Interobserver Variability in the Diagnosis of Ulcerative Colitis-Associated Dysplasia by Telepathology

Robert D. Odze, M.D., F.R.C.P.C., John Goldblum, M.D., Amy Noffsinger, M.D., Nada Alsaigh, M.D., Lyndo A. Rybicki, M.S., Franz Fogt, M.D., M.R.C. Path.

Departments of Pathology, University of Pennsylvania Presbyterian Medical Center (FF), Philadelphia, Pennsylvania, University of Cincinnati Medical Center (AN), Cincinnati, Ohio, Dianon Systems, Inc. (NA), Stratford; and the Brigham and Women's Hospital (RDO), Boston, Massachusetts; and the Departments of Anatomic Pathology and Biostatistics and Epidemiology, The Cleveland Clinic Foundation (JG, LAR), Cleveland, Ohio

Telepathology (TP) is the practice of remote diagnostic consultation of electronically transmitted, static, digitalized images. The diagnostic efficacy of TP-based consultation services has not been widely tested. Dysplasia that arises in association with chronic ulcerative colitis (CUC) is, at present, the most important marker of an increased risk of malignancy in patients with this disease. Unfortunately, dysplasia is difficult to diagnose histologically and, as a result, suffers from a significant degree of intra- and interobserver variability. Furthermore, it is often necessary to obtain expert consultation of potential CUC-associated dysplasia cases before treatment. Therefore, the aim of this study was to evaluate the utility and interobserver variability of diagnosing dysplasia in CUC with the use of TP. Static, electronically transmitted, digitalized images of 38 CUC cases with areas considered negative, indefinite, or positive for dysplasia (low or high grade) were evaluated independently by four gastrointestinal pathologists. All cases were then graded by each of the pathologists by light-microscopic examination of the hematoxylin and eosin–stained glass slides. The degree of interobserver variability was determined by κ statistics. Overall, there was a fair degree of agreement (κ = 0.4) among the four reviewing pathologists after analysis of the digitalized images. The poorest level of agreement was in the indefinite and low-grade dysplasia categories. Grouping together several diagnostic categories (for instance, indefinite and low-grade dysplasia, or low-grade dysplasia and high-grade dysplasia) had no effect on the overall level of agreement. The degree of variability in interpretation of glass slides was slightly better (κ = 0.43) but still remained fair. After reviewing all cases by glass slide analysis, the diagnosis was changed in 38% of the slides; in the majority of these, the grade of dysplasia was increased. Use of TP for consultation in CUC-associated dysplasia has a moderate level of interobserver agreement. Because of a variety of technical reasons, diagnoses rendered by evaluation of digitalized images tended to be of a lower grade than that observed after a review of the glass slides.

KEY WORDS: Dysplasia, Inflammatory bowel disease, Static-image telepathology, Ulcerative colitis.

Mod Pathol 2002;15(4):379–386

LIPPINCOTT WILLIAMS & WILKINS

Unauthorized Use Prohibited

Telepathology (TP) is the practice of remote diagnostic consultation of either electronically transmitted, static, digitized images or real-time pictures obtained with the use of remote robotic microscopes (1, 2). One of the major advantages of digital camera–based systems is that they are widely available and may be used with different types of microscopes (2). Furthermore, digital photographs can be stored and sent via e-mail to a single or multiple consultants, thereby avoiding financial costs associated with integrated site-specific robotic microscope and computer units (3, 4). The main advantage of stationary robotic units is that real-time interaction is possible simply by moving the light-microscope stage from the remote site (2, 5). However, this system requires that both the physician who is requesting a consult and the consultant possess the necessary hardware (5). Unfortunately, this equipment is costly and is usually connected to only one consultation center, which limits the number and availability of potential consultants (2).

Copyright © 2002 by The United States and Canadian Academy of Pathology, Inc.
VOL. 15, NO. 4, P. 379, 2002 Printed in the U.S.A.
Date of acceptance: December 20, 2001.
Address reprint requests to: Robert D. Odze, M.D., FRCP(c), Department of Pathology, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115; e-mail: rdodze@bics.bwh.harvard.edu; fax: 617-277-9015

Telepathology was first used in an intraoperative frozen-section consultation service in Norway in 1989 (6). Since then, other studies have shown that TP can be applied successfully to an intraoperative consultation service (7, 8). Recently, use of TP has expanded to include consultations for routinely processed cases (2–5, 9, 10). In fact, many organizations, such as the Armed Forces Institute of Pathology, now offer diagnostic services using TP, including stationary TP. Furthermore, the International Union Against Cancer has recently proposed using stationary TP as an inexpensive diagnostic aid to help classify tumors (see the Web page of the International Union Against Cancer for details, www.uicc.org/programmes/detection/tpcc.shtml). Unfortunately, the diagnostic efficacy of TP-based pathology consultation services has not been widely tested (2). This applies, in particular, to areas of pathology in which diagnoses are based on accurate and specific architectural and/or cytologic abnormalities and interobserver agreement with the results of conventional microscopic examination is high.

