Interobserver Agreement of Confocal Laser Endomicroscopy for Bladder Cancer

Timothy C. Chang, MS,¹ Jen-Jane Liu, MD,¹ Shelly T. Hsiao, BA,² Ying Pan, PhD,¹ Kathleen E. Mach, PhD,¹ John T. Leppert, MD, MS,^{1,2} Jesse K. McKenney, MD,³ Robert V. Rouse, MD,^{2,3} and Joseph C. Liao, MD^{1,2}

Abstract

Background and Purpose: Emerging optical imaging technologies such as confocal laser endomicroscopy (CLE) hold promise in improving bladder cancer diagnosis. The purpose of this study was to determine the interobserver agreement of image interpretation using CLE for bladder cancer.

Methods: Experienced CLE urologists (n=2), novice CLE urologists (n=6), pathologists (n=4), and nonclinical researchers (n=5) were recruited to participate in a 2-hour computer-based training consisting of a teaching and validation set of intraoperative white light cystoscopy (WLC) and CLE video sequences from patients undergoing transurethral resection of bladder tumor. Interobserver agreement was determined using the κ statistic.

Results: Of the 31 bladder regions analyzed, 19 were cancer and 12 were benign. For cancer diagnosis, experienced CLE urologists had substantial agreement for both CLE and WLC + CLE (90%, κ 0.80) compared with moderate agreement for WLC alone (74%, κ 0.46), while novice CLE urologists had moderate agreement for CLE (77%, κ 0.55), WLC (78%, κ 0.54), and WLC + CLE (80%, κ 0.59). Pathologists had substantial agreement for CLE (81%, κ 0.61), and nonclinical researchers had moderate agreement (77%, κ 0.49) in cancer diagnosis. For cancer grading, experienced CLE urologists had fair to moderate agreement for CLE (68%, κ 0.64), WLC (74%, κ 0.67), and WLC + CLE (53%, κ 0.33), as did novice CLE urologists for CLE (53%, κ 0.39), WLC (66%, κ 0.50), and WLC + CLE (61%, κ 0.49). Pathologists (65%, κ 0.55) and nonclinical researchers (61%, κ 0.56) both had moderate agreement for CLE in cancer grading.

Conclusions: CLE is an adoptable technology for cancer diagnosis in novice CLE observers after a short training with moderate interobserver agreement and diagnostic accuracy similar to WLC alone. Experienced CLE observers may be capable of achieving substantial levels of agreement for cancer diagnosis that is higher than with WLC alone.

Introduction

PROBE-BASED CONFOCAL LASER ENDOMICROSCOPY (CLE) is an emerging optical imaging technology that enables endoscopic microscopy of mucosal lesions, with dynamic, subsurface imaging of tissue microarchitecture and cellular features. The technology has been applied as an adjunct to standard white light endoscopy in the respiratory¹ and gastrointestinal tracts,^{2–4} and recently in the urinary tract.^{5–8} Particularly for bladder cancer, a new generation of imaging technologies such as CLE may augment the diagnostic accuracy of white light cystoscopy (WLC) through improved visualization of flat lesions, differentiation of benign from neoplastic tumors, and delineation of the tumor boundaries.⁹

Based on the well-established principle of confocal microscopy,^{10,11} CLE is performed using a fiberoptic, sterilizable probe that fits within the working channel of a standard cystoscope. Optical sectioning of the tissue of interest with micron-scale resolution is achieved using a 488 nm laser as the light source and fluorescein, an FDA-approved drug that may be administered intravesically or intravenously, as the contrast agent. Previously, we demonstrated the feasibility of using CLE to obtain real-time *in vivo* images of bladder tumors⁶ and developed diagnostic criteria for grading bladder tumors using CLE.⁵

Interobserver agreement studies are useful in determining the subjective variation in interpretation among observers in analyzing images.¹² Methods that demonstrate higher levels of agreement between observers are deemed more reliable.¹² In

¹Department of Urology, Stanford University School of Medicine, Stanford, California.

