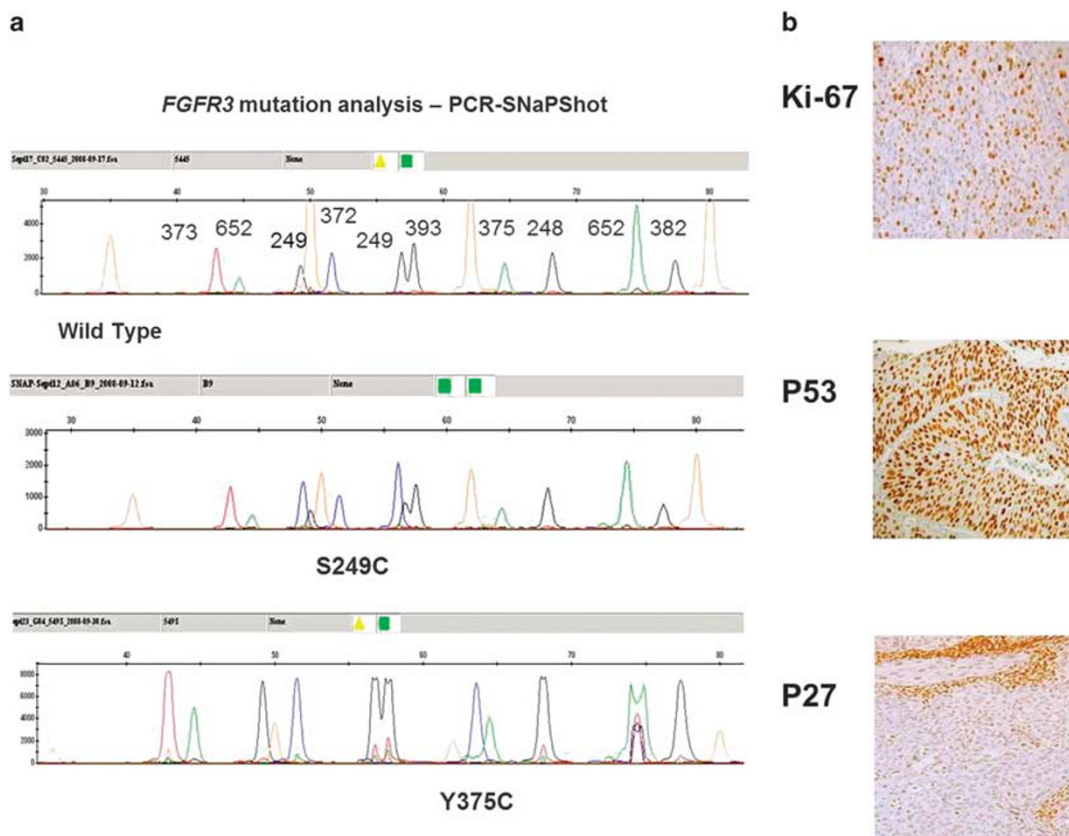


Figure 2 Examples of PCR-SNaPShot (*FGFR3* mutation analysis) and altered expression of Ki-67, P53 and P27. (a) Examples of the *FGFR3* mutation analysis are shown. The top-example shows the wild-type gene with the codon numbering according to the *FGFR3b* isoform. Below are the two most common *FGFR3* mutations in bladder cancer, ie, S249C and Y375C which we found in 126 (67%) and 30 (16%) tumors, respectively. The colors referring to specific nucleotides are as follows: G = blue, A = green, T = red and C = black. Orange is a size marker. (b) Examples of altered expression of Ki-67 ($>25\%$), P53 ($>10\%$) and P27 ($<50\%$) are shown. *FGFR3*, fibroblast growth factor receptor 3 gene.

Molecular Validation: WHO 1973 and 2004

The distribution of the molecular markers across the two WHO classifications is given in Table 3. In general, the frequency of *FGFR3* mutations decreased with higher grades and the expression of Ki-67, P53 and P27 increased with higher grades. In addition, comparing WHO1973 with 2004, *FGFR3* mutations were more frequent in HG (38%) than

Table 2 (A) Comparison of histological grade according to the WHO 1973 and 2004 classification systems. Please note that one pathologist reviewed both classification systems in two separate sessions. (B) Comparison of stage according to the WHO 1973 and 2004 classification systems

(A)	WHO 2004			Total
	LMP	LG	HG	
WHO 1973				
G1	56	29	3	88
G2	23	72	54	149
G3	0	0	88	88
Total	79	101	145	325

(B)	WHO 2004/1973			Total
	LMP/G1	LG/G2	HG/G3	
Stage				
pTa	79/88	65/75	27/8	171
pT1	0/0	36/74	118/80	154
Total	79/88	101/149	145/88	325

Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma. These tables show that the WHO2004 classification system is characterized by higher grading compared with WHO1973.

in G3 (18%). On the other hand, altered expression (Ki-67, P53 and P27) was more frequently found in G3 than in HG. Similar differences in molecular profile can be appreciated for LG and G2 but not for LMP and G1 (Table 3). We did not find a difference between laboratories and different antibody use (P27) in relation to pathological variables and clinical outcome (data not shown). Taken together, our results pointed to a significant stepwise increase in biological aggressiveness along the line LMP/G1, LG, G2, HG and G3.

