Int J Clin Exp Pathol 2012;5(9):882-891
www.ijcep.com /ISSN:1936-2625/IJCEP1208028

Original Article

Accuracy of grading of urothelial carcinoma on urine cytology: an analysis of interobserver and intraobserver agreement

Michelle D Reid[1], Adeboye O Osunkoya[1], Momin T Siddiqui[1], Stephen W Looney[2]

[1]Department of Pathology, Emory University School of Medicine and [2]Department of Biostatistics, Georgia Health Sciences University, Augusta, GA, USA

Received August 31, 2012; Accepted September 24, 2012; Epub October 20, 2012; Published October 30, 2012

Abstract: Background: Urine samples of known urothelial carcinoma were independently graded by 3 pathologists with (MS, MR) and without (AO) fellowship training in cytopathology using a modified version of the 2004 2-tiered World Health Organization classification system. By measuring interobserver and intraobserver agreement among pathologists, compared with the gold standard of biopsy/resection, specimen accuracy and reproducibility of grading in urine was determined. Methods: 44 urine cytology samples were graded as low or high-grade by 3 pathologists with a 2-3 week interval between grading. Pathologists were blinded to their and others' grades and histologic diagnoses. Coefficient kappa was used to measure interobserver and intraobserver agreement among pathologists. Accuracy was measured by percentage agreement with the biopsy/resection separately for each pathologist, and for all pathologists and occasions combined. Results: The overall accuracy was 77% (95% C.I., 72% – 82%). Pathologist AO was significantly more accurate than MR on occasion 1 ($p = 0.006$) and 2 ($p = 0.039$). No other significant differences were found among the observers. Interobserver agreement using coefficient kappa was unacceptably low, with all but one of the kappa value being less than 0.40, the cutoff for a "fair" degree of agreement. Intraobserver agreement, as measured by coefficient kappa, was adequate. Conclusions: Our study underscores the lack of precision and subjective nature of grading urothelial carcinoma on urine samples. There was poor inter- and intraobserver agreement among pathologists despite fellowship training in cytopathology. Clinicians and cytopathologists should be mindful of this pitfall and avoid grading urothelial carcinoma on urine samples, especially since grading may impact patient management.

Keywords: Urothelial carcinoma, accuracy of grading, urine cytology

Introduction

Cancer of the bladder and genitourinary tract has shown a significant increase in incidence in the United States over the years. In 2011 alone, 71,980 new cases were diagnosed and there were 15,510 related deaths (http://www.cancer.org/acs/groups/content/@epidemiology-surveilance/documents/document/acspc-029771.pdf). Cystoscopy is considered the gold standard for diagnosis and post-treatment surveillance of bladder cancer; however, because of the invasive nature of the procedure and prohibitive cost, it has been replaced by urine cytology, which is a cheaper screening and surveillance tool for early cancer detection.

Urine cytology was first described by Papanicolaou and Marshall in 1945 [1]. The urothelial cells in urine represent a much larger surface area of the genitourinary tract compared to biopsies, which only sample a limited area. In addition, cytologic samples allow for three-dimensional examination of urothelial cells, while tissue sections render only two-dimensional views [2].

However, despite its popularity, urine cytology is not an infallible test. The identification of low grade urothelial carcinoma is fraught with difficulty and low sensitivity rates [3-9]. As a result, numerous ancillary techniques have been developed to improve its sensitivity, including immunohistochemistry (cytokeratin 20 [10]

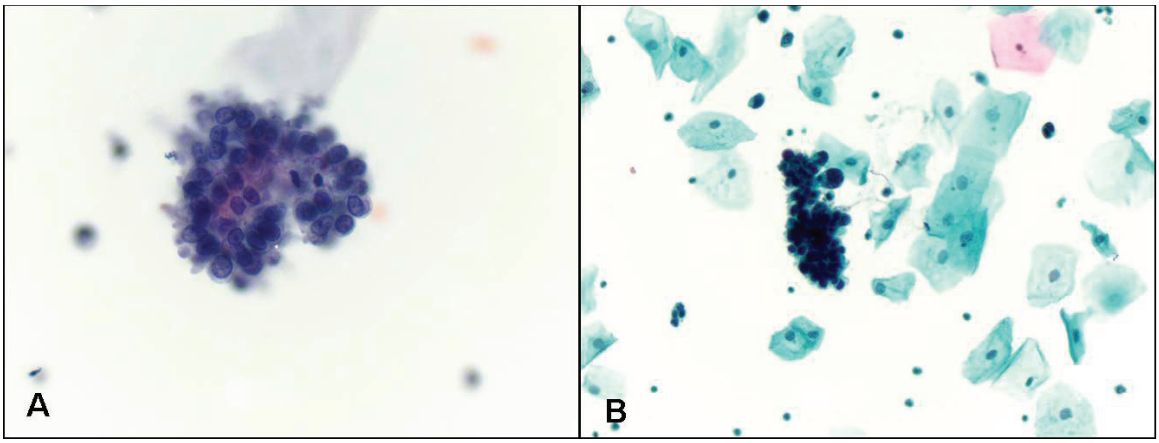

Figure 1. Low grade urothelial carcinoma. A. Papillary cluster of malignant cells with high nuclear to cytoplasmic ratio (Papanicolaou stain, magnification x 400). B. Three-dimensional cluster of malignant cells with nuclear pleomorphism, hyperchromasia and irregular nuclear borders (Papanicolaou stain, magnification x 200).