²Veterans Affairs Palo Alto Health Care System, Palo Alto, California.

³Department of Pathology, Stanford University School of Medicine, Stanford, California.

Chang TC, Liu JJ, Hsiao ST, Pan Y, Mach KE, McKenney J, Rouse R, and JC Liao presented a poster, Interobserver agreement and accuracy of confocal laser endomicroscopy for *in vivo* diagnosis of bladder cancer, at the Annual Meeting of the American Urological Association, Atlanta, GA, May 2012.

disciplines that require subjective interpretation of diagnostic imaging such as pathology^{13,14} and radiology,¹⁵ interobserver studies are commonly applied to determine reproducibility of the results from one observer to another. Interobserver agreement studies for CLE have been studied in the gastrointestinal tract, which have ranged from moderate to good agreement in the diagnosis of colorectal cancer¹⁶ to substantial and almost perfect agreement in Barrett's esophagus.¹⁷

Interobserver agreement of CLE in the urinary tract has not been previously examined, and can assess the reliability of the technology between observers and evaluate the adoptability of the CLE in novice users. The aim of this study was to determine the interobserver agreement and diagnostic accuracy of CLE for bladder cancer.

Methods

Observer recruitment

The study was approved by the Stanford University Institutional Review Board and the Veterans Affairs Palo Alto Health Care System (VAPAHCS) Research and Development Committee. The study consisted of four observer groups. The first was an experienced CLE urologist group consisting of a board certified urologist and a urology chief resident involved in the protocol development of CLE for the urinary tract. The remaining three groups were novice CLE observers with no previous experience with CLE, consisting of six board certified urologists, four pathologists, and five nonclinical researchers. The novice CLE urologists were expert WLC users, but the pathologists and nonclinical researchers had no experience with WLC. The nonclinical researchers ranged from an undergraduate student to PhD scientists and engineers. All groups participated in identical training sessions.

Teaching and validation set

The observers participated in a 2-hour, computer-based training that consisted of separate teaching and validation sets that featured intraoperative WLC and CLE videos from selected patients undergoing transurethral resection of bladder tumor at the VAPAHCS from 2008 to 2011.

The computer-based, interactive training was created in a website format compatible with common internet browsers for easy access and future scalability. The observers were first introduced to background information on bladder cancer and CLE technology (Fig. 1A). Using still images and video sequences, the observers were instructed to identify three microarchitectural features (flat *vs* papillary, tissue organization, and vascularity) and three cellular features (morphology, cohesiveness, and cellular borders) of benign and pathologic urothelium using CLE. Diagnostic criteria (Table 1) were developed that associated the six features with benign or neoplastic urothelium, and the observers were instructed to categorize each sequence as benign, low grade (LG), or high grade (HG) cancer. Fifteen CLE video sequences were provided in a teaching set for the observers to iteratively practice diagnosing the sequences (Fig. 1B).

After the teaching set, the observers proceeded to the validation set to evaluate 32 CLE video sequences consisting of 12 benign, 9 LG, and 11 HG images (Fig. 1C). The benign sequences were from biopsy-confirmed normal mucosa, inflammation, and papilloma; LG sequences were

from LG papillary tumors; and HG sequences were from HG papillary tumors and carcinoma-*in-situ*. The observers were able to pause and review each video clip frame-by-frame. All observers were blinded to patient history and final pathology and were asked to diagnose and grade each clip as benign, LG, or HG. Upon completing the 32 CLE sequences, the four pathologists and five nonclinical researchers concluded the study. The experienced CLE urologists and novice CLE urologists were asked to further diagnose an additional set of 32 corresponding WLC images in different order, followed by a third and final set of 32 images where the CLE and WLC were shown together. The data were gathered to determine the interobserver agreement and diagnostic accuracy.