Molecular Cross-Validation: WHO 1973 and 2004

The results from Table 3 prompted us to perform a cross-validation to investigate whether molecular differences can be found within one grade category itself for one WHO system by using the other WHO system. The molecular cross-validation of the WHO 2004 system with the WHO 1973 classification system is shown in Table 4A. When we tested the categories G1-2 vs G3 in the HG category, three of four molecular markers were significantly different. This implied that within the HG category, G3 has a more aggressive biological potential than cases graded G1-2. We also tested G1 vs G2-3 within the LG category. We found that two of the four molecular markers were significantly different (Table 4A). The molecular cross-validation of the WHO 1973 system with the WHO 2004 classification system is shown in Table 4B. Within the G2 category, three of the four molecular markers were significantly different for LMP-LG vs HG. Within the G1 category, none of 4 was different for LMP vs LG-HG. Taken together, the molecular cross-validations confirmed a significant stepwise increase in biological aggressiveness along the line LMP/G1, LG, G2, HG and G3.

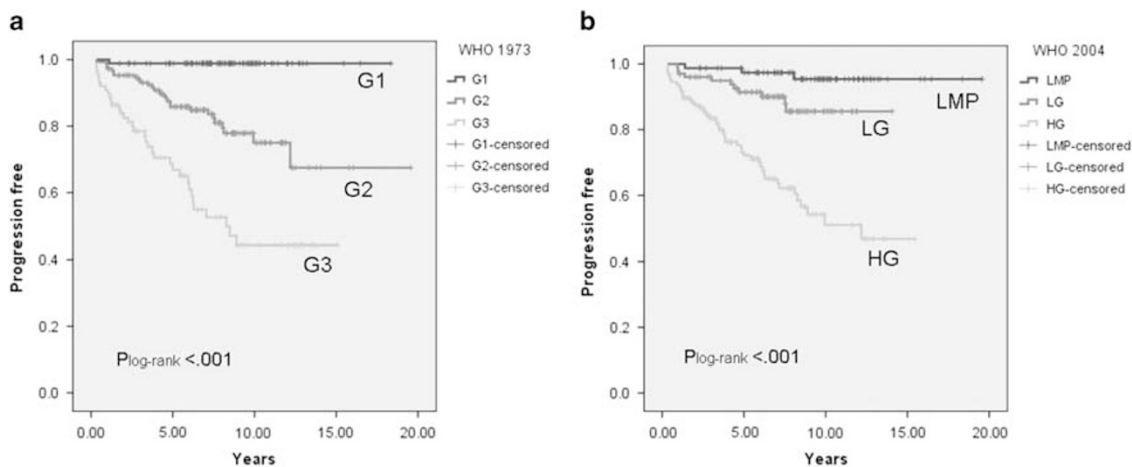


Figure 3 This figure shows the Kaplan–Meier curves providing clinical validation of the WHO 1973 (a) and 2004 (b) classification systems for grade to predict progression of non-muscle invasive bladder cancer. The *P*-values (log-rank) were highly significant ($P < 0.001$) for both grading systems. The *P*-values (log-rank) comparing G1/G2, G2/G3, LMP/LG and LG/HG were = 0.003, < 0.001, = 0.038 and < 0.001, respectively. LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; G1, grade 1; etc.; WHO, World Health Organization.

