



Interobserver reproducibility of a hybrid three-tier grading system of papillary nonmuscle invasive urothelial carcinoma: an international UroPathology study

Michelle R. Downes,^{1,2}  Katherine Lajkosz,³  Ferran Algaba,⁴ Yves Allory,⁵ 
 Mahul B. Amin,^{6,7}  Liang Cheng,^{8,9}  Sangeeta Desai,¹⁰  Markus Eckstein,^{11,12,13} 
 Arndt Hartmann,^{11,12,13}  Antonio Lopez-Beltran,¹⁴  Sebastian Mannweiler,¹⁵
 Maurits L. van Montfoort,¹⁶  Jon Oxley,¹⁷  Gladell P. Paner,¹⁸ Hemamali Samaratunga,¹⁹ 
 Steven Shen,²⁰ Joshua Warrick,²¹  Sean R. Williamson,²²  Sara E. Wobker²³  &
 Theodorus H. van der Kwast²⁴ 

¹Division of Anatomic Pathology, Department of Laboratory Medicine and Molecular Diagnostics, Precision Diagnostics and Therapeutics Program, Sunnybrook Health Sciences Centre, ²Laboratory Medicine and Pathobiology, University of Toronto, ³Department of Biostatistics, Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada, ⁴Pathology Department, Fundació Puigvert, Barcelona, Spain, ⁵Department of Pathology, Institut Curie, Paris, France, ⁶Department of Pathology and Laboratory Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, ⁷Department of Urology, USC Keck School of Medicine, Los Angeles, California, ⁸Department of Pathology and Laboratory Medicine, Brown University Warren Alpert Medical School, ⁹The Legorreta Cancer Center at Brown University, Providence, Rhode Island, USA, ¹⁰Department of Pathology, Tata Memorial Centre, Mumbai, India, ¹¹Institute of Pathology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), ¹²Comprehensive Cancer Center EMN, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, ¹³Bavarian Center for Cancer Research (BZKF), Erlangen, Germany, ¹⁴Unit of Anatomical Pathology, Faculty of Medicine, Cordoba University, Cordoba, Spain, ¹⁵Institute of Pathology, Medical University of Graz, Graz, Austria, ¹⁶Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands, ¹⁷Department of Cellular Pathology, North Bristol NHS Trust, Bristol, UK, ¹⁸Departments of Pathology and Surgery, Section of Urology, University of Chicago, Chicago, Illinois, USA, ¹⁹Department of Pathology, Aquesta UroPathology and University of Queensland, Brisbane, Queensland, Australia, ²⁰Department of Pathology Genomic Medicine, Houston Methodist Hospital and Weill Cornell Medical College, Houston, Texas, ²¹Departments of Pathology and Urology, Penn State College of Medicine, Hershey, Pennsylvania, ²²Pathology and Laboratory Medicine Institute, The Cleveland Clinic, Cleveland, Ohio, ²³Lineberger Comprehensive Cancer Center and Department of Pathology and Laboratory Medicine University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA and ²⁴Laboratory Medicine Program, Princess Margaret Cancer Center, University Health Network, Toronto, Ontario, Canada

Date of submission 27 November 2024

Accepted for publication 15 February 2025

Downes M R, Lajkosz K, Algaba F, Allory Y, Amin M B, Cheng L, Desai S, Eckstein M, Hartmann A, Lopez-Beltran A, Mannweiler S, van Montfoort M L, Oxley J, Paner G P, Samaratunga H, Shen S, Warrick J, Williamson S R, Wobker S E & van der Kwast T H

(2025) *Histopathology* 87, 649–659. <https://doi.org/10.1111/his.15438>

Interobserver reproducibility of a hybrid three-tier grading system of papillary nonmuscle invasive urothelial carcinoma: an international UroPathology study

Abbreviations: AUA, American Urological Association; CIS, carcinoma in situ; EAU, European Association of Urology; H&E, haematoxylin and eosin; HG, high grade; ISUP, International Society of Urological Pathology; LG, low grade; LIS, laboratory information systems; NMIBC, nonmuscle invasive bladder cancer; PUNLMP, papillary urothelial neoplasm of low malignant potential; WHO, World Health Organization.

Aims: A hybrid-three tier system with low grade (LG), high grade- G2 (HG-G2), high grade- G3 (HG-G3) has been proposed in recognition of, and to help address, the clinical heterogeneity within high grade WHO 2004/2022. We assessed interobserver reproducibility amongst international uropathologists using this three-tier approach.

Methods and Results: Papillary Ta nonmuscle invasive bladder cancer (NMIBC) specimens ($n = 30$) were selected and graded by two uropathologists and assessed using WHO 2004/2022 and WHO 1973 and categorized as LG ($n = 15$), HG-G2 ($n = 8$), HG-G3 ($n = 7$), and photographed at 10 $\times$ and 20 $\times$ magnification. Images were circulated via Survey Monkey to invited uropathologists who determined: (1) that image was LG or HG, and (2) if HG, assigned to G2 or G3. Model-based kappa measure of association was used to assess interrater agreement. Eighteen uropathologists (eight North

American, eight European, two other) assessed 60 images with 1076 gradings for analysis. The kappa value amongst Europeans versus North Americans was 0.663 versus 0.647 for 10 $\times$ images and 0.682 versus 0.623 for 20 $\times$ images. At 10 $\times$, agreement for LG, HG-G2, and HG-G3 was 74.6%, 63.6%, and 92.0%, and at 20 $\times$ was 64.3%, 63.9%, and 95.2% respectively.