From the 32 sequences reviewed by the observers, one of the original HG sequences was excluded because of a discrepancy in the final pathologic diagnosis. The resulting 31 responses of benign, LG, or HG from each of the observers for the 31 images with correlating histopathologic information were used to generate the interobserver agreement and diagnostic accuracy data analyses. For cancer diagnosis, all 31 responses were used. For cancer grading, however, a subset of the 19 cancer (9 LG and 10 HG) sequences found on histopathology were used.

In addition to diagnosing and grading each of the sequences as the other groups did, the experienced CLE urologist group was asked to also identify the six key CLE features (Table 1) for each sequence.

Statistical analysis

Interobserver agreement was assessed using the Fleiss κ statistic. The description for κ statistic developed by Landis and Koch with 0.00 to 0.20 as slight, 0.21 to 0.40 as fair, 0.41 to 0.60 as moderate, 0.61 to 0.80 as substantial, and 0.81 to 1.00 as almost perfect levels of agreement was used in this study.¹⁸

To determine diagnostic accuracy, a dedicated pathologist (R.V.R.) who was blinded to the clinical history and not included in the pathologist group reviewed all the histopathologic slides corresponding to each bladder lesion. Sensitivity and specificity was calculated using the histopathologic results as the standard.

Results

Table 2 shows the percent agreement and κ statistic for the two experienced CLE urologists for each of the six features. Tissue organization, vascular features, cellular morphology, and cellular borders had substantial levels of agreement. The experienced CLE urologists had moderate agreement in the ability to determine flat *vs* papillary using CLE, while they had a fair level of agreement in characterizing cellular cohesiveness.

Table 3 shows the interobserver agreement and diagnostic accuracy for cancer diagnosis. Experienced CLE urologists had substantial to almost perfect levels of agreement for CLE and WLC+CLE, which were both greater than WLC alone. The novice CLE urologists had moderate levels of agreement for all groups. The sensitivity and specificity for all groups were between 73 and 89%.

Table 4 shows the interobserver agreement and diagnostic accuracy for cancer grading. All groups had decreased levels of agreement compared with cancer diagnosis except for nonclinical researchers who had a slightly increased κ for

FIG 1. Computer-based, interactive training module for confocal laser endomicroscopy of the bladder. **(A)** Confocal laser endomicroscopy (CLE) training: Observers were trained to identify six key CLE features. **(B)** Teaching set: Teaching set consisted of 15 video sequences to iteratively practice diagnosis and grading of CLE sequences. **(C)** Validation set: All observer groups reviewed and diagnosed CLE sequences. The experienced CLE and novice CLE urologist groups continued on to diagnose two additional sets of white light cystoscopy (WLC) and WLC+CLE images.

grading compared with cancer diagnosis (0.56 *vs* 0.49). There was also an unexpected decrease in κ for experienced CLE urologists from WLC to WLC+CLE (0.67 to 0.33). For LG cancer, experienced CLE urologists had increased sensitivity and specificity with the addition of CLE to WLC (sensitivity 64%, specificity 81%) compared with standard WLC alone (sensitivity 55%, specificity 50%).

Novice CLE urologists showed an increase in specificity with the addition of CLE to WLC as well (specificity 65%) compared with WLC alone (specificity 54%). For LG, the novice groups all had similar diagnostic accuracy using CLE alone (sensitivity 49-52%, specificity 66-75%). For HG cancer, experienced CLE urologists showed an increase in sensitivity for WLC+CLE (sensitivity 69%) compared with WLC (sensitivity 50%) while maintaining specificity (specificity 73% for both) with the addition of CLE to standard WLC. For novice CLE

urologists, there was an increase in sensitivity but a decrease in specificity with the addition of CLE to standard WLC.