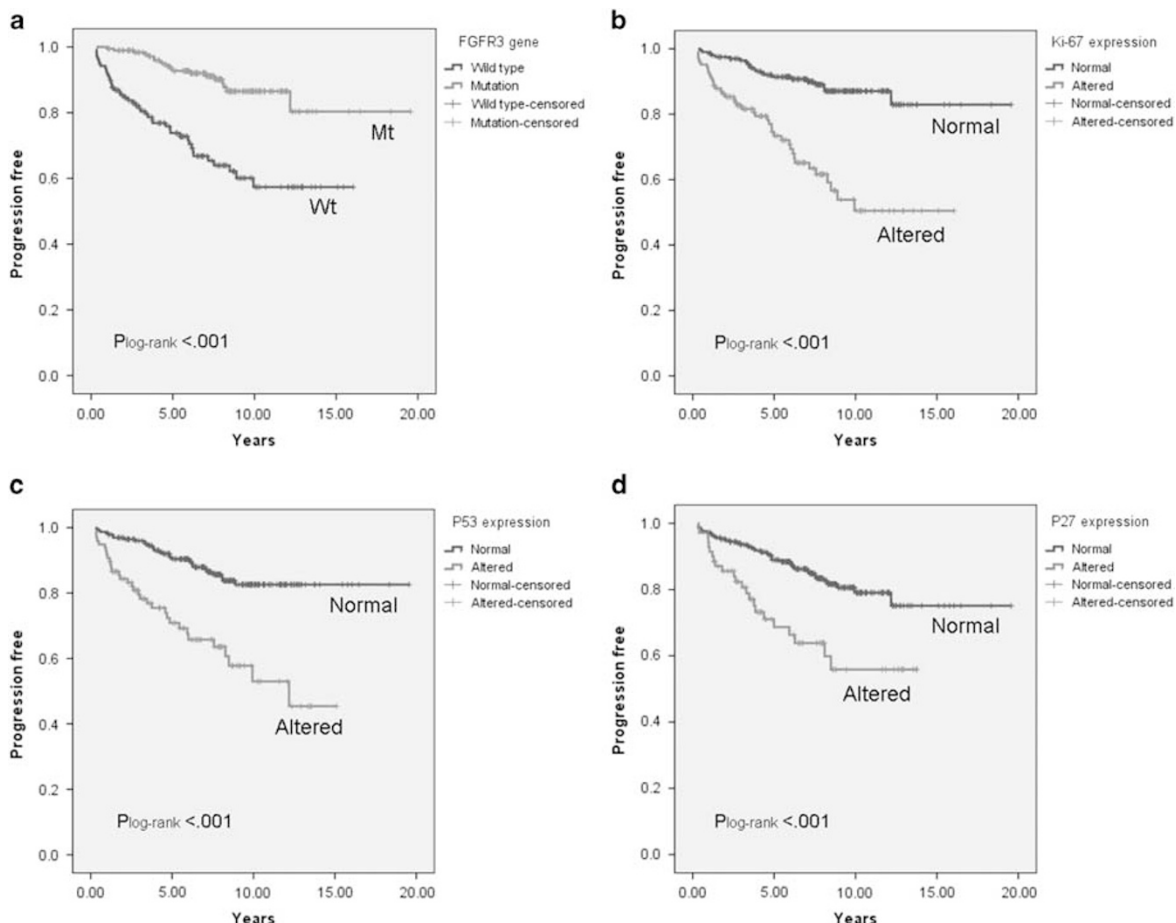


Figure 4 This figure shows the Kaplan–Meier curves providing clinical validation of the *FGFR3* gene status (a), Ki-67 expression (b), P53 expression (c) and P27 expression (d) to predict clinical progression of non-muscle invasive bladder cancer. The *P*-values (log-rank) were highly significant ($P < 0.001$) for the four molecular markers. Please note that *FGFR3* mutations were related to favorable disease whereas altered expression of Ki-67, P53 and P27 was associated with worse outcome. Ki-67, P53 and P27 expression was considered as altered if $> 25\%$, $> 10\%$ and $< 50\%$, respectively. *FGFR3*, fibroblast growth factor receptor 3 gene; mt, mutant; wt, wild type (no mutation).

Table 3 Molecular validation of the WHO 1973 and 2004 classification systems

	<i>FGFR3</i> (mt) (%)	<i>P</i> ^a	<i>Ki-67</i> > 25% (%)	<i>P</i> ^a	<i>P53</i> > 10% (%)	<i>P</i> ^a	<i>P27</i> < 50% (%)	<i>P</i> ^a	Total
Grade WHO 1973									
G1	78 (89)	<0.001	5 (6)	<0.001	2 (2)	<0.001	3 (3)	<0.001	88
G2	92 (62)		54 (36)		37 (25)		33 (22)		149
G3	17 (18)	<0.001	65 (75)	<0.001	58 (66)	<0.001	34 (40)	= 0.007	88
Grade WHO 2004									
LMP	67 (85)	= 0.438	5 (7)	= 0.005	2 (2)	= 0.005	3 (4)	= 0.004	79
LG	80 (78)		22 (23)		15 (15)		18 (17)		101
HG	40 (38)	<0.001	97 (65)	<0.001	80 (56)	<0.001	49 (35)	= 0.006	145
Total	187 (58)		124 (38)		97 (30)		70 (21)		325

Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; *FGFR3*, fibroblast growth factor receptor 3 gene; mt, mutant.
 Two *P*-values per molecular marker per grading system are given. The *P*-values refer to comparisons between the lowest/intermediate and the highest/intermediate grade. All but one comparisons were significant, ie, only the frequency of *FGFR3* mutations was equal in LMP and LG lesions (WHO 2004). The expression was considered altered if Ki-67 expression $> 25\%$, P53 expression $> 10\%$ and P27 expression $< 50\%$.
^aTwo-sided Fisher's exact test (%) refers to the percentage of the cancer-related molecular alteration per grade category.

Table 4A Molecular cross-validation of the WHO 2004 system with the WHO 1973 classification system

		WHO 1973												
		FGFR3			Ki-67			P53			P27			Total
		wt	mt	P ^a	≤25%	>25%	P ^a	≤10%	>10%	P ^a	≥50%	<50%	P ^a	
WHO 2004														
HG	G1-2	34	23	=0.008	25	32	=0.031	35	22	=0.002	42	15	=0.152	57
	G3	71	17		23	65		30	58		54	34		88
LG	G1	3	26	=0.114	28	1	=0.003	28	1	=0.061	28	1	=0.020	29
	G2-3	18	54		51	21		58	14		55	17		72

Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; *FGFR3*, fibroblast growth factor receptor 3 gene; mt, mutant; wt, wild type (no mutation).

The expression was considered altered if Ki-67 expression >25%, P53 expression >10% and P27 expression <50%.

The following cross-validations per category were based on Figure 1 and the results in Tables 2 and 3: The categories G1-2 vs G3 (WHO1973) were tested for the four molecular markers in HG (WHO2004). The categories G1 vs G2-3 (WHO1973) were tested for the four molecular markers in LG (WHO 2004).

^aTwo-sided Fisher's exact test.