Conclusion: Three-tier grading of papillary Ta NMIBC had substantial interobserver agreement amongst international uropathologists. The recognition of the HG-G3 case reached the highest concordance. North American uropathologists had comparable kappa scores (substantial agreement) to Europeans, despite being unaccustomed to separating HG cases into G2 and G3, demonstrating three-tier grading could be “quickly” adopted by genitourinary experts if endorsed and required by the relevant bodies in their jurisdiction of practice.

Keywords: bladder cancer, hybrid grading, interobserver variation, nonmuscle invasive, papillary

Introduction

Globally, bladder cancer is recognized as the tenth most common malignancy and sixth most frequent in men, with approximately three-quarters of patients having nonmuscle invasive bladder cancer (NMIBC) at initial presentation.¹ The term NMIBC encompasses three distinct pathologic entities, namely, carcinoma *in situ* (CIS, Tis), papillary noninvasive carcinoma (Ta disease), and carcinoma (papillary and nonpapillary) with invasion limited to the lamina propria (T1 disease). Grading of papillary NMIBC is useful in determining a patient's risk of progression^{2,3} and is a key prognostic variable that is incorporated into various risk stratification tools.^{4,5}

There are two World Health Organization (WHO) grading schemes that are widely used for papillary NMIBC: the original two-tier WHO 2004 system (now known as WHO 2004/2022)⁶ and the three-tier WHO 1973 system.⁷ The European Association of Urology (EAU) guidelines recommend grading with both WHO 2004/2022 and WHO 1973,⁵ while the American Urological Association (AUA) guidelines use the WHO 2004/2022 grading alone.⁴ Two published studies have shown that utilizing a hybrid three-tier grading system^{2,3} is a better prognosticator for determining progression to muscle invasion and beyond ($\geq T2$).

There are limited original studies assessing interobserver reproducibility in assigning grade to papillary NMIBC⁸ and the published studies have been variable in terms of design, types of cases assessed, and number of cases included. The interobserver reproducibility assessed by kappa statistics is slightly higher for WHO 2004/2022 than WHO 1973 and improves when the category of papillary urothelial neoplasm of low malignant potential (PUNLMP) is excluded from analysis. Nevertheless, reproducibility is at best moderate, irrespective of grading scheme.⁸ Intraobserver reproducibility is similar for both WHO 1973 and WHO 2004/2022.

In September 2022, a consensus meeting on bladder pathology was organized by the International Society of Urological Pathology (ISUP) in Basel, Switzerland, which was informed by premeeting surveys of the ISUP membership and the EAU membership.⁹ Working group 1 reviewed bladder cancer grading system performance, and based on the premeeting surveys of both memberships and the in-conference voting, there was a preference to refine grading of papillary NMIBC into a three-tier system.^{8,10} As such, the existing category of “low grade” (LG) in WHO 2004/2022 would be retained and the “high grade” (HG) WHO 2004/2022 would be divided into high grade- grade 2 (HG-G2) and

high grade- grade 3 (HG-G3) in line with the WHO 1973 system.

The main purpose of this current study was to explore the interobserver reproducibility of such a three-tier hybrid grading scheme amongst a group of international uropathologists. A secondary aim was to determine whether the magnification, either 10× or 20×, used in grading cases had an impact on the grade assigned.

Methods

The laboratory information systems (LIS) of Sunnybrook Health Sciences Centre and University Health Network, two academic institutions in Toronto, Canada, were searched for papillary, NMIBC diagnosed at biopsy or transurethral resection between the years 2003–2022. Research Ethics Board approval was granted at both centres; REB 187-2016 and CAPCR 20-5817, respectively.

The criteria for inclusion were: papillary carcinoma, staging category Ta (noninvasive) with minimal cautery artefact and variability in haematoxylin and eosin (H&E) staining. Each case was reviewed in a consensus manner by two uropathologists (M.R.D., T.v.D.K.) working cooperatively to determine cases to be selected for study inclusion. The cases were selected to represent a spectrum of diagnostic difficulties from those that were deemed straightforward to others that were more challenging. Each case was assessed using two grading schemes: WHO 2004/2022 (cases designated as LG, HG), and a hybrid three-tier grading scheme (cases designated as LG, HG-G2, or HG-G3). For each case, a representative area was annotated and then photographed at 10× and 20× magnification using an Olympus BX53 microscope (Tokyo, Japan) with mounted Leica DFC320 camera (Wetzlar, Germany). The areas that were selected for image capture had to be free of cautery artefact, have reasonable quality H&E, with good nuclear and cytoplasmic staining, and minimal histologic artefacts. Each selected area of image capture was felt to be representative of the overall grade assigned at prior case sign out.

Each image was assigned a random number from 1 to 60 and imported as JPEG images into Survey Monkey (www.surveymonkey.com, California, USA) to create 60 unique survey questions. For each image, there were two questions: 1- is the carcinoma HG or LG using WHO 2004/2022, and 2- if HG is it a HG-G2 or a HG-G3?