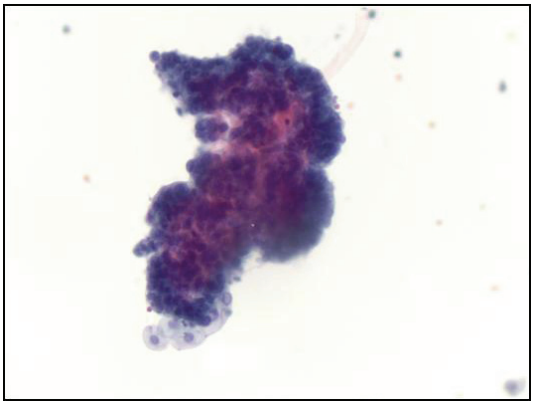

Figure 2. Low grade urothelial carcinoma. Urine sample contains papillary cluster of cells with a suggestion of a central fibrovascular core (Papanicolaou stain, magnification x 200).

and p53 [11]), DNA image cytometry [2], microsatellite analysis [12] and Urovysion fluorescence in situ hybridization (FISH) [13], to name a few. Urovysion FISH is the most widely used ancillary test for diagnosis; however, it also has limitations due to false-positive results in non-urothelial tumors including primary bladder squamous cell carcinoma, adenocarcinoma, as well as metastatic prostate, colon and renal cell carcinoma [14, 15].

We graded 44 urine samples of urothelial carcinoma, all of which had been confirmed histologically. These cases were independently graded by 3 pathologists with and without fellowship training in cytopathology using the most recent 2004 two-tiered World Health Organization (WHO) classification system [16]. We determined the accuracy and reproducibility of the grading of urothelial carcinoma in urine by measuring interobserver and intraobserver agreement among pathologists, compared with the gold standard follow-up biopsy or resection specimen.

While other studies have examined interobserver variability in the histologic grading of urothelial carcinoma, none have focused on cytologic grading. Our purpose was not to refine existing criteria for the grading of urothelial carcinoma on cytology, nor was it to evaluate the sensitivity and specificity of the current WHO grading system. The aim of this study was to determine if accurate and reproducible cytologic grading was possible in urine samples and if the process was comparable to histologic grading. This has implications for the management of patients with bladder and genitourinary tract cancer and could impact the way cytopathologists interpret and report urine cytology results.

Materials and methods

A search of the Emory University Pathology Department's archives revealed 44 urine cytology specimens diagnosed as positive for urothelial carcinoma that had corresponding histologic confirmation of diagnosis and tumor grade. These cases were diagnosed between June 2004 and June 2010. This final number of cases included all urothelial carcinomas diagnosed on cytology with corresponding histologic confirmation. No attempt was made to influ-

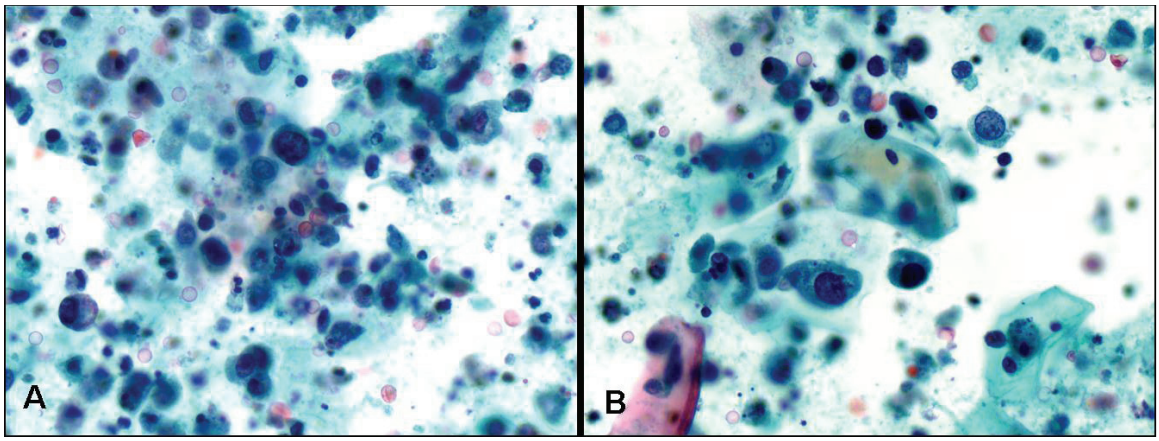

Figure 3. High grade urothelial carcinoma. Note numerous large single malignant cells with high nuclear to cytoplasmic ratio, coarse chromatin, irregular nuclear borders, single cell necrosis and background necrosis (Papanicolaou stain, magnification x 400).