Dysplasia that arises in association with chronic ulcerative colitis (CUC) is defined as unequivocal intraepithelial neoplastic epithelium (11, 12). At present, the histologic diagnosis of dysplasia is the most important marker of an increased risk of malignancy in CUC and is critical in guiding patient management (11, 13). Dysplasia is generally categorized as negative, indefinite, or positive (low or high grade; 11). Because histologic distinction between these categories is often subtle and suffers from a significant degree of intra- and interobserver variability (12, 14), it is often necessary to obtain confirmation of the diagnosis before treatment.

Therefore, the aim of this study was to evaluate the utility and interobserver variability of diagnosing dysplasia in CUC with the use of TP, and to compare it with the interobserver variability obtained with the use of microscopic slides.

LIPPINCOTT WILLIAMS & WILKINS
Unauthorized Use Prohibited

MATERIALS AND METHODS

Analysis of Digitalized Images

Four dedicated gastrointestinal (GI) pathologists were involved in this study. Each pathologist selected 9–10 representative cases of CUC-related dysplasia (low and high grade) or cases considered negative or indefinite for dysplasia from archival mucosal biopsy pathology material at his or her institute of practice. In total, 38 cases were selected. One representative microscopic slide containing the diagnostic area of each case was sent to a fifth (reference) GI pathologist who subsequently established a working diagnosis of each case based on an analysis of the glass slides, using previously published criteria (12). The reference pathologist established a diagnosis of negative for dysplasia (reactive) in 11 cases, of indefinite in 5, of low-grade dysplasia (LGD) in 12, and of high-grade dysplasia (HGD) in 10 cases. The reference pathologist photographed the histologic area of interest using a Twin Cam MX-700 digital camera (Fuji film) with a modified (for use with a microscope) ocular attachment (I. Miller Precision Optical Instruments Inc., Philadelphia, PA). The photographs were stored on a 16-MB SmartMedia disc (Fuji film). Three photographs of each case were taken at different magnifications (high [40×], medium [20×], and low [10×] power) at 1289 and 1024 resolutions. The photographs were then converted to 114 JPEG images ranging from 143 to 864 KB in size. Because of the large number of images in this study, the photographs were stored on a read-only compact disc (650 MB) that was then sent to each of the other four participating pathologists for review. In this manner, the reference pathologist was simulating a method that a consulting pathologist would use to obtain an expert opinion with the use of TP and thus, did not participate in the digitalized image analysis. Each reviewer was asked to establish a diagnosis of negative (reactive), indefinite, or positive for dysplasia (low or high grade) based on an evaluation of each of the three photographs from each case. Ease reviewer was also asked to comment on difficulties encountered on analysis of the digitalized images. All results and comments were sent to the reference pathologist for statistical analysis.

Analysis of Microscopic Slides

In the second part of the study, each of the four reviewers was asked to reevaluate all cases by light-microscopic examination of the hematoxylin and eosin–stained glass slides and to select the most appropriate diagnostic category (negative, indefinite, LGD, HGD). Slides were reviewed at least 1 month after analysis of the digitalized images.

Statistics

The kappa statistic was used to assess interobserver variability based on analysis of digitalized images or by microscopic slides among the four reviewing pathologists; results were compared between the two methods using the z test. The kappa statistic was also used to assess interobserver variability between digitalized images and microscopic slides separately for each of the four reviewing pathologists.

Kappa measures agreement beyond that which is expected by chance alone. κ values of >0.75 indicate excellent agreement beyond chance; values be-

tween 0.40 and 0.75 indicate fair to good agreement beyond chance; and values of <0.40 indicate poor agreement (15). The results of the kappa analysis are summarized as the kappa value, 95% confidence interval, and *P* value from the *z* test, which determines whether agreement is better than that which is expected by chance alone. Individual kappas indicate agreement within individual diagnostic categories, whereas the overall kappa is a weighted average of the individual kappas.

Kappa statistics were calculated for cases using digitalized images and compared with those obtained from nonproblematic cases using the *z* test. Problematic cases are defined as ones in which at least one of the four reviewing pathologists noted that there were difficulties in making a diagnosis based on analysis of the digitalized image.

All analyses were conducted using SAS software (SAS Institute, Cary, NC); $P < .05$ was the criterion for statistical significance.