Discussion

Our single-center study indicates that CLE image interpretation of bladder cancer is adoptable by novice observers through a 2-hour computer-based interactive training. We created a training session centered on guiding observers to identify six key CLE features in analyzing bladder cancer microarchitecture and cellular morphology. There were substantial levels of agreement for four of the six features, suggesting that these features can be identified reliably with proper training. The moderate level of agreement for the flat *vs* papillary feature was not unexpected because it is a feature more easily identified on a macroscopic level by WLC than on a

TABLE 1. DIAGNOSIS TABLE

	<i>Benign</i>			<i>Cancer</i>		
	<i>Normal</i>	<i>Papilloma</i>	<i>Inflammmatory</i>	<i>Low grade</i>	<i>High grade</i>	
					<i>Papillary</i>	<i>CIS</i>
Architectural						
Flat <i>vs</i> papillary	Flat	Papillary	Flat	Papillary	Papillary	Flat
Organization	Organized	Organized, normal thickness	Loose cells in LP	Organized, increased thickness	Disorganized	Disorganized
Vascular features	Capillary network in LP	Fibrovascular stalk	n/a	Fibrovascular stalk	Tortuous vessels in fibrovascular stalk	n/a
Cellular						
Morphology	Monomorphic	Monomorphic	Small, monomorphic	Monomorphic	Pleomorphic	Pleomorphic
Cohesiveness	Cohesive	Cohesive	Small, clustered cells	Cohesive	Not cohesive	Not cohesive
Borders	Distinct	Distinct	Distinct	Distinct	Indistinct	Indistinct

microarchitectural level. The cohesiveness feature had a fair level of agreement, suggesting that further refinement of the criteria may improve the reliability of this feature.

Once the observers were trained to identify these features, they were introduced to a table correlating the six features to bladder cancer diagnoses. This table was developed based on the 2004 World Health Organization (WHO) Classification of Bladder Tumours,¹⁹ and further refinement of the table is expected as the technology matures and is adopted by additional end users. Because most of the observer groups with the exception of the pathologist group do not diagnose bladder cancer routinely using microscopy, the table served as a useful guide for novice observers, particularly nonclinical researchers, to diagnose bladder cancer.

The interobserver agreement and diagnostic accuracy results were reported separately for CLE, WLC, and WLC + CLE. Data on CLE alone were useful in assessing the technology itself and observing its stand-alone performance without the bias of WLC. The WLC information provided baseline data, because it is the standard imaging modality for bladder cancer. The practical use of CLE in the clinical setting, however, would involve WLC for the initial survey of the bladder and guiding of the CLE probe to the area of interest. Thus, the WLC + CLE data provide the most clinically relevant information. Because pathologists and nonclinical researchers had no previous training in WLC, they were only asked to review CLE alone.

The ability of novice observers to learn CLE image interpretation was demonstrated by the performance of the various groups after a 2-hour training session. For cancer diagnosis (Table 3), novice CLE urologists had similar moderate levels of agreement and diagnostic accuracy for CLE, WLC, and WLC + CLE, indicating that the 2-hour training was sufficient for CLE to reliably diagnose cancer with diagnostic accuracy comparable to standard WLC. Interestingly, nonclinical researchers who had no previous CLE or clinical experience also had a moderate level of agreement for cancer diagnosis, similar to the other novice CLE observer groups, while also maintaining similar levels of diagnostic accuracy. Thus, both the novice CLE urologists demonstrated compa-

table agreement for WLC and CLE, while concurrently, nonclinical researchers with no previous clinical experience demonstrated similar levels of agreement for CLE as other clinically trained novice groups. These results provide evidence on the relative ease in training the novice observers to interpret CLE images. It is notable that these results are in line with the moderate levels of interobserver agreement reported for CLE imaging of colorectal neoplasia.¹⁶

Experienced CLE urologists obtained substantial to nearly perfect levels of agreement for CLE and WLC + CLE with κ of 0.80 while also maintaining the level of sensitivity and specificity compared with WLC. CLE outperformed WLC for the experienced CLE urologists group with respect to interobserver agreement. The result suggests that for cancer diagnosis, CLE may provide added value compared with WLC because it has greater interobserver reliability without sacrificing diagnostic accuracy. In short, the results indicate that a 2-hour training session may enable novice users to achieve similar levels of interobserver reliability and diagnostic accuracy as WLC for cancer diagnosis, while greater reliability is seen for CLE compared with WLC in experienced CLE urologists.