ence the cases in any manner, as all "qualifiers" within the time period specified were accepted for the study. The following data were collected: patient demographics; cytologic and histologic diagnoses; and time from urine collection to biopsy/resection. All samples were processed according to standard guidelines; cytospin and Thinprep® slides were prepared and each slide was given a unique accession number. A log form was then created to facilitate the entry of the 3 reviewers' grades (low- versus high-grade urothelial carcinoma). The 3 reviewers included 2 fellowship-trained cytopathologists (MR and MS) and 1 general surgical pathologist with sign-out responsibility in cytopathology (AO). In addition, the latter pathologist (AO) had fellowship training in urologic pathology. The 3 reviewers had previously agreed on the criteria for the grading of urothelial carcinoma (using cytologic and modified histologic 2004 World Health Organization criteria) (**Figures 1**, **2** and **3**). For low-grade urothelial carcinoma, these included cytoplasmic homogeneity, high nuclear-to-cytoplasmic ratio, irregular nuclei (**Figure 1**), as well as papillae with and without fibrovascular cores (**Figure 2**) and irregular cell clusters (**Figure 1**). For high-grade urothelial carcinoma, features included high nuclear-to-cytoplasmic ratio, marked nuclear hyperchromasia, coarse chromatin, irregular nuclei, plus or minus large nucleoli and isolated malignant cells (**Figure 3**).

All 44 cytologic samples and 2 log forms were then given to each reviewer for grading on 2 occasions, with a 2-3 week interval between grading. This interval was chosen with the aim of identifying intraobserver variability. Each pathologist was blinded to his/her first and second round grades, the grades of other reviewers, as well as the patients' histories and histologic diagnoses.

Finally, all log forms were collected and tabulated, along with additional information including patient demographics, date of urine collection, date and type of follow-up biopsy/resection, histologic diagnosis and grade. Tissue diagnosis and grade were considered the gold standard for diagnosis.

Coefficient kappa was then used to measure interobserver and intraobserver agreement among pathologists. Accuracy was measured by percentage agreement with the gold standard (tissue biopsy or resection) for each pathologist on each occasion and for all pathologists and occasions combined.

Statistical analysis

Cohen's kappa is the generally accepted method for assessing agreement between two dichotomous variables, neither of which can be assumed to be the gold standard, but several deficiencies have been noted. [17] For example, the value of kappa is affected by any discrepancy in the relative frequencies of the two categories being scored (in this case, "high grade" vs. "low grade"): The higher the discrepancy, the smaller the value of kappa. To adjust for these deficiencies, the prevalence-adjusted

Table 1. Descriptive statistics

Characteristic	Categories	n	Mean ± SD	Minimum	Maximum	Percent
Age (yr)	---	44	70.6± 9.4	47	92	---
Gender	Female	12	---	---	---	27.3
	Male	32	---	---	---	72.7
Grade	Low	3	---	---	---	6.8
	High	41	---	---	---	93.2
Days Between Urine Sample & Biopsy	---	44	13.1 ± 22.4	1	86	---
Histologic Diagnosis	HGPUC	23	---	---	---	52.3
	CIS	9	---	---	---	20.5
	HGUC	6	---	---	---	13.6
	Other*	6	---	---	---	13.6
Lamina propria Invasion	LPI	22	---	---	---	49.9
	Suspicious for LPI	1	---	---	---	2.2
	No LPI	21	---	---	---	47.7
Source	Bladder BX	19	---	---	---	43.2
	TURBT	11	---	---	---	25.0
	Renal Pelvis BX	4	---	---	---	9.1
	Ureter BX	2	---	---	---	4.6
	Nephroureterectomy	2	---	---	---	4.6
	Other*	6	---	---	---	13.6

*Includes one each of bladder BX/Renal pelvis BX, Cystoprostatectomy, Kidney BX, Ureteral orifice biopsy, uretero-ileal anastomosis, Urethral Tumor BX.

Table 2. Percent correct by observer and occasion

Observer	Occasion	Number (n)	Number Correct	Percent Correct	95% Confidence Interval
MS	1	44	32	73	58 - 84
MS	2	44	35	80	66 - 90
MS	Both	88	67	76	67 - 85
AO	1	44	37	84	71 - 93
AO	2	44	39	89	77 - 96
AO	Both	88	76	86	79 - 94
MR	1	44	28	64	49 - 77
MR	2	44	32	73	58 - 84
MR	Both	88	60	68	58 - 78
All	Trial 1	132	97	73	66 -81
All	Trial 2	132	106	80	74 -87
All	Both	264	203	77	72 - 82

and bias-adjusted kappa (PABAK), which is equivalent to the proportion of "agreements" between the variables minus the proportion of "disagreements", may also be reported in addition to kappa [17].

In the present study, both coefficient kappa and PABAK, along with 95% confidence intervals (C.I.), were used to measure intraobserver and interobserver agreement for all cytology samples. P-values were also calculated for testing the null hypothesis that the true level of agreement is zero. Interobserver agreement was analyzed separately for each pair of observers and each occasion. Intraobserver agreement between the two occasions was analyzed separately for each observer.