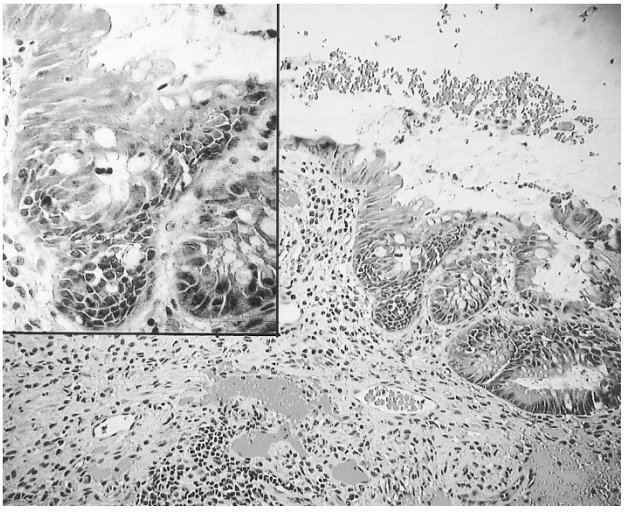

FIGURE 1. Low- and high-power (inset) digitalized images of a case considered negative for dysplasia by all four reviewers.

RESULTS

Interobserver Variability of Digitalized Images

Table 1 provides a summary of the degree of interobserver variability regarding interpretation of digitalized photographs in the 38 study cases. Overall, there was fair agreement ($\kappa = 0.4$, $P < .001$) among the four reviewing pathologists concerning the ability to differentiate negative, indefinite, LGD, and HGD in CUC. Figures 1–6 show reprints of representative digitalized photographs from each diagnostic category. Of the four diagnostic categories, the poorest agreement occurred in the indefinite ($\kappa = 0.18$) and LGD ($\kappa = 0.36$) categories. Cases of negative ($\kappa = 0.51$) or HGD ($\kappa = 0.54$) showed a considerably higher level of agreement.

Separate statistical analyses were performed after grouping together some of the diagnostic categories. However, in these instances, the degree of interobserver variability did not improve significantly and remained in the fair range. For instance, when the categories of indefinite for dysplasia and LGD were grouped together, or when LGD and HGD were combined, the degree of agreement remained fair ($\kappa = 0.43$ and 0.46, respectively; results not shown in tables).

Interobserver Variability of Microscopic Slides

The bottom half of Table 1 provides a summary of the degree of interobserver variability regarding interpretation of microscopic slides. Although the level of agreement among the four reviewing pathologists increased slightly ($\kappa = 0.43$, $P < .001$), it remained fair and was not significantly greater than the level of agreement obtained by analysis of digitalized images ($P = .62$). Once again, the lowest level of agreement was observed in the indefinite category ($\kappa = 0.27$). The highest level of agreement was observed in biopsies considered to be LGD ($\kappa = 0.51$), negative for dysplasia ($\kappa = 0.46$), or HGD ($\kappa = 0.45$).

LIPPINCOTT WILLIAMS & WILKINS
Unauthorized Use Prohibited

TABLE 1. Kappa Indices for Interobserver Agreement among Four Gastrointestinal Pathologists

Category	Kappa	*P*	95% Confidence Interval	Interpretation[a]
Digitalized images				
Negative	0.51	<0.001*	0.38–0.64	Good
Indefinite	0.18	0.008*	0.05–0.31	Poor
LGD	0.36	<0.001*	0.23–0.49	Poor
HGD	0.54	<0.001	0.41–0.67	Good
Overall	0.40	<0.001*	0.32–0.48	Fair
Microscopic slides				
Negative	0.46	<0.001*	0.33–0.59	Fair
Indefinite	0.27	<0.001*	0.14–0.40	Poor
LGD	0.51	<0.001*	0.38–0.64	Good
HGD	0.45	<0.001*	0.32–0.58	Fair
Overall	0.43	<0.001*	0.35–0.50	Fair

LGD, = low-grade dysplasia; HGD, = high-grade dysplasia.
* Significant agreement beyond chance.
[a] Poor: kappa < 0.40, fair to good: 0.40 ≤ kappa ≤ 0.75, excellent: kappa > 0.75.