In regard to cancer grading, the interobserver agreement for cancer grading was in the fair to moderate range for CLE, WLC, and WLC + CLE, which was generally lower than for cancer diagnosis. No clear patterns were noted on diagnostic

TABLE 2. INTEROBSERVER AGREEMENT OF FEATURES OBSERVED ON CONFOCAL LASER ENDOMICROSCOPY

	<i>p_a</i>	κ (95% CI)
Microarchitectural		
Flat <i>vs</i> papillary	77%	0.54 (0.19–0.89)
Organization	87%	0.70 (0.35–1.00)
Vascular features	81%	0.74 (0.49–0.99)
Cellular		
Morphology	84%	0.66 (0.31–1.00)
Cohesiveness	74%	0.33 (–0.03–0.68)
Borders	87%	0.74 (0.38–1.00)

p_a = percent agreement; CI = confidence interval.

TABLE 3. INTEROBSERVER AGREEMENT AND DIAGNOSTIC ACCURACY FOR CANCER DIAGNOSIS

	CLE		WLC		WLC + CLE	
Interobserver agreement	p _a	κ (95% CI)	p _a	κ (95% CI)	p _a	κ (95% CI)
Experienced CLE urologists	90%	0.80 (0.45–1.00)	74%	0.46 (0.10–0.81)	90%	0.80 (0.45–1.00)
Novice CLE urologists	77%	0.55 (0.46–0.64)	78%	0.54 (0.45–0.63)	80%	0.59 (0.50–0.68)
Pathologists	81%	0.61 (0.47–0.76)	-	-	-	-
Nonclinical researchers	77%	0.49 (0.38–0.60)	-	-	-	-
Diagnostic accuracy	Sn	Sp	Sn	Sp	Sn	Sp
Experienced CLE urologists	84%	88%	89%	83%	89%	88%
Novice CLE urologists	75%	81%	86%	79%	89%	83%
Pathologists	84%	81%	-	-	-	-
Nonclinical researchers	89%	73%	-	-	-	-

CLE=confocal laser endomicroscopy; WLC=white light cystoscopy; CI=confidence interval; p_a=percent agreement; Sn=sensitivity; Sp=specificity.

accuracy. The results suggest that CLE may not be as reliable for cancer grading when compared with cancer diagnosis. The interobserver agreement for cancer grading using CLE is similar to pathology literature using the 2004 WHO classification system. May and associates¹⁴ reported κ of 0.30 to 0.52 for interobserver agreement and van Rhijn and colleagues¹³ reported κ of 0.14 to 0.58 and 0.55 to 0.81 for interobserver and intraobserver agreement, respectively. Moreover, a comparison of the cancer grading results in our study derived from the electronic medical records (by multiple pathologists) with the results of a single pathologist (R.V.R.) with expertise in bladder cancer showed a κ of 0.58. These findings illustrate the inherent challenges of bladder cancer grading with CLE or standard pathology.

Our study has several limitations. First, our interobserver agreement for cancer grading analysis was a subset analysis of the confirmed cancers from the original data set. When reviewing the sequences, the observers were given the option to grade the sequences as benign, LG, or HG, rather than simply LG or HG. The additional choice may have contributed to the overall lower interobserver agreement compared with cancer diagnosis. Nevertheless, the occurrences were few, and as the data analysis reflects a more clinically relevant and practical

scenario of clinicians grading an unknown lesion, the study was designed accordingly. Second, there may be selection bias of the CLE video sequences, which were edited offline and chosen nonrandomly by a member of our team with the subjective criteria of image quality (fair to good) and roughly equal distribution of benign, LG, and HG lesions. This person did not participate as an observer for the study. Third, there may be recall bias from the experienced CLE urologists who acquired the original CLE sequences. The use of an independent member from our team to select the images and video sequences mitigates, but does not eliminate, this potential bias. Fourth, an inherent limitation of using CLE for bladder cancer diagnosis is the reliance on microarchitectural and cellular features, whereas pathology uses additionally nuclear morphology (eg, size, mitotic figures). Nuclear features are not routinely seen under CLE, as fluorescein is used as the contrast agent, which stains the extracellular matrix nonspecifically.^{5,6}