Table 4B Molecular cross-validation of the WHO 1973 system with the WHO 2004 classification system

		WHO 2004												
		FGFR3			Ki-67			P53			P27			Total
		wt	mt	P ^a	≤25%	>25%	P ^a	≤10%	>10%	P ^a	≥50%	<50%	P ^a	
WHO 1973														
G1	LMP	5	51	=0.486	53	3	=0.999	55	1	=0.999	54	2	=0.999	56
	LG-HG	5	27		30	2		31	1		31	1		32
G2	LMP-LG	25	70	<0.001	72	23	<0.001	80	15	=0.001	77	18	=0.224	95
	HG	32	22		23	31		32	22		39	15		54

Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; *FGFR3*, fibroblast growth factor receptor 3 gene; mt, mutant; wt, wild type (no mutation).

The expression was considered altered if Ki-67 expression >25%, P53 expression >10% and P27 expression <50%.

The following cross-validations per category were based on Figure 1 and the results in Tables 2 and 3: The categories LMP vs LG-HG (WHO2004) were tested for the four molecular markers in G1 (WHO1973). The categories LMP-LG vs HG (WHO2004) were tested for the four molecular markers in G2 (WHO1973).

^aTwo-sided Fisher's exact test.

Clinical Cross-Validation: WHO 1973 and 2004

Analogous to the molecular cross-validation of the two WHO systems, we also performed clinical cross-validations. As expected, no differences were found for recurrence. For progression, the same cross-validations that were significant in molecular cross-validations were also clinically significant. Figure 5a–c shows that within HG, LG and G2, the corresponding grades from the other WHO classification behaved differently. No difference was found within the G1 category ($P_{\log\text{-rank}} = 0.186$). In conclusion, these clinical cross-validations confirmed a significant stepwise increase in progression along the line LMP/G1, LG, G2, HG and G3 as suspected from the molecular cross-validations.

WHO 1973 and 2004 in pTa and pT1

We also performed the clinical and molecular (cross) validations in pTa and pT1 non-muscle invasive bladder cancer. None of the pT1 tumors was classified as LMP or G1. Therefore, cross-validations in

pT1 were limited to the G2 and HG categories. In clinical validations, both WHO classifications were significant predictors for progression in pTa. In pT1, WHO1973 significantly predicted progression ($P_{\log\text{-rank}} = 0.027$) whereas WHO2004 did not ($P_{\log\text{-rank}} = 0.098$). In molecular validations for pTa and pT1, the same molecular markers (see Table 3) were significant. Molecular cross-validations showed significant differences for *FGFR3*, Ki-67 and P53 within the pT1-G2 category, for P53 within pT1-HG and for Ki-67 and P53 within pTa-LG and pTa-HG, respectively. Clinical cross-validations were significant for progression within pTa-LG ($P_{\log\text{-rank}} = 0.041$) and pTa-HG ($P_{\log\text{-rank}} = 0.043$), respectively. The subgroup analyses in pTa and pT1 confirmed the conclusions made in the paragraphs above.

Multivariable Analyses

Multivariable analyses were performed to compare the prognostic value of the different classification