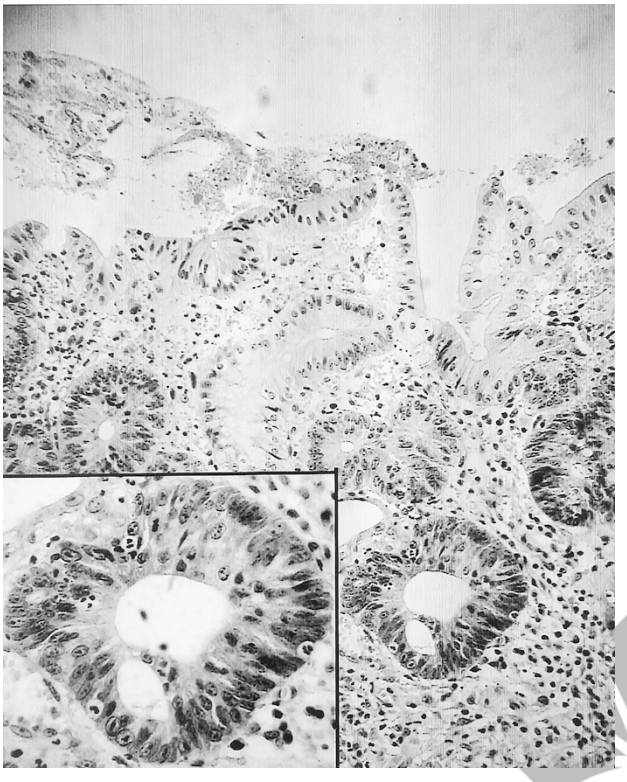

FIGURE 2. Low- and high-power (inset) digitalized images of a case considered negative for dysplasia by one reviewer and indefinite for dysplasia by three. After a review of the microscopic slide, none of the four reviewing pathologists changed their opinion.

FIGURE 3. Low- and high-power (inset) digitalized images of a case considered high-grade dysplasia by all four reviewers by telepathology and slide review.

LIPPINCOTT WILLIAMS & WILKINS
Unauthorized Use Prohibited

When separate statistical analyses were performed after grouping together some of the diagnostic categories, the degree of agreement did not exceed a good level. For instance, when the categories of indefinite for dysplasia and LGD were grouped together, or when LGD and HGD were combined, the degree of overall agreement remained fair to good (κ = 0.41 and 0.50, respectively).

Table 2 provides a summary of the kappa values obtained between all 6 potential pairs of pathologists in interpretation of digitalized images or microscopic slides. The range of kappa values was 0.24–0.53 for analysis of digitalized images and 0.11–061 for evaluation of microscopic slides.

Results of Analysis of Digitalized Images *Versus* Microscopic Slides

Table 3 provides a summary of the level of intraobserver agreement for each of the four reviewing pathologists between analysis of digitalized images and microscopic slides. Intraobserver agreement ranged from κ = 0.22 to 0.68. Interestingly, for each individual pathologist, the highest level of agreement occurred in the negative for dysplasia and, with the exception of one reviewer, the high-grade dysplasia category.

Table 4 provides more detail regarding diagnostic differences between digitalized images and microscopic slides. Line 1 in Table 4 indicates that by analysis of microscopic slides, Pathologist 1 disagreed with his or her digitalized image diagnosis in 7 of 38 (18%) cases. Of these 7 cases, the pathologist changed the diagnosis to a higher diagnostic category (*e.g.,* indefinite to LGD) in 4 cases after evaluation of the glass slides, and in the other 3, it was changed to a lower diagnostic category. The results of the other three pathologists are also noted in Table 3. Interestingly, of the cases in which pathologists changed their diagnoses, the change was to a higher diagnostic category of dysplasia in 57–95% of cases (77% overall).

A summary of the reviewing pathologist's comments regarding potential limitations of evaluating digitalized photographs are indicated in Table 5. Overall, the number of comments was minimal. The most frequently reported limitations included poor resolution of the digitalized photograph and the impression that some of the digitalized images were not representative of the most severe area of atypia. In several instances, the reason for difficulty in evaluating digitalized images was related to the fact that surface epithelium was not clearly visible in the photograph. Some pathologists felt that the histologic areas of interest appeared more atypical when analyzed by light microscopy in comparison

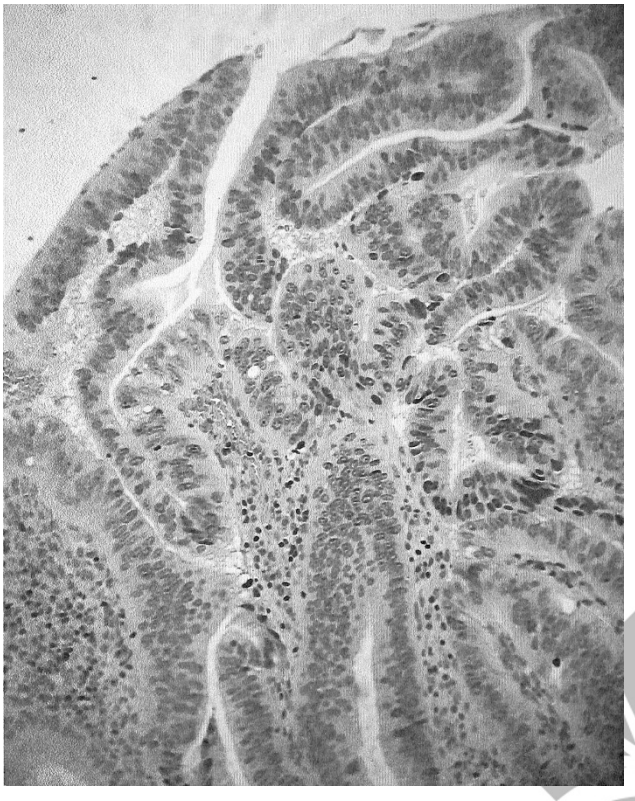

FIGURE 4. All four reviewers considered this high-magnification digitalized image as representative of high-grade dysplasia.