Overall, novice CLE observers demonstrated the ability to use CLE as an adjunct to WLC to diagnose bladder cancer after a brief training. Nonclinical researchers with no clinical training were able to diagnose bladder cancer to a comparable level as clinically trained novice CLE observers, highlighting the adoptability and translatability of the CLE technology to a

TABLE 4. INTEROBSERVER AGREEMENT AND DIAGNOSTIC ACCURACY FOR CANCER GRADING

	CLE		WLC		WLC + CLE	
Interobserver agreement	p _a	κ (95% CI)	p _a	κ (95% CI)	p _a	κ (95% CI)
Experienced CLE urologists	68%	0.64 (0.41–0.87)	74%	0.67 (0.35–0.98)	53%	0.33 (–0.03–0.70)
Novice CLE urologists	53%	0.39 (0.31–0.47)	66%	0.50 (0.41–0.60)	61%	0.49 (0.41–0.57)
Pathologists	65%	0.55 (0.43–0.68)	-	-	-	-
Nonclinical researchers	63%	0.56 (0.47–0.65)	-	-	-	-
Diagnostic accuracy	Sn	Sp	Sn	Sp	Sn	Sp
Low grade						
Experienced CLE urologists	50%	94%	55%	50%	64%	81%
Novice CLE urologists	52%	75%	59%	54%	55%	65%
Pathologists	50%	66%	-	-	-	-
Nonclinical researchers	49%	73%	-	-	-	-
High grade						
Experienced CLE urologists	75%	64%	50%	73%	69%	73%
Novice CLE urologists	46%	74%	44%	76%	54%	67%
Pathologists	50%	66%	-	-	-	-
Nonclinical researchers	63%	60%	-	-	-	-

CLE=confocal laser endomicroscopy; WLC=white light cystoscopy; CI=confidence interval; p_a=percent agreement; Sn=sensitivity; Sp=specificity.

wide range of novice users. Our results indicate that further studies are warranted that refine the CLE features used in diagnosing and grading bladder cancer as well as multicenter studies that validate the translatability of this study.

Future directions include prospective multicenter studies to investigate the overall clinical utility of CLE, as well as cost-benefit analyses, which will be necessary for widespread adoption of CLE for bladder cancer diagnosis and grading. In addition, CLE, as a microscopic imaging modality, may be combined with other new macroscopic imaging technologies (ie. photodynamic diagnosis and narrow band imaging) already in clinical use to improve the overall optical diagnosis of bladder cancer.⁹

Conclusion

CLE is an adoptable technology for novice CLE observers after a training session with moderate interobserver agreement and diagnostic accuracy similar to WLC alone. Experienced CLE observers may be capable of achieving substantial levels of agreement for cancer diagnosis that is markedly higher than with WLC alone. Fair to moderate levels of agreement are achieved for cancer grading, although literature suggests the variability may in part be attributable to the grading classification system.

Acknowledgments

The authors would like to acknowledge our colleagues who participated in the study (H.G., J.D.B., C.V.C., M.E., E.S., M.S., D.B., R.M., C.Z., H.R., J.H., S.O., and A.S.). We also thank Mauna Kea Technologies for technical support and helpful discussions. This work was supported in part by the U.S. National Institutes of Health (NIH) R01 CA160986 (J.C.L.).

Disclosure Statement

No competing financial interests exist.