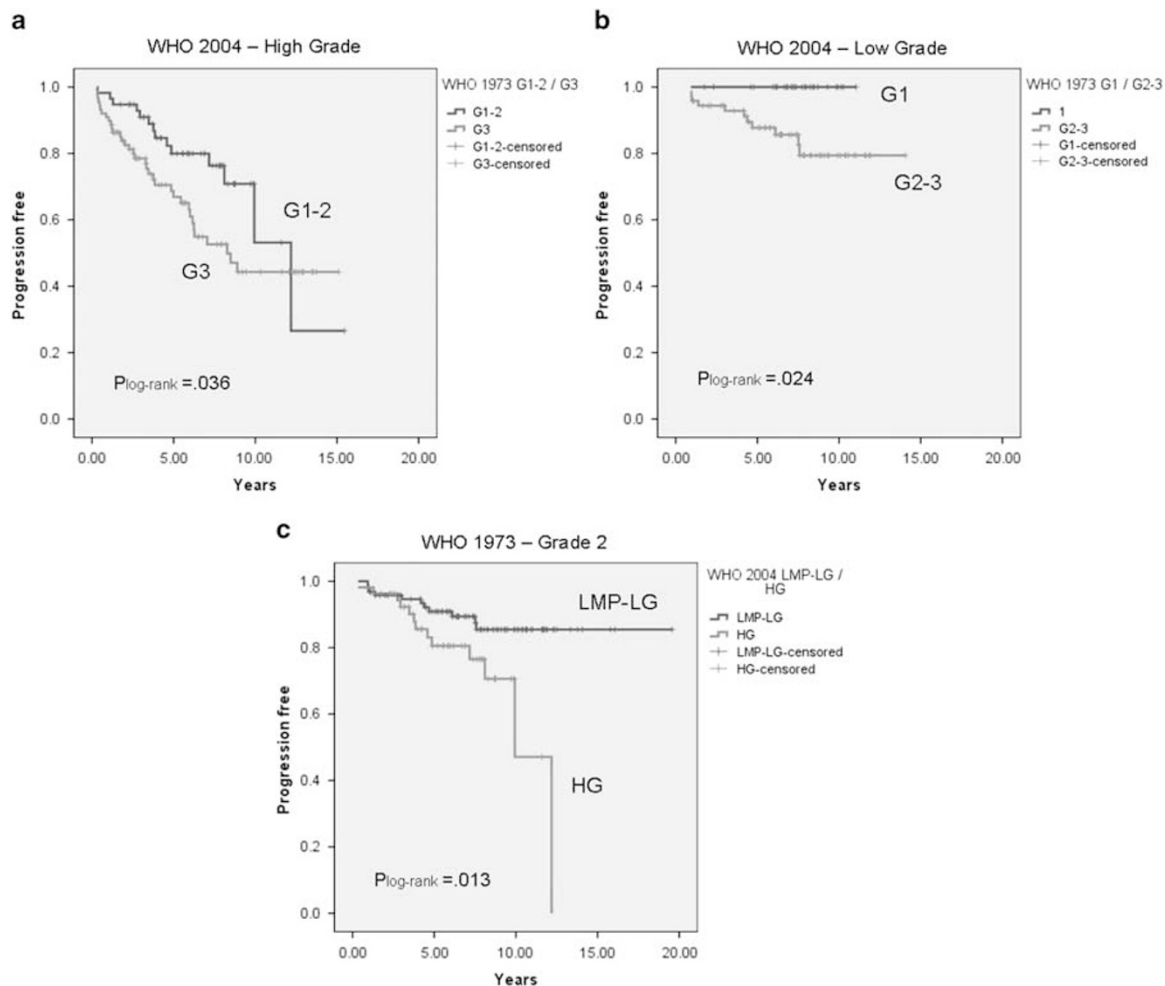


Figure 5 This figure shows the Kaplan–Meier curves providing clinical cross-validation of the WHO 1973 and 2004 classification systems for grade predicting progression. (a) The groups HG—G1/2 and HG—G3 behave differently ($P_{\log\text{-rank}} = 0.036$). (b) Significantly different progression rates for LG—G1 and LG—G2/3 ($P_{\log\text{-rank}} = 0.024$) are shown. (c) A significant difference in progression for G2—LMP/LG and G2—HG ($P_{\log\text{-rank}} = 0.013$) is shown. Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; G1, grade 1; etc.; WHO, World Health Organization.

systems for grade. For these analyses, the four-tiered system was tested against the two three-tiered systems. Within the proposed four-tiered system, the LMP/LG/G1, LG/G2, HG/G2 and HG/G3 categories comprised 85, 72, 54 and 88 patients, respectively (Table 2A). Twenty-six cases (23 LMP/G2 and 3 HG/G1) were excluded from the multivariable analyses because these could not be assigned to one of the four-tiered categories. Multiplicity, tumor size, CIS, grade (WHO1973, WHO2004 and combined four-tiered grading system) and stage were entered in the multivariable models. In univariable analyses for recurrence, multiplicity, tumor size, carcinoma *in situ*, Ki-67 and P53 were significant. Of these variables, multiplicity and tumor size were significant in multivariable analysis (Table 5A). In univariable analyses for progression, multiplicity, carcinoma *in situ*, grade WHO 1973, grade WHO2004, the combined 4-tiered grading system, stage, *FGFR3*, Ki-67, P53 and P27 were significant

(Figure 6). Of these variables, the combined four-tiered grading system and stage were significant in multivariable analysis (Table 5B). If only one of the classification systems for grade was entered in the multivariable analysis for progression, all systems (WHO1973, $P = 0.009$; WHO2004, $P = 0.022$; combined 4-tiered system, $P = 0.001$) were significant. The addition of the four molecular markers (*FGFR3*, Ki-67, P53 and P27) as separate variables to the multivariable analyses to predict recurrence and progression did not change the outcome of these analyses. In univariable analyses to predict stage (pTa/pT1), multiplicity, tumor size, carcinoma *in situ*, grade WHO1973, grade WHO2004, the combined four-tiered grading system, stage, *FGFR3*, Ki-67, P53 and P27 were significant. In the logistic regression models (with and without molecular markers as variables) to predict stage (pTa/pT1), the combined four-tiered grading system ($P < 0.001$) and P27 ($P = 0.028$) were significant (Table 5C).

Table 5 Significant variables, hazard ratios and 95% confidence intervals in the multivariate analyses for the prediction of recurrence (A) and progression (B) are shown. For the prediction of stage (pTa/pT1) (C), significant variables, odds ratios and 95% confidence intervals in the logistic regression analyses are shown

Recurrence		
(A)	P-value	HR (95% CI)
Multiplicity	= 0.001	1
Solitary		1.6 (1.2–2.1)
Multiple		
Size (cm)	= 0.015	1
≤ 3		1.4 (1.1–1.9)
> 3		
Progression		
(B)	P-value	HR (95% CI)
Four-tiered grading system	= 0.002	1
LMP/LG/G1		5.0 (1.4–18.0)
LG/G2		7.9 (2.2–28.5)
HG/G2		10.7 (3.0–37.6)
HG/G3		
Stage	= 0.043	1
pTa		1.8 (1.1–3.0)
pT1		
Prediction of stage pT1		
(C)	P-value	OR (95% CI)
Four-tiered grading system	< 0.001	1
LMP/LG/G1		1.7 (1.3–4.0)
LG/G2		4.5 (2.8–9.3)
HG/G2		8.9 (4.1–15.8)
HG/G3		
P27 expression	= 0.028	1
Normal		1.6 (1.2–2.1)
Altered		

Abbreviations: LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; CIS, carcinoma *in situ*; HR, hazard ratio; OR, odds ratio; CI, confidence interval.