Eighteen international uropathologists with expertise in bladder pathology, (FA, YA, MBA, LC, SD, ME, AH, AL-B, SM, MLvM, JO, GPP, HS, JW, SRW, SEW,

and one North American pathologist who requested to remain anonymous) were invited to participate in the study. They were provided with the following instructions: "Images are a mix of 10× and 20× objective magnification." (Participants were not informed that the same cases were photographed at both 10× and 20×.) The cases had been collected from different institutions, and therefore they should expect some variability in H&E staining. When assessing the HG cases, (they were instructed) to use the criteria of HG-G3 being a carcinoma at the far end of the grading spectrum, based on disorder, marked variation in nuclear size, irregular nuclear contours, severity of nuclear atypia, nuclear hyperchromasia, and mitotic activity (Table 1). Each individual survey

Table 1. Architectural and cytological features used to assign grade in each listed grading scheme

Grading scheme	Architectural and cytological features
WHO 2004	
Low grade	Delicate papillae, orderly cohesive cells with enlarged oval nuclei showing some variability in nuclear size/shape. Infrequent, basal mitoses.
High grade	Complex papillae with fusion and branching showing disordered, crowded cells with loss of polarity. Nuclear pleomorphism with prominent nucleoli and frequent, multi-level mitoses.
Revised WHO 1973	
Grade 1	Ordered cells with mild nuclear variation, absence of hyperchromasia, maturation to umbrella cell layer and minimal, basally located mitoses
Grade 2	Not 1 or 3
Grade 3	Variable polarity with disordered layers, hyperchromasia, marked variation in nuclear size, absent umbrella cells and prominent mitotic activity
Hybrid grade	
Low grade	Delicate papillae, orderly cohesive cells with enlarge oval nuclei showing some variability in nuclear size/shape. Infrequent, basal mitoses.
High grade-G2	Disordered polarity with some nuclear variability and prominent nucleoli, absence of large, hyperchromatic pleomorphic nuclei, prominent mitotic activity
High grade- G3	Variable polarity with disordered layers, hyperchromasia, marked variation in nuclear size, absent umbrella cells and prominent mitotic activity

also included three general questions at the start of the survey, which were: 1- location of practice (North America, Europe, or other), 2- what grading system they used in daily practice for papillary NMIBC (WHO 2004/2022, WHO 1973 or other), and 3- what magnification they preferred for grading (10×, 20×, both or other). The survey remained open for 2 weeks. The responses were downloaded into excel file format.

Model-based kappa (κ) measures of association were used to assess interrater agreement.¹¹ The κ estimates and their 95% confidence intervals (CIs) were stratified by magnification (10× vs. 20×) and practice location (North America vs. Europe). Categories for kappa agreement were as follow: 0–0.20 indicate slight agreement, 0.21–0.40 indicate fair agreement, 0.41–0.60 indicate moderate agreement, 0.61–0.80 indicate substantial agreement, and 0.81–1 indicate almost perfect agreement. See Table S1 for variance and intraclass coefficient estimates from models used in kappa estimation. Statistical analyses were conducted using R v. 4.3.0 (Vienna, Austria).

Results

There were 18 survey respondents: eight North American, eight European, one Australian, and one Indian pathologist. With respect to daily practice, 10 respondents (55.6%) used WHO 2004/2022 grading and eight (44.4%) used both WHO 2004/2022 and WHO 1973 for reporting papillary NMIBC cases. All the European participants used dual/hybrid grading in their daily practice. The majority reported using both 10× and 20× magnification when grading ($n = 11$, 61.1%), while four (22.2%) used 10× alone, two (11.1%) used 20× alone, and one (5.6%) pathologist used 4× and 10× in their daily practice. One survey respondent skipped two questions and two survey respondents each skipped a single question, leaving 1076/1080 (99.6%) image grading responses for assessment.

There were 30 cases selected, generating 60 images in total. The breakdown of the original assigned grades was as follows: LG, $n = 15$ (30 images, 50.0%), HG-G2, $n = 8$ (16 images, 26.7%), and HG-G3, $n = 7$ (14 images, 23.3%). Table 2 shows the distribution of grade assigned amongst the 1076 responses.

Figures 1–3 show examples of cases with the high-level agreement for LG, HG-G2, and HG-G3. Figure 4 shows examples of low agreement for LG and Figures 5 and 6 for HG-G2 bladder cancers.

Table 2. Responses for each of the 30 cases/60 images ($n = 1076$ gradings)

Study ID	Magnification	LG	HG G2	HG G3
1	10×	15	3	0
	20×	14	4	0
2	10×	14	3	0
	20×	16	2	0
3	10×	14	4	0
	20×	10	8	0
4	10×	4	14	0
	20×	7	10	1
5	10×	12	6	0
	20×	9	9	0
6	10×	18	0	0
	20×	16	2	0
7	10×	6	12	0
	20×	5	13	0
8	10×	17	0	0
	20×	17	1	0
9	10×	15	3	0
	20×	16	2	0
10	10×	18	0	0
	20×	17	1	0
11	10×	8	10	0
	20×	6	12	0
12	10×	13	4	0
	20×	7	11	0
13	10×	16	2	0
	20×	11	7	0
14	10×	18	0	0
	20×	17	1	0
15	10×	12	6	0
	20×	5	12	0
16	10×	2	14	2
	20×	1	16	1

Continued

Table 2. (Continued)

Study ID	Magnification	LG	HG G2	HG G3
17	10×	6	12	0
	20×	3	12	1
18	10×	13	5	0
	20×	7	11	0
19	10×	0	6	12
	20×	0	5	13
20	10×	3	13	1
	20×	4	13	1
21	10×	1	16	1
	20×	0	13	5
22	10×	1	11	6
	20×	1	14	3
23	10×	1	14	2
	20×	0	8	10
24	10×	0	2	16
	20×	0	0	18
25	10×	0	3	15
	20×	0	0	18
26	10×	0	3	15
	20×	0	1	17
27	10×	0	1	17
	20×	0	4	14
28	10×	0	0	18
	20×	0	1	17
29	10×	0	1	17
	20×	0	0	18
30	10×	0	0	18
	20×	0	0	18

Assigned grade.