Statistical methods for clustered data [18] 0 were used to estimate the percent correctly

Table 3. Coefficient kappa for interobserver variability

Observer	Occasion	Kappa	95% C.I.	*p*-Value
MS vs. AO	1	0.05	(-0.26, 0.36)	1.000
MS vs. MR	1	0.22	(-0.08, 0.51)	0.235
AO vs. MR	1	0.37	(0.09, 0.66)	0.013
MS vs. AO	2	0.15	(-0.19, 0.50)	0.573
MS vs. MR	2	0.57	(0.30, 0.84)	< 0.001
AO vs. MR	2	0.29	(-0.01, 0.59)	0.053

Table 4. PABAK coefficient for interobserver variability

Observer	Occasion	PABAK	95% C.I.	*p*-Value
MS vs. AO	1	0.41	(0.12, 0.65)	< 0.001
MS vs. MR	1	0.36	(0.07, 0.61)	0.007
AO vs. MR	1	0.50	(0.21, 0.72)	0.016
MS vs. AO	2	0.55	(0.26, 0.76)	< 0.001
MS vs. MR	2	0.68	(0.42, 0.86)	< 0.001
AO vs. MR	2	0.50	(0.21, 0.72)	< 0.001

PABAK: prevalence-adjusted and bias-adjusted kappa.

classified for each observer on each occasion, as well as for both occasions combined for each observer, for all observers combined for each occasion, and for all observers and occasions combined. Confidence intervals for percent correct were calculated after adjusting for the clustering of occasions within observers. The exact version of Cochran's Q test, with mid-p correction [19], was used to compare the observers in terms of percent accurately classified, and to compare interobserver agreement between each pair of observers and between occasions and to compare intraobserver agreement among the 3 observers.

Results

Table 1 contains descriptive statistics for the study. The results from the gold standard histologic diagnosis were used to evaluate the accuracy of the cytological grading by calculating the total percentage correct for each observer on each occasion (**Table 2**). The low grade (n = 3) and high-grade (n = 41) carcinomas were combined for this analysis. In order to obtain estimates of overall accuracy, the results for both occasions were combined for each observer separately, the results for all 3 observers were combined for each occasion, and all results for all observers and all occasions were also combined (**Table 2**). Each of the confidence intervals for these overall estimates of accuracy were adjusted for the clustering of occasions within observers, as described in the Statistical Analysis section. The overall accuracy was 77% (95% C.I., 72% – 82%). The accuracy rates were better on the second occasion for all three reviewers (**Table 2**). Observer AO was significantly more accurate than observer MR on Occasion 1 (p = 0.006) and Occasion 2 (p = 0.039). No other significant differences were found among the observers in terms of accuracy.

Interobserver agreement, as measured by coefficient kappa, was unacceptably low (**Table 3**). All but one of the kappa values were less than 0.40, which is the cutoff for a "fair" degree of agreement [20]. Only two of the interobserver agreement coefficients were significantly different from zero, indicating relatively better agreement, and these were the agreement between AO and MR on occasion 1 and the agreement between MS and MR on occasion 2. Only one of the 95% confidence intervals in **Table 3** included the value 0.75, which is generally considered to be the minimally acceptable value for a reliable clinical measurement.

However, because of the discrepancy in the frequencies of low grade (n = 3) and high grade (n = 41) carcinomas in the sample, the value of kappa was likely to be attenuated. Accordingly, the PABAK coefficient was also calculated as a measure of interobserver agreement (**Table 4**). As expected, the PABAK coefficients indicated higher levels of interobserver agreement; however, only two of the 95% confidence intervals (in **Table 4**) included the value 0.75, again indicating less than adequate reliability. The levels of interobserver agreement, as measured by PABAK, were generally comparable, regardless of which pair of observers was being compared, or which occasion. The only exception was the interobserver agreement between observers MR and MS; the PABAK coefficient measuring agreement between these two observers was significantly higher at occasion 2 than at occasion 1 (p = 0.022).

The level of intraobserver agreement, as measured by coefficient kappa, was adequate (**Table 5**). For observers AO and MS, the kappa value exceeded the 0.75 cutoff. In fact, for

Table 5. Coefficient kappa for intraobserver variability

Observer	Kappa	95% C.I.	*p*-Value
AO	0.83	(0.60, 1.00)	< 0.001*
MS	0.78	(0.55, 1.00)	< 0.001*
MR	0.58	(0.32, 0.84)	< 0.001*

observers AO and MS, the 95% confidence interval included the maximum possible value of 1.00 (perfect agreement between occasions 1 and 2). Even though the kappa value for observer MR was less than 0.75, the 95% confidence interval had an upper limit of 0.84, which indicates "almost perfect" agreement according to the Landis and Koch criteria [20]. All of the intraobserver kappa coefficients were significantly greater than zero. None of the observers differed significantly from either of the others in terms of their level of intraobserver agreement. The PABAK coefficients for intraobserver agreement differed very little from the kappa coefficients.