References

- Thiberville L, Salaün M, Lachkar S, et al. Human in vivo fluorescence microimaging of the alveolar ducts and sacs during bronchoscopy. *Eur Respir J* 2009;33:974–985.
- Dunbar KB, Okolo P 3rd, Montgomery E, Canto MI. Confocal laser endomicroscopy in Barrett's esophagus and endoscopically inapparent Barrett's neoplasia: A prospective, randomized, double-blind, controlled, crossover trial. *Gastrointest Endosc* 2009;70:645–654.
- Pech O, Rabenstein T, Manner H, et al. Confocal laser endomicroscopy for in vivo diagnosis of early squamous cell carcinoma in the esophagus. *Clin Gastroenterol Hepatol* 2008;6:89–94.
- Goetz M, Kiesslich R, Dienes HP, et al. In vivo confocal laser endomicroscopy of the human liver: A novel method for assessing liver microarchitecture in real time. *Endoscopy* 2008;40:554–562.
- Wu K, Liu JJ, Adams W, et al. Dynamic real-time microscopy of the urinary tract using confocal laser endomicroscopy. *Urology* 2011;78:225–231.
- Sonn GA, Jones SN, Tarin TV, et al. Optical biopsy of human bladder neoplasia with in vivo confocal laser endomicroscopy. *J Urol* 2009;182:1299–1305.
- Sonn GA, Mach KE, Jensen K, et al. Fibered confocal microscopy of bladder tumors: An ex vivo study. *J Endourol* 2009;23:197–201.
- Adams W, Wu K, Liu JJ, et al. Comparison of 2.6- and 1.4-mm imaging probes for confocal laser endomicroscopy of the urinary tract. *J Endourol* 2011;25:917–921.
- Liu JJ, Droller MJ, Liao JC. New optical imaging technologies for bladder cancer: Considerations and perspectives. *J Urol* 2012;188:361–368.
- Robinson JP. Principles of confocal microscopy. In: Darzynkiewicz, Z, ed. *Methods in Cell Biology*. Waltham, MA: Academic Press. 2001, pp 89–106.
- Helmchen F. Miniaturization of fluorescence microscopes using fibre optics. *Exp Physiol* 2002;87:737–745.
- Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. *Fam Med* 2005;37:360–363.
- van Rhijn BW, van Leenders GJ, Ooms BC, et al. The pathologist's mean grade is constant and individualizes the prognostic value of bladder cancer grading. *Eur Urol* 2010;57:1052–1057.
- May M, Brookman-Amissah S, Roigas J, et al. Prognostic accuracy of individual uropathologists in noninvasive urinary bladder carcinoma: A multicentre study comparing the 1973 and 2004 World Health Organisation classifications. *Eur Urol* 2010;57:850–858.
- Tekes A, Kamel I, Imam K, et al. Dynamic MRI of bladder cancer: Evaluation of staging accuracy. *AJR Am J Roentgenol* 2005;184:121–127.
- Gómez V, Buchner AM, Dekker E, et al. Interobserver agreement and accuracy among international experts with probe-based confocal laser endomicroscopy in predicting colorectal neoplasia. *Endoscopy* 2010;42:286–291.
- Wallace MB, Sharma P, Lightdale C, et al. Preliminary accuracy and interobserver agreement for the detection of intraepithelial neoplasia in Barrett's esophagus with probe-based confocal laser endomicroscopy. *Gastrointest Endosc* 2010;72:19–24.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
- Montironi R, Lopez-Beltran A. The 2004 WHO classification of bladder tumors: A summary and commentary. *Int J Surg Pathol* 2005;13:143–153.

Address correspondence to:
Joseph C. Liao, MD
Department of Urology
Stanford University School of Medicine
300 Pasteur Dr., Room S-287
Stanford, CA 94305-5118
E-mail: jlliao@stanford.edu

Abbreviations

CLE = confocal laser endomicroscopy
HG = high grade
LG = low grade
VAPAHCS = Veterans Affairs Palo Alto Health Care System
WLC = white light cystoscopy