The variables entered in the multivariable analyses were multiplicity, tumor size, the presence of carcinoma *in situ* (CIS), grade (WHO 1973, WHO2004 and combined four-tiered grading system) and stage. The addition of the four molecular markers (*FGFR3*, Ki-67, P53 and P27) as separate variables to the multivariable analyses did not change the outcome of these analyses except for the prediction of pT1.

Discussion

The WHO 2004 classification system was adopted to reduce observer variability and provide better prognostic information for non-muscle invasive bladder cancer patients. The main advantage of the WHO 2004 system over the WHO 1973 system is that histological criteria have been defined for each category.^{19,23} One of the criticisms to the new system is that it has been accepted in spite of lack of clinical evidence and proper studies to assess reproducibility and prognostic value.^{22–24} So far, studies on observer variability have shown no

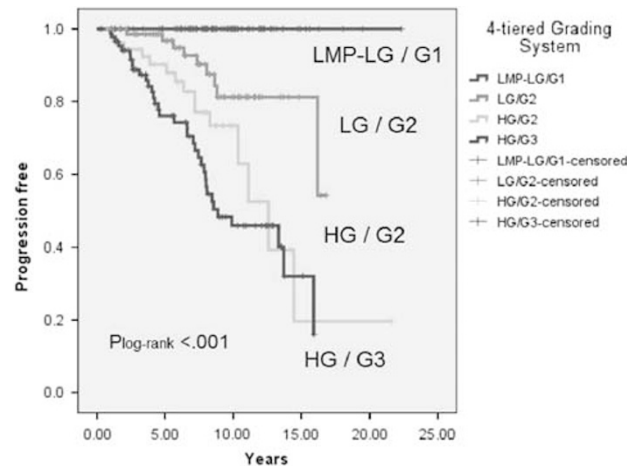


Figure 6 This figure shows the Kaplan–Meier curve for the four-tiered grading system based on the WHO 1973 and 2004 classification systems to predict clinical progression of non-muscle invasive bladder cancer. The *P*-value (log-rank) was highly significant ($P < 0.001$). Please note that 26 cases (23 LMP/G2 and 3 HG/G1) were excluded because these could not be assigned to one of the four-tiered categories. LMP, papillary urothelial neoplasia of low malignant potential; LG, low grade—papillary urothelial carcinoma; HG, high grade—papillary urothelial carcinoma; G1, grade 1; etc.; WHO, World Health Organization.

benefit of the WHO2004 grading system.^{25,26} Recurrence in LMP, which is considered as a benign lesion, varied between 25 and 60% in 11 studies²³ indicating that LMP requires surveillance as low-risk non-muscle invasive bladder cancer.²⁷ Our study with long follow-up also showed no significant difference in recurrence between LMP and LG or HG non-muscle invasive bladder cancer. Our smaller sample size and the notion that grade is a better predictor for progression than recurrence may explain the absence of significance for recurrence in our study.^{6,23,25,28} Previously, progression was found more frequently in G3 compared with HG as the HG group is more heterogeneous.²³ We confirmed this in a larger series with one pathologist reviewing both WHO classifications. Otto *et al*²⁹ specifically analyzed the two WHO classifications in 310 patients with primary pT1 bladder cancer and found that the WHO 1973 system was superior to the WHO2004 classification. Holmäng *et al*³⁰ re-graded 255 pTaG1 into LMP and LG. Their number of progressive cases ($N = 6$) was too low to find a significant difference. The three-tiered WHO2004 system is therefore not superior to the WHO1973 classification from clinical point of view.^{14,23,25,29} Consequently, management and/or follow-up schemes of non-muscle invasive bladder cancer patients have not changed.^{25,27} One of the solutions may be a classification system for grade that includes elements from both WHO classification systems. Our study using clinical cross-validation confirms this concept as we found a significant stepwise increase in progression along the line LMP/G1, LG, G2, HG and G3.

Next to clinical comparisons, we also validated both WHO systems with molecular markers. Like for grade, molecular markers were more indicative for progression than for recurrence. As expected, we found that the frequency of *FGFR3* mutations decreased with higher grades and altered expression of Ki-67, P53 and P27 increased with higher grades. Analogous to the clinical cross-validations, the cross-validations using the four molecular markers confirmed a stepwise increase in biological aggressiveness along the line LMP/G1, LG, G2, HG and G3. Only few studies that compare molecular markers with progression in both WHO1973 and 2004 classifications are available.^{14,31} In general, *FGFR3*, Ki-67, P53 and CK20 were found to predict progression in both WHO classifications.^{14,31} *FGFR3* was also significant in the pT1 and HG subgroups¹⁴ and P53 in G2 non-muscle invasive bladder cancer.³¹ A systematic molecular cross-validation was not reported in these studies.