Discussion

Our study utilized cytomorphologic characteristics to grade urothelial carcinoma in urine samples into low and high grade tumors, thus assessing the reliability and reproducibility of the current 2004 WHO grading system in urine samples. We evaluated interobserver and intraobserver agreement in the grading of these tumors among 3 pathologists, with and without fellowship training in cytopathology. We found that the overall accuracy of grading on urine cytology was unacceptably low at 77%. Interobserver agreement of urothelial carcinoma grade, as measured by coefficient kappa, was unacceptably low among the trio. In addition, the PABAK coefficients also indicated poor reliability in cytologic grading. The level of intraobserver agreement, as measured by coefficient kappa, was adequate among the 3 reviewers.

The histopathologic classification of urothelial carcinoma has been practiced since the 1920s [21] and revised by Bergkvist [22], Mostofi [23], Pauwels [24] and the WHO [25]. Mostofi's 1973 WHO system was once the most commonly used and had remained unchanged for 30 years, despite problems with its reproducibility [26-28]. In 1998, the WHO/international Society of Urologic Pathology (ISUP) introduced a modified 3-tiered classification system for urothelial neoplasms [25]. However, problems remained with the system, as evidenced by problematic interpretation of heterogeneous tumors [29] and interobserver variability in grading [30]. This variability in grading was seen not only with the 1998 WHO system [25] but also with other grading systems as well [22, 23]. Given its strong prognostic impact in predicting biologic aggression, tumor grading strongly influences clinical management and prediction of patient outcome. However, there remains an inherent degree of subjectivity in the grading of urothelial carcinoma, resulting in significant interobserver variability [26, 28, 31, 32].

In 2004, the WHO agreed on a more standardized classification of urothelial tumors that improved reproducibility in diagnosis and grading [33]. It included a more detailed description of categories of non-invasive papillary urothelial tumors, and converged the former 3-tiered grading system into a two-tiered (low grade versus high grade) one [32]. However, although better than its 1973 and 1998 predecessors, some pathologists still found the new system difficult to histologically classify heterogeneous lesions as low-grade or high-grade [34].

While the accuracy and reproducibility of numerous grading systems for urothelial carcinoma have been studied histologically, [26-29, 31, 32, 34, 35] the cytologic grading of urothelial carcinoma has only been rarely attempted [6, 36-38]. Wojcik *et al* used digitalized computer-assisted quantitative nuclear grading to differentiate low- and high grade urothelial carcinoma from normal urothelium on cytology [38]. They used 38 nuclear features (including size, shape and chromatin organization) which they found helpful in differentiating low from high grade carcinoma and normal urothelium. This method, however, is expensive and is also not universally available. Vom Dorp *et al* combined a cytologic-cytometric grading system with the measurement of nuclear diameter and circumference [37] and claimed to be able to distinguish grade 1 tumors from grade 2 and 3 carcinomas on cytology. Using the 1998 WHO/ISUP system, Whisnant *et al* attempted to distinguish PUNLMP from low grade urothelial carcinoma on cytology and found that while urine was sen-