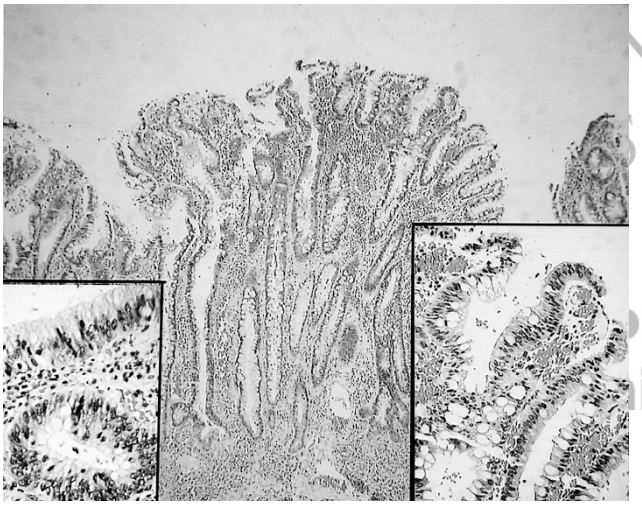

FIGURE 5. Low-, medium- (inset, right) and high-power (inset, left) digitalized images of a case considered negative for dysplasia by one pathologist and as low-grade dysplasia by three pathologists. After a review of the glass slide, one pathologist changed the diagnosis from negative to low-grade dysplasia, one maintained a diagnosis of low-grade dysplasia, and two changed their diagnosis from low- to high-grade dysplasia.

to a review of the digitalized photograph. Other less frequent comments are also noted in this table. Significantly lower interobserver agreement was noted between 16 cases in which problems were noted and 22 cases in which no problems were identified ($\kappa = 0.12$ and 0.59, respectively, $P < .001$).

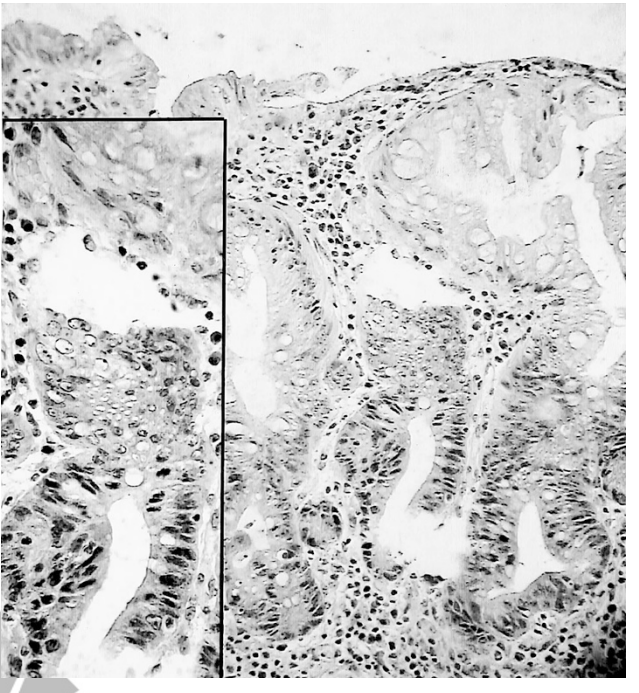

FIGURE 6. These digitalized images of a case were interpreted as negative by one pathologist and indefinite by three pathologists. After glass slide analysis, two pathologists changed their diagnosis to a higher grade (both indefinite to high-grade dysplasia).

DISCUSSION

The widespread availability of computer networks and fast data transmission has resulted in the capability to obtain diagnostic consultations from remote sites (2, 3). For instance, in static TP, digitalized images are sent via e-mail to consultants for review (16). With the growing popularity of this method of consultation, it is important to ensure that diagnostic accuracy remains at a high level. Although TP has been actively discussed among pathologists for >10 years, it is still not widely used despite its many conveniences, such as low cost, rapid consultation turnaround time, and relative ease of use (2). Because of the small number of interobserver TP-based studies, there may be a common belief that the diagnostic accuracy of TP is low. Thus, this study was performed to determine the efficacy of TP in an area of pathology particularly prone to a high level of interobserver variability, and one that important clinical treatment decisions are based on (11, 12, 14). For instance, in CUC, a confirmed diagnosis of dysplasia, particularly high grade, is usually a strong indication for colectomy, whereas cases considered negative or indefinite for dysplasia are normally managed by regular or increased surveillance, respectively (11, 13). In fact, Riddell *et al.* (12), in their original study that standardized the current classification of CUC-associated dysplasia, recommended that a case considered dysplasia by one pathologist should be