%, percentage; G2, grade 2; G3, grade 3; HG, high grade; LG, low grade; ×, magnification.

The κ value for the 10× images ($n = 537$) was 0.641 (95% CI: 0.557–0.725) and 20× images ($n = 539$) was 0.623 (95% CI: 0.533–0.714), which is substantial agreement (Table 3) with both North

American and European pathologists achieving substantial agreement values with North American κ for 10× images = 0.647 (95% CI: 0.552–0.741) and for 20× images = 0.623 ([95% CI: 0.518–0.728]) The European pathologists' κ for 10× images was = 0.663 (95% CI: 0.585–0.741) and for 20× images was 0.682 (95% CI: 0.593–0.772). The lowest observed κ was for assessment of LG 20× images at $\kappa = 0.241$ (95% CI: 0.112–0.369), which represented fair agreement, whereas the HG images at 20× had $\kappa = 0.549$ (95% CI: 0.417–0.680) (moderate agreement). Kappa could not be estimated separately for HG-G2 and HG-G3 cases, as all HG-G3 cases were classified as either HG-G2 (16/252) or HG-G3 (236/252).

Discussion

Papillary NMIBC tumours show variability in terms of recurrence and progression. Grade is not a significant factor in determining recurrence; however, grade is highly significant in predicting progression ($\geq T2$).^{12,13} Irrespective of the grading system used, both WHO 2004/2022 and WHO 1973 show differences in progression within the carcinoma categories, with LG progressing less than HG (WHO 2004/2022), while progression rates in G3>G2>G1 using WHO 1973. The issue of which system performs better in determining progression was addressed in previous publications,^{2,12,14} which found that a hybrid approach, combining WHO 2004/2022 and WHO 1973 outperformed either system alone. An independent study from North America also showed hybrid grading to be preferential in determining progression, particularly in the HG group, where there was a significant difference between HG-G2 and HG-G3 and less of a difference in the LG-G1 and LG-G2.³ This suggested that a three-tier hybrid grading scheme retaining LG as one category and dividing HG into two categories may be sufficient. The WHO 2004/2022 HG group is a clinically and molecularly^{15–17} heterogeneous group, with those at the extreme HG (G3 in WHO 1973) showing faster and higher progression rates than those HG cases that correlate with G2 in WHO 1973.¹⁸ Therefore, there is merit in subclassifying these HG cases to inform management decisions and ensure appropriate intervention.

The major criticism of WHO 1973 has rightly been the lack of histologic criteria to enable accurate classification of cases. At the Basel ISUP Consensus Meeting, there was general agreement amongst participants that all existing grading schemes could

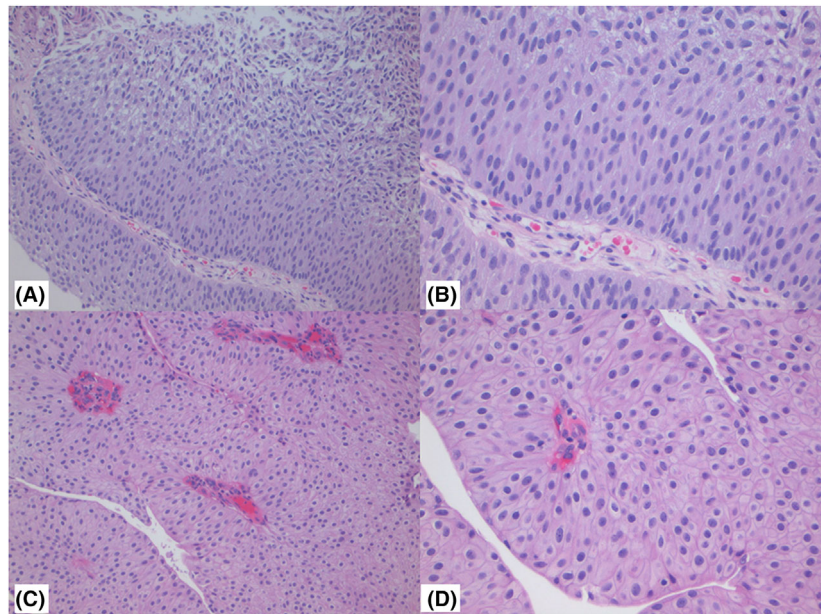


Figure 1. Low grade papillary urothelial carcinomas with highest agreement on both 10 $\times$ and 20 $\times$ images. (A) A case at 10 $\times$ and (B), the same case at 20 $\times$ magnification. All participants (100%) agreed with low grade on 10 $\times$. One participant graded as HG-G2 on 20 $\times$. The images show orderly arrangement of cells that maintain polarity and show mild variation in nuclear size with fine chromatin. No mitoses are seen in the images. (C) A second case at 10 $\times$ and (D), the same case at 20 $\times$ magnification. All participants (100%) agreed with low grade on 10 $\times$. One participant graded as HG-G2 on 20 $\times$. Similar histologic features are seen as in parts (A and B). Note the difference in H&E staining between the two cases.