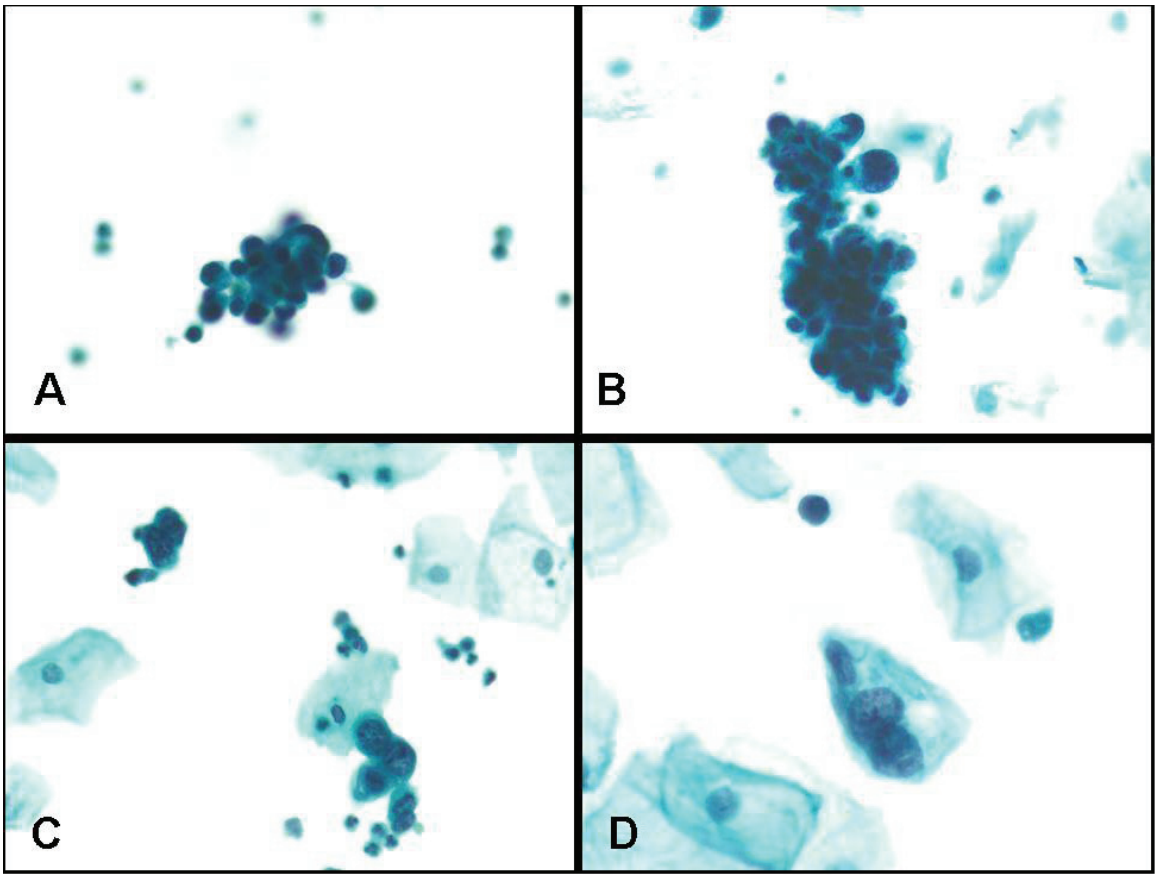

Figure 4. Case of urothelial carcinoma showing mixed low and high grade features. Note tight clusters of malignant hyperchromatic low grade tumor cell clusters with dense cytoplasm and irregular nuclear borders (in A and B), and rare, isolated clusters of high grade malignant cells (in C and D) (Papanicolaou stain, magnification x 200).

sitive in detecting abnormal cells in both lesions, it was not specific in distinguishing the two entities [6]. Rubben *et al* claimed that the accuracy for grade 1 tumors was as high as 78% in urine cytology samples [36], however others have not shown similar results.

An ideal grading system should be relatively simple and highly reproducible, irrespective of the level of expertise and sub-specialization of its users. Our study looked at cytomorphology as the sole method for distinguishing between low and high grade urothelial carcinoma and we found that there was poor interobserver agreement in the grading of urothelial carcinoma. Furthermore, there was some intraobserver variation in grading as well. One possible reason for the poor interobserver agreement was level of experience of the reviewer. We addressed this potential confounder by comparing the more junior to the senior cytopathologist and found that the senior cytopathologist (MS) with 7 more years' experience was more accurate than the junior one (MR). Therefore it would appear that level of experience may contribute somewhat to interobserver variation and accuracy of grading. However, despite this, the overall grading accuracy for all reviewers was minimally acceptable at best and this was independent of specialty training in cytopathology. All 3 pathologists had better accuracy rates on the second round of scoring compared to the first. This may be related to more learned familiarity with the cytologic differences between the low and high grade tumors after the first go-round. Interestingly, of the 3 pathologists, the general surgical pathologist (AO) performed best at differentiating low grade from high grade urothelial carcinoma. Of all our findings, this was the most surprising, since AO had no formal training in cytopathology. However it should be noted that in addition to

routinely signing out cytology specimens, AO also had specialized training as a urologic pathologist, which may or may not have given him an advantage over the other observers.

Other potential confounders in the poor grading of cytology samples include a lack of careful review of slides, reviewer fatigue (which we could not accurately test for), as well as tumor heterogeneity. Fifty-one percent (n=21/41) of the high-grade tumors examined were considered "mixed low and high grade tumors" by at least 2 reviewers (**Figure 4**), leading to an increased potential for grading variability. Failure to identify low or high grade malignant cells could also have occurred because of low specimen cellularity, obscuring inflammation, blood or squamous contaminants.

The diagnosis of low grade urothelial carcinoma in urine samples is difficult and is a source of frustration for practicing cytopathologists. Low grade carcinoma often resembles normal or reactive urothelium with only subtle differences. Chu *at al* [7] and Raab *et al* [39] have attempted to separate low grade tumors from their normal and high-grade counterparts by suggesting cytoplasmic vacuolization/homogeneity, nuclear to cytoplasmic ratio and nuclear membrane irregularity as features of low grade tumors. Others suggest that papillary fragments with fibrovascular cores, especially when found in voided urine, are worrisome for low grade urothelial carcinoma. However, none of these are specific. In fact, some cytopathologists believe that normal urothelium and low-grade urothelial carcinoma cannot be distinguished by light microscopy alone [40].