Considering studies that reported on molecular or clinical (progression) validation using WHO1973 and 2004,^{5,14,23,25,29–32} it is striking that the WHO1973 grade was only reviewed in one previous study.²⁹ We reported that the pathologist's mean grade is constant (≤ 0.1 grade-point) in both WHO classifications whereas differences in mean grade among four pathologists were as high as 0.7 grade point.²⁶ Of note, the mean grade of the WHO2004 classification was 0.4 grade point higher than the mean grade of the WHO1973 classification. Moreover, progression in each grade category highly depended on the pathologist's mean grade.²⁶ This indicates that (central) pathology review by one pathologist is crucial if both WHO systems are compared with each other in one study. This may also be the most important reason that previous studies have not systematically reported on cross-validation as we did. Therefore, our study provides unique and original molecular and clinical data on both current WHO classifications for grade.

For clinicians, the G2 category in WHO1973 and the HG category in WHO2004 are troublesome. Although earlier efforts to subdivide the WHO1973 G2 NMI-BCs into G2a and G2b resulted in useful prognostic information in a four-tiered classification,³³ this subdivision did not gain widespread acceptance among pathologists. The WHO1999 classification subdivided HG into two separate categories⁸ but was also not accepted or universally used¹⁹ despite a recent editorial by Liedberg *et al*³⁴ who made a plea using clinical and molecular data to distinguish between G2 and G3 as proposed in the WHO1999 classification. On the basis of molecular biology and clinical/multivariable data, our results support a four-tiered grading system using the 1973 and 2004 WHO classifications with one low-grade (LMP/LG/G1) category which includes LMP, two intermediate grade (LG/G2 and HG/G2) categories and one high-grade (HG/G3) category. In clinical practice and to avoid outliers

like LMP/G2, we recommend to first grade a non-muscle invasive bladder cancer according to WHO2004 and subsequently according to WHO1973. All LMP should be graded as G1, the LG carcinomas as G1 or G2 and the HG carcinomas as G2 or G3.

The clinical recommendations in guidelines based on three-tiered grading systems are currently not clear for all risk groups.^{3–5} The proposed four-tiered system may help to improve treatment strategies for the G2 category in WHO1973 and the HG category in WHO2004. For example, LMP/LG/G1 may be managed by TUR and a single instillation of chemotherapy, LG/G2 by additional intravesical instillations of chemotherapy or BCG, HG/G2 by additional intravesical instillations of BCG and in case of HG/G3, early radical cystectomy may be considered. In this context, it is also important if the pathologist is a high or low non-muscle invasive bladder cancer grader (high or low 'mean' grade).²⁶ However, the four-tiered system may also suffer from reduced reproducibility compared with the three-tiered grading system. It is therefore important that pathologists and urologists closely work together to ensure optimal outcomes for non-muscle invasive bladder cancer patients.

Acknowledgments

We would like to thank Simone Russell, Rati Vajpeyi, Sally Hanna, Roni Sambas, Cynthia Kuk and Gina Lockwood for help and advice. The medical-ethical committees of the Erasmus MC, Rotterdam (MEC 168.922/1998/55) and the University Health Network, Toronto (02-0515-C and 08-0263-T) gave approval. Financial support was given by the University of Toronto and by a grant from the Dutch Cancer Society—KWF Kanker-bestrijding.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

- 1 Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013;63:11–30.
- 2 Sobin DH, Wittekind Cheds. *TNM Classification of Malignant Tumours*, 6th edn. Wiley-Liss: New York; 2002, pp 199–202.
- 3 Hall MC, Chang SS, Dalbagni G, *et al*. Guideline for the management of non-muscle invasive bladder cancer (stages Ta, T1, and Tis): 2007 update. *J Urol* 2007;178: 2314–2330.
- 4 Babjuk M, Burger M, Zigeuner R, *et al*. EAU guidelines on non-muscle-invasive urothelial carcinoma of the bladder: update 2013. *Eur Urol* 2013;64:639–653.
- 5 van Rhijn BW, Burger M, Lotan Y, *et al*. Recurrence and progression of disease in non-muscle-invasive bladder cancer: from epidemiology to treatment strategy. *Eur Urol* 2009;56:430–442.