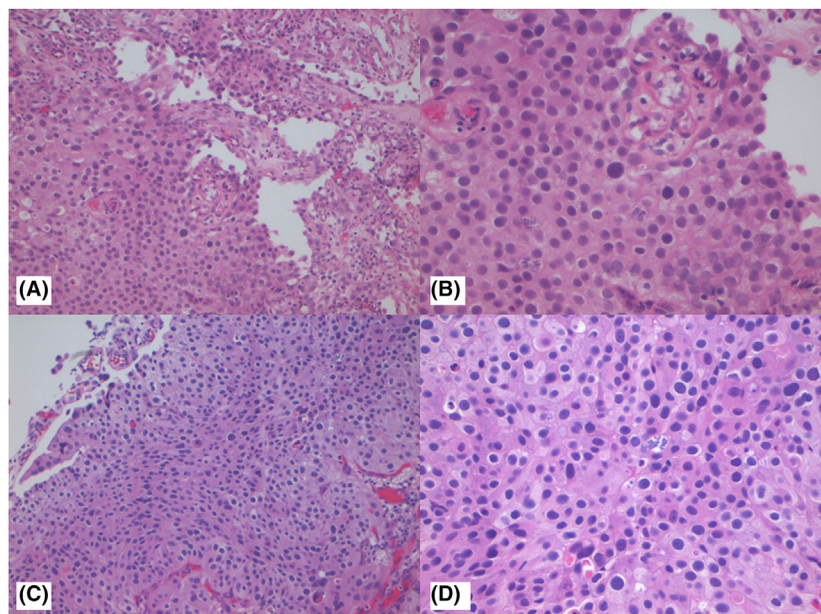


Figure 2. High grade-G2 papillary urothelial carcinomas with highest agreement on both 10 $\times$ and 20 $\times$ images. (A) A case at 10 $\times$ and (B), the same case at 20 $\times$ magnification. Two participants graded as low grade on 10 $\times$ and one on 20 $\times$, while two graded as high grade-G3 on 10 $\times$ and one on 20 $\times$. (C) A case at 10 $\times$ and (D), the same case at 20 $\times$ magnification. One participant graded as low grade on 10 $\times$, while one graded as high grade-G3 on 10 $\times$, and five on 20 $\times$. The histologic features in all images show increased cellular disorder, loss of normal polarity, more variability in nuclear size and shape and increased nuclear hyperchromasia compared with the cases in Figure 1. At 20 $\times$ magnification, mitotic figures can be appreciated in both cases. Note the difference in H&E staining between the two cases.

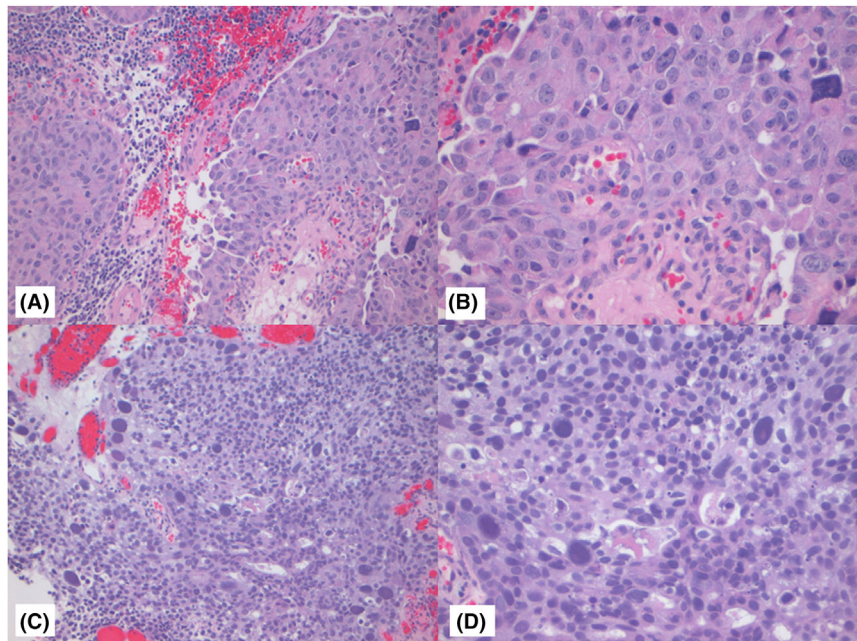


Figure 3. High grade-G3 papillary urothelial carcinomas with highest agreement on both 10× and 20× images. (A) A case at 10× and (B), the same case at 20× magnification. All participants agreed with HG-G3 (100%) on 10× and 20×. (C) A case at 10× and (D), the same case at 20× magnification. One participant graded as high grade-G2 on 10×. These cases show the extreme end of the high-grade spectrum with marked architectural disorder, lack of maturation, extreme nuclear pleomorphism coupled with marked nuclear hyperchromasia. In (B), prominent nucleoli are appreciated. Only occasional mitotic figures are seen (D).