Our study underscores the lack of precision and subjective nature of grading of urothelial carcinoma on urine samples, and highlights the poor inter- and intraobserver agreement that may be seen among pathologists with and without formal training in cytopathology. Clinicians and cytopathologists alike should be mindful of this pitfall and thus avoid grading urothelial carcinoma on urine samples, especially since grading may impact clinical management. Based on our findings, one can conclude that the current two-tiered WHO histologic grading system is difficult to reproduce in urine specimens and we propose that for positive cytologic samples one should merely use the diagnostic category "positive for malignant cells" without attempting to grade these tumors. Future focus should be placed on the development of non-morphologic ancillary tests that might improve the diagnostic sensitivity of urine cytology, especially for low-grade urothelial carcinoma.

Address correspondence to: Dr. Michelle D Reid, Department of Pathology, Emory University School of Medicine, 1364 Clifton Rd NE, Room H190, Atlanta, GA 30322, USA. E-mail: michelle.reid@emory.edu

References

[1] Papanicolaou GN and Marshall VF. Urine Sediment Smears as a Diagnostic Procedure in Cancers of the Urinary Tract. Science 1945; 101: 519-520.

[2] Forte JD, Croker BP and Hendricks JB. Comparison of histologic and cytologic specimens of urothelial carcinoma with image analysis. Implications for grading. Anal Quant Cytol Histol 1997; 19: 158-166.

[3] Brown FM. Urine cytology. It is still the gold standard for screening? Urol Clin North Am 2000; 27: 25-37.

[4] Curry JL and Wojcik EM. The effects of the current World Health Organization/International Society of Urologic Pathologists bladder neoplasm classification system on urine cytology results. Cancer 2002; 96: 140-145.

[5] Sullivan PS, Nooraie F, Sanchez H, Hirschowitz S, Levin M, Rao PN and Rao J. Comparison of ImmunoCyt, UroVysion, and urine cytology in detection of recurrent urothelial carcinoma: a "split-sample" study. Cancer 2009; 117: 167-173.

[6] Whisnant RE, Bastacky SI and Ohori NP. Cytologic diagnosis of low-grade papillary urothelial neoplasms (low malignant potential and low-grade carcinoma) in the context of the 1998 WHO/ISUP classification. Diagn Cytopathol 2003; 28: 186-190.

[7] Chu YC, Han JY, Han HS, Kim JM and Suh JK. Cytologic evaluation of low grade transitional cell carcinoma and instrument artifact in bladder washings. Acta Cytol 2002; 46: 341-348.

[8] Wiener HG, Vooijs GP and van't Hof-Grootenboer B. Accuracy of urinary cytology in the diagnosis of primary and recurrent bladder cancer. Acta Cytol 1993; 37: 163-169.

[9] Tetu B, Tiguert R, Harel F and Fradet Y. ImmunoCyt/uCyt+ improves the sensitivity of urine cytology in patients followed for urothelial carcinoma. Mod Pathol 2005; 18: 83-89.

[10] Soyuer I, Tokat F and Tasdemir A. Significantly increased accuracy of urothelial carcinoma detection in destained urine slides with com-