- 6 Sylvester RJ, van der Meijden AP, Oosterlinck W, *et al*. Predicting recurrence and progression in individual patients with stage Ta T1 bladder cancer using EORTC risk tables: a combined analysis of 2596 patients from seven EORTC trials. *Eur Urol* 2006;49:466–477.
- 7 Fernandez-Gomez J, Solsona E, Unda M, *et al*. Prognostic factors in patients with non-muscle-invasive bladder cancer treated with bacillus Calmette-Guérin: multivariate analysis of data from four randomized CUETO trials. *Eur Urol* 2008;53:992–1001.
- 8 Busch C, Algaba F. The WHO/ISUP 1998 and WHO 1999 systems for malignancy grading of bladder cancer. Scientific foundation and translation to one another and previous systems. *Virchows Arch* 2002;441:105–108.
- 9 Vajo Z, Francomano CA, Wilkin DJ. The molecular and genetic basis of fibroblast growth factor receptor 3 disorders: The achondroplasia family of skeletal dysplasias, Muenke craniosynostosis and Crouzon syndrome with acanthosis nigricans. *Endocr Rev* 2000;21:23–39.
- 10 Cappellen D, De Oliveira C, Ricol D, *et al*. Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nat Genet* 1999;23:18–20.
- 11 Billerey C, Chopin DK, Aubriot-Lorton MH, *et al*. Frequent FGFR3 mutations in papillary non-invasive bladder (pTa) tumors. *Am J Pathol* 2001;158:1955–1959.
- 12 van Rhijn BW, Vis AN, van der Kwast TH, *et al*. Molecular grading of urothelial cell carcinoma with fibroblast growth factor receptor 3 and MIB-1 is superior to pathological grade for the prediction of clinical outcome. *J Clin Oncol* 2003;21:1912–1921.
- 13 Hernández S, López-Knowles E, Lloreta J, *et al*. Prospective study of FGFR3 mutations as a prognostic factor in nonmuscle invasive urothelial bladder carcinomas. *J Clin Oncol* 2006;24:3664–3671.
- 14 Burger M, van der Aa MN, van Oers JM, *et al*. Prediction of progression of non-muscle-invasive bladder cancer by WHO 1973 and 2004 grading and by FGFR3 mutation status: a prospective study. *Eur Urol* 2008;54:835–843.
- 15 van Rhijn BW, Zuiverloon TCM, Vis AN, *et al*. Molecular Grade (FGFR3/MIB-1) and EORTC risk-scores are predictive in primary non-muscle invasive bladder cancer. *Eur Urol* 2010;58:433–441.
- 16 van Rhijn BW, Lurkin I, Chopin DK, *et al*. Combined microsatellite and FGFR3 mutation analysis enables a highly sensitive detection of urothelial cell carcinoma in voided urine. *Clin Cancer Res* 2003;9:257–263.
- 17 Junker K, van Oers JM, Zwarthoff EC, *et al*. Fibroblast growth factor receptor 3 mutations in bladder tumors correlate with low frequency of chromosome alterations. *Neoplasia* 2008;10:1–7.
- 18 Shariat SF, Ashfaq R, Sagalowsky AI, *et al*. Predictive value of cell cycle biomarkers in non-muscle invasive bladder transitional cell carcinoma. *J Urol* 2007;177:481–487.
- 19 Eble J, Sauter G, Epstein J, *et al*. Pathology and Genetics of Tumours of the Urinary System and Male Genital Organs. IARC Press: Lyon, 2004.
- 20 Mostofi FK, Sobin LH, Torloni H. Histological typing of urinary bladder tumours. International Classification of Tumours, No. 10. World Health Organization: Geneva, 1973;21–31.
- 21 Epstein JI, Amin MB, Reuter VR, *et al*. The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder. *Am J Surg Pathol* 1998;22:1435–1448.
- 22 Lopez-Beltran A, Montironi R. Non-invasive urothelial neoplasms: according to the most recent WHO classification. *Eur Urol* 2004;46:170–176.
- 23 Hofmann T, Knüchel-Clarke R, Hartmann A, *et al*. Clinical implications of the 2004 WHO histological classification on non-invasive tumours of the urinary bladder. *EAU-EBU Update Series* 2006;4:83–95.
- 24 Harnden P. A critical appraisal of the classification of urothelial tumours: time for a review of the evidence and a radical change? *BJU Int* 2006;99:723–725.
- 25 MacLennan GT, Kirkali Z, Cheng L. Histologic grading of noninvasive papillary urothelial neoplasms. *Eur Urol* 2007;51:889–897.
- 26 van Rhijn BW, van Leenders GJLH, Ooms ECM, *et al*. The pathologist's mean grade is constant and individualizes the prognostic value of bladder cancer grading. *Eur Urol* 2010;57:1052–1057.
- 27 Montironi R, Lopez-Beltran A, Scarpelli M, *et al*. (2004) World Health Organization classification of the noninvasive urothelial neoplasms: inherent problems and clinical reflections. *Eur Urol Suppl* 2009;8:453–457.
- 28 Pan CC, Chang YH, Chen KK, *et al*. Prognostic significance of the 2004 WHO/ISUP classification for prediction of recurrence, progression, and cancer-specific mortality of non-muscle-invasive urothelial tumors of the urinary bladder: a clinicopathologic study of 1,515 cases. *Am J Clin Pathol* 2010;133:788–795.
- 29 Otto W, Denzinger S, Fritsche HM, *et al*. The WHO classification of 1973 is more suitable than the WHO classification of 2004 for predicting survival in pT1 urothelial bladder cancer. *BJU Int* 2011;107:404–408.
- 30 Holmäng S, Hedelin H, Anderström C, *et al*. Recurrence and progression in low grade papillary urothelial tumors. *J Urol* 1999;162:702–707.
- 31 Holmäng S, Andius P, Hedelin H, *et al*. Stage progression in Ta papillary urothelial tumors: relationship to grade, immunohistochemical expression of tumor markers, mitotic frequency and DNA ploidy. *J Urol* 2001;165:1124–1128.
- 32 Cao D, Vollmer RT, Luly J, *et al*. Comparison of 2004 and 1973 World Health Organization grading systems and their relationship to pathologic staging for predicting long-term prognosis in patients with urothelial carcinoma. *Urology* 2010;76:593–599.
- 33 Malmström P-U, Busch C, Norlen BJ. Recurrence, progression and survival in bladder cancer. A retrospective analysis of 232 patients with greater than or equal to 5-year follow-up. *Scand J Urol Nephrol* 1987;21:185–195.
- 34 Liedberg F, Lauss M, Patschan O, *et al*. The importance of being grade 3: WHO 1999 versus WHO 2004 pathologic grading. *Eur Urol* 2012;62:620–623.