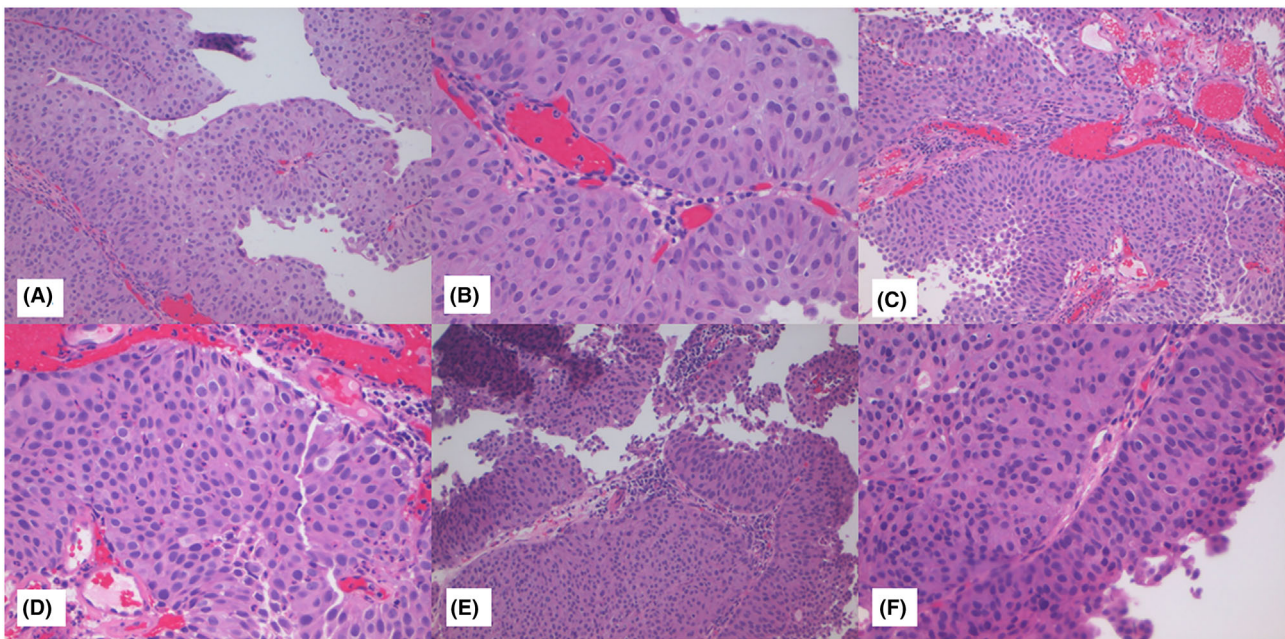


Figure 4. Low grade papillary urothelial carcinomas with the lowest agreement on both 10× and 20× images. (A) A case at 10× and (B), the same case at 20× magnification. Ten participants graded as HG-G2 on 10× and 12 on 20×. (C) A second case at 10× and (D), the same case at 20× magnification. Twelve participants graded as HG-G2 on 10× and 13 on 20×. (E) A third case at 10× and (F), the same case at 20× magnification. Fourteen participants graded as HG-G2 on 10× and 10 on 20×. Note that all cases were graded as high grade-G2, and none as high grade-G3, highlighting both the downside of using a two-tier grading system and benefit of using a three-tier system.

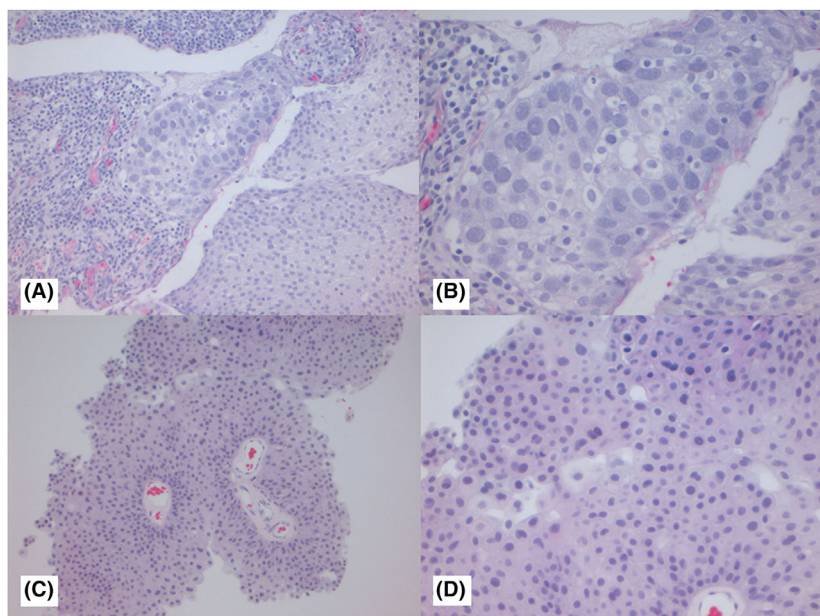


Figure 5. High grade-G2 papillary urothelial carcinomas with the lowest agreement on both 10× and 20× images. (A) A case at 10× and (B), the same case at 20× magnification. Twelve participants graded as high grade-G3 on 10× and thirteen on 20×. (C) A case at 10× and (D), the same case at 20× magnification. Thirteen participants graded as low grade on 10×, and seven on 20×. The above cases highlight how a binary “low” and “high” grade system may be too definitive and using a high grade-G2 category provides a more nuanced assessment around the diagnostic threshold.

be improved upon and that differentiating LG and HG cases remained a challenge. Further, there was interest in moving towards a hybrid grading scheme.⁸ Acknowledging the existing challenges in grade assignment, an expert opinion paper published prior to the Basel consensus meeting produced more detailed histologic criteria for distinguishing the WHO 1973 categories.¹⁹ These criteria were circulated to the study urothologists as part of the study instructions.

Half of the participants used WHO 2004/2022 in their day-to-day practice and not WHO 1973 or hybrid grading; however, the kappa values achieved were substantial across all participants. When sub-analysed by location of practice (Europe vs. North America), the kappa values, while slightly higher amongst Europeans, were substantial in both cohorts. This suggests that despite many not using hybrid grading currently, it has the potential to be adopted reasonably quickly. A key finding from our study was the near unanimous agreement in identifying the HG-G3 cases, arguably the most detrimental cases. Not surprising, and in keeping with survey data from the Basel consensus meeting,⁸ was the difficulty in separating HG-G2 and LG cases. The advantage of hybrid- three-tier grading is that when a case is assigned HG-G3, there is no uncertainty on the

treating clinician's part as to whether it is an LG or HG case, whereas a HG-G2 conveys that the carcinoma is closer to the border of LG.