NOT
& WILKINS
Unauthorized Use
Prohibited

TABLE 2. Kappa Indices for Interobserver Agreement among Pairs of Pathologists

Category	Kappa Values between Pairs of Pathologists[a]						Overall Kappa Range
	1:2	1:3	1:4	2:3	2:4	3:4	
Digitalized images	0.38	0.52	0.42	0.24	0.28	0.53	0.24–0.53
Microscopic slides	0.43	0.47	0.11	0.61	0.42	0.46	0.11–0.61

[a] Four reviewing pathologists, designated 1–4.

confirmed by another before definitive management.

The results of this study showed that there was an overall fair level of agreement ($\kappa = 0.4$) among four GI pathologists in interpreting dysplasia in CUC by evaluation of electronically transmitted digitalized images. The level of agreement was highest for cases considered negative or HGD. This latter result is not surprising given that "negative for dysplasia" and "HGD" represent two extremes of a continuous spectrum of atypical changes that occur in CUC. Statistically, as summarized by Riddell in an excellent review of the problems that may be encountered in grading dysplasia in CUC (17), when multiple pathologists are asked to impart an absolute diagnosis to a disorder that develops as a gradually increasing spectrum of atypia, then their readings will form a distribution curve with the means representing the ultimate correct diagnosis. Thus, readings at either extreme, such as "negative for dysplasia" or "HGD," can have only one tail to the distribution curve of diagnoses. Therefore, the overall level of disagreement would be expected to be less than for lesions considered either "indefinite for dysplasia" or "LGD."

Of particular significance is the fact that the level of agreement achieved by a review of digitalized images in this study was only slightly lower that that obtained after reviewing the original glass slides. Similar to digitalized image evaluation, agreement was lower for indefinite samples. However, one must be cautious in interpreting these results for the following reasons. First, glass slide review has the advantage of enabling the observer the ability to evaluate tissue reactions both near and distant to the focal atypical area of interest. Second, the study may be biased because of the fact that the number of cases evaluated in each diagnostic category is not necessarily an accurate representation of the actual proportion of cases pathologists may see in clinical practice. However, for both the interobserver and intraobserver analyses, separate kappa values were obtained for each individual diagnostic group.

The level of interobserver agreement by analysis of digitalized images of CUC-related dysplasia in this study is similar, or perhaps even slightly higher,

LIPPINCOTT WILLIAMS & WILKINS

TABLE 3. Kappa Indices for Intraobserver Agreement between Readings by Digitalized Images and Microscopic Slides

Readings by Pathologist	Individual Kappa	*P*	95% Confidence Interval	Interpretation[a]
Pathologist 1				
Negative	0.78	<0.001*	0.47–1.00	Excellent
Indefinite	0.47	0.004*	0.15–0.79	Fair
LGD	0.73	<0.001*	0.41–1.00	Good
HGD	0.91	<0.001*	0.59–1.00	Excellent
Overall	0.50	<0.001*	0.54–0.93	Good
Pathologist 2				
Negative	0.62	<0.001*	0.30–0.93	Good
Indefinite	−0.16	0.310	−0.48–0.15	Poor
LGD	0.01	0.940	−0.30–0.33	Poor
HGD	0.42	<0.010*	0.10–0.74	Fair
Overall	0.22	<0.018*	0.04–0.41	Poor
Pathologist 3				
Negative	0.67	<0.001*	0.35–0.99	Good
Indefinite	0.52	0.001*	0.21–0.84	Good
LGD	0.75	<0.001*	0.44–1.00	Excellent
HGD	0.76	<0.001*	0.43–1.00	Excellent
Overall	0.68	<0.001*	0.49–0.87	Good
Pathologist 4				
Negative	0.52	0.001*	0.21–0.84	Good
Indefinite	0.19	0.250	−0.13–0.50	Poor
LGD	0.37	0.024*	0.05–0.68	Poor
HGD	0.27	0.100	−0.05–0.59	Poor
Overall	0.33	<0.001*	0.14–0.51	Poor

Unauthorized Use Prohibited

LGD, = low-grade dysplasia; HGD, = high-grade dysplasia.

* Significant agreement beyond chance.

[a] Poor: kappa $<$ 0.40, fair to good: 0.40 $\leq$ kappa $\leq$ 0.75, excellent: kappa $>$ 0.75.