bined analysis of standard cytology and CK-20 immunostaing. Acta Cytol 2009; 53: 357-360.
[11] Righi E, Rossi G, Ferrari G, Dotti A, De Gaetani C, Ferrari P and Trentini GP. Does p53 immunostaining improve diagnostic accuracy in urine cytology? Diagn Cytopathol 1997; 17: 436-439.
[12] Wild PJ, Fuchs T, Stoehr R, Zimmermann D, Frigerio S, Padberg B, Steiner I, Zwarthoff EC, Burger M, Denzinger S, Hofstaedter F, Kristiansen G, Hermanns T, Seifert HH, Provenzano M, Sulser T, Roth V, Buhmann JM, Moch H and Hartmann A. Detection of urothelial bladder cancer cells in voided urine can be improved by a combination of cytology and standardized microsatellite analysis. Cancer Epidemiol Biomarkers Prev 2009; 18: 1798-1806.
[13] Sokolova IA, Halling KC, Jenkins RB, Burkhardt HM, Meyer RG, Seelig SA and King W. The development of a multitarget, multicolor fluorescence in situ hybridization assay for the detection of urothelial carcinoma in urine. J Mol Diagn 2000; 2: 116-123.
[14] Reid-Nicholson MD, Motiwala N, Drury SC, Peiper SC, Terris MK, Waller JL and Ramalingam P. Chromosomal abnormalities in renal cell carcinoma variants detected by Urovysion fluorescence in situ hybridization on paraffin-embedded tissue. Ann Diagn Pathol 2011; 15: 37-45.
[15] Reid-Nicholson MD, Ramalingam P, Adeagbo B, Cheng N, Peiper SC and Terris MK. The use of Urovysion fluorescence in situ hybridization in the diagnosis and surveillance of non-urothelial carcinoma of the bladder. Mod Pathol 2009; 22: 119-127.
[16] Lopez-Beltran ASG, Gasser T, et al. Tumors of the Urinary System. In: Eble JNSG, Epstein JI, Sesterhenn IA, editors. World Health Organization Classification of Tumours. Pathology and Genetics> Tumors of the Urinary Sytem and Male Genital Organs. Lyon, France: IARC Press; 2004. p. 89-109.
[17] Byrt T, Bishop J and Carlin JB. Bias, prevalence and kappa. J Clin Epidemiol 1993; 46: 423-429.
[18] Donner A, Shoukri MM, Klar N and Bartfay E. Testing the equality of two dependent kappa statistics. Stat Med 2000; 19: 373-387.
[19] Lehmann E. Nonparametrics: Statistical Methods Based on Ranks. San Francisco: Holden-Day, 1975.
[20] Landis JR and Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159-174.
[21] Broders AC. Epithelioma of the Genito-Urinary Organs. Ann Surg 1922; 75: 574-604.
[22] Bergkvist A, Ljungqvist A and Moberger G. Classification of bladder tumours based on the cellular pattern. Preliminary report of a clinical-pathological study of 300 cases with a minimum follow-up of eight years. Acta Chir Scand 1965; 130: 371-378.
[23] Mostofi FK, Sobin LH, Torloni H. Histologic typing of urinary bladder tumors. In: editors. World Health Organization. Geneva: p. 1973.
[24] Pauwels RP, Schapers RF, Smeets AW, Debruyne FM and Geraedts JP. Grading in superficial bladder cancer. (1). Morphological criteria. Br J Urol 1988; 61: 129-134.
[25] Epstein JI, Amin MB, Reuter VR and Mostofi FK. The World Health Organization/International Society of Urological Pathology consensus classification of urothelial (transitional cell) neoplasms of the urinary bladder. Bladder Consensus Conference Committee. Am J Surg Pathol 1998; 22: 1435-1448.
[26] Ooms EC, Anderson WA, Alons CL, Boon ME and Veldhuizen RW. Analysis of the performance of pathologists in the grading of bladder tumors. Hum Pathol 1983; 14: 140-143.
[27] Ooms EC, Kurver PH, Veldhuizen RW, Alons CL and Boon ME. Morphometric grading of bladder tumors in comparison with histologic grading by pathologists. Hum Pathol 1983; 14: 144-150.
[28] Tosoni I, Wagner U, Sauter G, Egloff M, Knonagel H, Alund G, Bannwart F, Mihatsch MJ, Gasser TC and Maurer R. Clinical significance of interobserver differences in the staging and grading of superficial bladder cancer. BJU Int 2000; 85: 48-53.
[29] Cheng L, Neumann RM, Nehra A, Spotts BE, Weaver AL and Bostwick DG. Cancer heterogeneity and its biologic implications in the grading of urothelial carcinoma. Cancer 2000; 88: 1663-1670.
[30] Campbell PA, Conrad RJ, Campbell CM, Nicol DL and MacTaggart P. Papillary urothelial neoplasm of low malignant potential: reliability of diagnosis and outcome. BJU Int 2004; 93: 1228-1231.
[31] Cheng L, Neumann RM, Weaver AL, Cheville JC, Leibovich BC, Ramnani DM, Scherer BG, Nehra A, Zincke H and Bostwick DG. Grading and staging of bladder carcinoma in transurethral resection specimens. Correlation with 105 matched cystectomy specimens. Am J Clin Pathol 2000; 113: 275-279.
[32] Engers R. Reproducibility and reliability of tumor grading in urological neoplasms. World J Urol 2007; 25: 595-605.
[33] Eble JN, Sauter G, Epstein JL and Sesterhenn IA. World Health Organization Classification of Tumours. Lyon, France: IARC Press, 2004.
[34] Cao D, Vollmer RT, Luly J, Jain S, Roytman TM, Ferris CW and Hudson MA. 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.

[35] Kruger S, Thorns C, Bohle A and Feller AC. Prognostic significance of a grading system considering tumor heterogeneity in muscle-invasive urothelial carcinoma of the urinary bladder. Int Urol Nephrol 2003; 35: 169-173.

[36] Rubben H, Bubenzer J, Bokenkamp K, Lutzeyer W and Rathert P. Grading of transitional cell tumours of the urinary tract by urinary cytology. Urol Res 1979; 7: 83-91.

[37] Vom Dorp F, Pal P, Tschirdewahn S, Rossi R, Borgermann C, Schenck M, Becker M, Szarvas T, Hakenberg OW and Rubben H. Correlation of pathological and cytological-cytometric grading of transitional cell carcinoma of the urinary tract. Urol Int 2011; 86: 36-40.

[38] Wojcik EM, Miller MC, O'Dowd GJ and Veltri RW. Value of computer-assisted quantitative nuclear grading in differentiation of normal urothelial cells from low and high grade transitional cell carcinoma. Anal Quant Cytol Histol 1998; 20: 69-76.

[39] Raab SS, Lenel JC and Cohen MB. Low grade transitional cell carcinoma of the bladder. Cytologic diagnosis by key features as identified by logistic regression analysis. Cancer 1994; 74: 1621-1626.

[40] Renshaw AA, Nappi D and Weinberg DS. Cytology of grade 1 papillary transitional cell carcinoma. A comparison of cytologic, architectural and morphometric criteria in cystoscopically obtained urine. Acta Cytol 1996; 40: 676-682.