Another finding from this work was the stated preference of pathologists to grade using two magnifications (10× and 20×). Unlike other genitourinary cancers such as renal cell carcinoma²⁰ and prostatic adenocarcinoma,^{21,22} there have been no recommendations on what magnification(s) to use to grade papillary NMIBC. While agreement using 10× and 20× were similar, there were slightly better overall kappa values for 10× image gradings compared with the 20× images gradings.

The three major limitations of this study were the inclusion of expert urologic pathologists with interest in bladder pathology, which raises the issue of whether the same results would be found amongst a larger group of nonspecialized pathologists. The cases were also selected from only two institutions, so while cases with variable H&E staining were included, in the real-world setting there would inevitably be greater variability in the intensity/quality of the staining, which could potentially impact grading assessments. Papillary urothelial neoplasm of low malignant potential (PUNLMP) cases are infrequent and were not included in this study. However, they would be considered part of the LG spectrum of cases

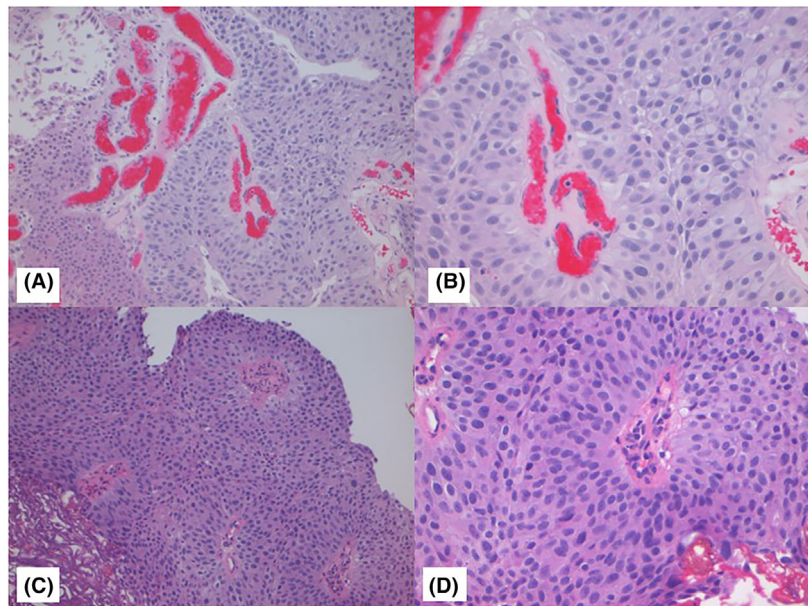


Figure 6. Cases with borderline agreement for diagnosis of HG-G2. (A) (10×) and (B) (20×) An HG-G2 (assigned grade). On 10× 6/18 participants graded this case as LG and 12/18 as HG-G2, whereas at 20× 3/18 graded as LG, 12/18 as HG-G2, and 3/18 as HG-G3. Note the degree of architectural disorder with loss of normal polarity and moderate nuclear pleomorphism, which is beyond that seen in LG cases. Conversely, there are no hyperchromatic nuclei, marked pleomorphism, or frequent mitotic figures. On the 20× image, nucleoli are more easily visible (although the H&E stain is pale) and on the higher magnification, the smaller field of view suggests a greater degree of atypia than on 10×, which if viewed in isolation could lead to a diagnosis of HG-G3. (C) and (D) An HG-G2, which 13 /18 called HG-G2, 3/18 LG and 1/18 HG-G3. (D) (20× same area), 13/18 called HG-G2, and 4/18 called LG. There is loss of normal polarity, best appreciated in (C), with moderate nuclear pleomorphism and in (D) occasional nucleoli are visible. A single large nucleus is appreciated to the right of both images, which is likely an umbrella cell, as there is a reasonable amount of surrounding cytoplasm. The challenge in this case is mostly distinguishing from LG. The loss of polarity along with the increased nuclear atypia are the features that put this case into the HG-G2 category.

in terms of grading¹⁹ and some pathologists who use the terminology of PUNLMP may have called some cases this entity if that option had been available and the age/clinical scenario were appropriate. Finally, we selected static images as the modality of choice to ensure the images were reviewed by participants at the 10× and 20× magnifications. Supplying scanned images could have introduced an element of zooming and toggling between multiple magnifications. It is possible that using a less rigid modality than forced grading on static images may have improved the results, but this could only be addressed in a dedicated digital versus static images study. In real-world practice, grading will likely be performed by most pathologists using both 10× and 20× magnification, similar to our study participants.

In conclusion, hybrid grading of papillary Ta NMIBC showed substantial agreement as assessed by interrater kappa values. This was higher than reported in prior studies that assessed agreement using WHO 1973 and WHO 2004/2022.⁸

Pathologists readily recognize HG- G3 cases, but challenges still remain within the HG-G2 category. We also show that similar kappa values for grading were achieved using 10× and 20× magnification and we are, to our knowledge, the first group to specifically assess magnification as a factor in grade assignment in papillary urothelial carcinomas. Our results show that urologic pathologists can readily adapt to using a three-tier hybrid grading system with good interrater agreement. Our positive findings should encourage much needed further clinicopathologic studies using these well-defined criteria so that future iterations of classification system revisions can incorporate refinements based on demonstrated clinical impact.