TABLE 4. Disagreement between Digitalized Images and Microscopic Slides

Pathologist	No. (%) Cases in which Reviewer Changed Diagnosis after Analysis of Microscopic Slides	Grade Change, (%)	
		Higher	Lower
1	7/38 (18)	4/7 (57)	3/7 (43)
2	22/38 (54)	16/22 (73)	6/22 (27)
3	9/38 (24)	6/9 (67)	3/9 (33)
4	19/38 (50)	18/19 (95)	1/19 (5)
Total	57/152 (38)	44/57 (77)	13/57 (23)

TABLE 5. Summary of Reviewers' Comments Regarding Limitations of Evaluating Digitalized Photographs

Pathologists' Comments	No. Observations ($N = 152$)[a]
Poor resolution of photograph	5
Inappropriate field selection	5
Surface not clearly visible	4
Slide "more atypical" than photograph	4
Lack of clarity at low power	3
Lack of depth perception	1
Poor nuclear detail	1
Poor stain	1

[a] Total number of observations (152) = no. of pathologists (4) × no. of cases (38).

compared with that obtained in other interobserver variability studies based upon microscopic analysis of glass slides (14, 18). For instance, in an interobserver variability study by Dixon *et al.* (14) that included 100 cases, four diagnostic categories (hyperplasia, reactive atypia, LGD, and HGD) and six participating pathologists, the overall level of agreement among all pairs of observers was considered fair ($\kappa = 0.41$), with a wide range of kappa values (0.29–0.58). Similarly, in a study by Melville *et al.* (18) that used 5 experienced pathologists, all of whom analyzed 207 specimens from 85 CUC patients for "no dysplasia, indefinite for dysplasia, LGD, HGD or cancer," the kappa values were generally poor, varying between 0.21 and 0.45 for each pair of observers. Although some studies suggest that there is a much higher level of interobserver agreement in interpretation of CUC-related dysplasia, these studies used a different statistical approach to the interpretation of their data, which makes comparison with our results quite difficult (12, 19, 20). For example, observations made by individual pathologists were compared with that of a numerical group average, a process that statistically tends to minimize discrepancies between observers, as originally pointed out by Melville *et al.* (18). Thus, although interobserver agreement in the diagnosis of dysplasia in CUC is, at best, fair, and needs significant improvement, based on the criteria available at this time, diagnostic consultation of these cases by TP is an acceptable alternative method to glass slide analysis.

This is the first study to evaluate interobserver variability of a difficult diagnostic disease entity by the use of static-image TP. However, other studies have been performed that attempted to validate, or determine, the diagnostic accuracy of the static image TP-based consultation system (3–5, 9, 16). For instance, in a study of 35 GI biopsy specimens evaluated in the form of digitalized images that were transmitted through the internet to one dedicated GI pathologist, concordance between the original slide diagnosis and the consultant's TP diagnosis was 94% (16). In another study involving 200 consecutive cases of prostate needle biopsies analyzed by static TP by one highly experienced observer, Weinstein *et al.* (9) achieved 90.5% diagnostic accuracy. Similarly, several other static-image TP studies, some performed on frozen-section tissues and others on a variety of general surgical pathology cases, have demonstrated a generally high level of diagnostic accuracy, with rates varying between 88–100% (2–5). However, none of these studies evaluated interobserver variability. Nevertheless, the data derived from these, and many other similarly performed studies, support the use of high-resolution TP as an easily obtainable, user-friendly type of technology that yields excellent quality and relatively inexpensive images for rapid consultation (2).

There are three major sources of potential diagnostic error when evaluating digitalized images by TP, as pointed out by Weinstein *et al.* (2) in an extensive current review of TP-based pathology. One type of error occurs as a result of misinterpretation, as often occurs with glass slide analysis as well (5). Other sources of error may be poor image quality or inappropriate field selection (4, 8, 9, 21). Inappropriate field selection is particularly problematic in static TP, where the representative area of interest is usually photographed by one observer and then sent to another for consultation (4). Thus, this system relies heavily on the experience and degree of knowledge of the referring pathologist regarding the important diagnostic criteria of the disorder under scrutiny. In fact, this potential source of error has been cited in 4–8% of static TP-based consultation cases (4, 8, 9). For instance, inappropriate field selection represented the primary reason for an erroneous diagnosis in 50% of misdiagnosed cases in one international referral study of 144 cases by Halliday *et al.* (4).