Author contributions

M.R.D. and T.H.v.d.K. designed the study, performed the research, analysed and interpreted the data, drafted and critically revised the article. K.L. analysed the data and contributed to the review and editing of

Table 3. Assessment of interrater agreement of images at 10× (30 images) and 20× (30 images)

Group	Observations	Raters	Kappa	95% CI
All images at 10×	537	18	0.641	(0.557, 0.725)
North America	237	8	0.647	(0.552, 0.741)
Europe	240	8	0.663	(0.585, 0.741)
Low grade	268	18	0.296	(0.166, 0.426)
High grade	269	18	0.529	(0.393, 0.666)
All images at 20×	539	18	0.623	(0.533, 0.714)
North America	240	8	0.623	(0.518, 0.728)
Europe	239	8	0.682	(0.593, 0.772)
Low grade	239	18	0.241	(0.112, 0.369)
High grade	270	18	0.549	(0.417, 0.680)

Categories for kappa agreement: 0–0.20 indicate slight agreement, 0.21–0.40 indicate fair agreement, 0.41–0.60 indicate moderate agreement, 0.61–0.80 indicate substantial agreement, and 0.81–1 indicate almost perfect agreement.

CI, confidence interval; ×, magnification.

the article. F.A., Y.A., M.B.A., L.C., S.D., M.E., A.H., A.L.B., S.M., M.L.v.M., J.O., G.P.P., H.S., S.S., J.W., S.R.W., and S.E.W. performed the research and contributed to the review and editing of the article. All authors have read and approved the submitted and final versions of the article.

Acknowledgements

This work was presented, in part, at the United States and Canadian Academy of Pathology Annual Meeting, Baltimore, Maryland on March 27th 2024.

Funding information

None.

Conflict of interest

The authors have no conflicting interests to disclose.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author.

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394–424.
- Beijert IJ, Hagberg O, Gardmark T *et al.* The importance of being grade 3: a plea for a three-tier hybrid classification system for grade in primary non-muscle-invasive bladder cancer. *Eur Urol* 2024; **86**: 391–399.
- Downes MR, Lajkosz K, Kuk C, Gao B, Kulkarni GS, van der Kwast TH. The impact of grading scheme on non-muscle invasive bladder cancer progression: potential utility of hybrid grading schemes. *Pathology* 2022; **54**: 425–433.
- Holzbeierlein JM, Bixler BR, Buckley DI *et al.* Diagnosis and treatment of non-muscle invasive bladder cancer: aua/suo guideline: 2024 amendment. *J Urol* 2024; **211**: 533–538.
- Babjuk M, Burger M, Capoun O *et al.* European association of urology guidelines on non-muscle-invasive bladder cancer (ta, t1, and carcinoma in situ). *Eur Urol* 2022; **81**: 75–94.
- WHO. *Who classification of tumours*. Vol. 8. 5th ed. Lyon: IARC Publications, 2022.
- Epstein JI, Amin MB, Reuter VR, Mostofi FK, Committee TBCC. 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.
- Downes MR, Hartmann A, Shen S *et al.* Comparison of bladder cancer grading system performance. *Am J Surg Pathol* 2024; **48**: e1–e10.
- van der Kwast TH, Bubendorf L, Cheng L. International society of urological pathology (isup) consensus conference on current issues in bladder cancer: progresses and challenges. *Am J Surg Pathol* 2024; **48**: 1–3.
- Beijert IJ, Cheng L, Liedberg F *et al.* International opinions on grading of urothelial carcinoma: a survey among european association of urology and international society of urological pathology members. *Eur Urol Open Sci* 2023; **52**: 154–165.
- Mitani AA, Freer PE, Nelson KP. Summary measures of agreement and association between many raters' ordinal classifications. *Ann Epidemiol* 2017; **27**: 677–685.
- 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**: 465–466.
- Soukup V, Capoun O, Cohen D *et al.* Prognostic performance and reproducibility of the 1973 and 2004/2016 world health organization grading classification systems in non-muscle-invasive bladder cancer: a european association of urology non-muscle invasive bladder cancer guidelines panel systematic review. *Eur Urol* 2017; **72**: 801–813.
- van Rhijn BWG, Hentschel AE, Brundl J *et al.* Prognostic value of the who1973 and who2004/2016 classification systems for grade in primary ta/t1 non-muscle-invasive bladder cancer: a multicenter european association of urology non-muscle-invasive bladder cancer guidelines panel study. *Eur Urol Oncol* 2021; **4**: 182–191.
- van Rhijn BW, Musquera M, Liu L *et al.* Molecular and clinical support for a four-tiered grading system for bladder cancer based on the who 1973 and 2004 classifications. *Mod Pathol* 2015; **28**: 695–705.

16. Zhao J, Richter J, Wagner U *et al*. Chromosomal imbalances in noninvasive papillary bladder neoplasms (pta). *Cancer Res* 1999; **59**; 4658–4661.
17. 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.
18. Beijert IJ, Hentschel AE, Brundl J *et al*. Prognosis of primary papillary ta grade 3 bladder cancer in the non-muscle-invasive spectrum. *Eur Urol Oncol* 2023; **6**; 214–221.
19. van der Kwast T, Liedberg F, Black PC *et al*. International society of urological pathology expert opinion on grading of urothelial carcinoma. *Eur Urol Focus* 2022; **8**; 438–446.
20. Delahunt B, Chevillie JC, Martignoni G *et al*. Members of the IRTTP: the international society of urological pathology (isup) grading system for renal cell carcinoma and other prognostic parameters. *Am J Surg Pathol* 2013; **37**; 1490–1504.
21. Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: current state and prognostic implications. *Diagn Pathol* 2016; **11**; 25.
22. Histological DFG. Histological grading and staging of prostatic carcinoma. In Tannenbaum M ed. *The prostate*. Philadelphia: Lea and Febiger, 1977.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Variance and intraclass coefficient estimates from models used in kappa estimation.