In our study, the reviewing pathologists cited comments regarding potential limitations of evaluating digitalized images in only 24 (16%) of a potential 152 observations (Table 5). Of the limitations noted, inappropriate field selection and/or an inability to view mucosa distant from the atypical focus accounted for the majority (20%), whereas in another 16% of instances, the reviewing pathologist commented that the surface of the mucosa was not clearly visible. The presence or absence of surface

LIPPINCOTT WILLIAMS & WILKINS
Unauthorized Use Prohibited

epithelial maturation is considered an important diagnostic feature when evaluating mucosal biopsies for dysplasia in CUC(12). Although the initial reference pathologist in our study was an experienced GI pathologist, well schooled in the diagnostic criteria for dysplasia in CUC, inappropriate field selection still remained an important issue and, thus, remains a potential limitation when using static TP consultation services. Other less common comments in this study were related mainly to various parameters of image quality that the reviewing pathologist felt hindered the establishment of an accurate diagnosis. It is reasonable to assume that as the technology related to TP advances with time, and as users become more knowledgeable and better acquainted with the particular operating characteristics of the system, this potential source of error may be minimized.

Finally, another important result of this study relates to the change in diagnostic interpretation noted when the pathologists' digitalized image diagnoses were compared with their glass slide diagnoses. This result may be explained on the basis of the following important reason. As mentioned previously, microscopic evaluation of the glass slides enabled the observers to evaluate atypical areas in the context of the surrounding mucosal changes, a feature that, at least in the case of CUC-related dysplasia, is known to have a significant impact on the final diagnosis (12). Although this reason was not indicated by the pathologists in their comments regarding potential limitations, statements such as "surface not clearly visible" or "slide more atypical than photograph" may be indirectly linked to this concept. Interestingly, in cases for which re-review of the glass slides resulted in a change of diagnosis, the majority were changed to a higher grade. In fact, one pathologist altered 95% of changed diagnoses to a higher grade on review of the glass slides in selected cases. Thus, at least in this study, it appears that the tendency was to underdiagnose dysplasia in CUC (false negative) by the use of TP. However, this is not necessarily a limitation because overdiagnosis of dysplasia in CUC by the use of TP (false positive) might lead to more serious clinical consequences (*e.g.,* unnecessary colectomy) compared with underdiagnosed cases.

In summary, the results of this study suggest that static TP-based consultation services may be used cautiously for evaluating CUC-associated dysplasia. As performed in this study, we recommend that both low- and high-power photomicrographs be used for consultation to minimize the potential limitation of inappropriate and/or insufficient field selection. Pathologists using this system should be aware that there is a tendency to underdiagnose dysplasia in comparison to the results obtained by microscopic analysis of glass slides. Better, more precise, criteria for dysplasia in CUC are needed to increase the degree of agreement among observers in this disorder.

REFERENCES

1. Weinstein RS, Bloom KJ, Rozek LS. Telepathology and the networking of pathology diagnostic servcies. Arch Pathol Lab Med 1987;111:646–52.
2. Weinstein RS, Bhattacharyya AK, Graham AR, et al. Telepathology: A ten-year progress report. Hum Pathol 1997;28:1–7.
3. Eusebi V, Foschini L, Erde S, et al. Transcontinental consults in surgical pathology via the Internet. Hum Pathol 1997;28: 13–6.
4. Halliday BE, Bhattacharyya AK, Graham AR, et al. Diagnostic accuracy of an international static-imaging telepathology consultation service. Hum Pathol 1997;28:17–21.
5. Dunn BE, Almagro UA, Choi H, et al. Dynamic-robotic telepathology: Department of Veterans Affairs feasibility study. Hum Pathol 1997;28:8–12.
6. Eide TJ, Nordrum I. Current status of telepathology. APMIS 1994;102:881–91.
7. Nordrum I, Engum B, Rinde E, et al. Remote frozen section service: telepathology project in Northern Norway. Hum Pathol 1991;22:514–8.
8. Weinstein LJ, Epstein JI, Edlow D, *et al.* Static Image analysis of skin specimens: the application of telepathology to frozen section evaluation. Hum Pathol 1997;28:30–5.
9. Weinstein MH, Epstein JI. Telepathology diagnosis of prostate needle biopsies. Hum Pathol 1997;28:22–9.
10. Callas PW, Leslie KO, Mattia AR, et al. Diagnostic accuracy of a rural live video telepathology system. Am J Surg Pathol 1997;21:812–9.
11. Goldman H. Significance and detection of dysplasia in chronic colitis. Cancer 1996;78:261–3.
12. Riddell RH, Goldman H, Ransohoff DF, *et al.* Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. Hum Pathol 1983;14: 931–68.
13. Riddell RH. Implications of a diagnosis of dysplasia in ulcerative colitis. J Gastroenterol 1995;30(Suppl VIII):25–9.
14. Dixon MF, Brown LJR, Gilmour HM, *et al.* Observer variation in the assessment of dysplasia in ulcerative colitis. Histopathology 1988;13:385–97.
15. Fleiss J. Statistical methods for rates and proportions. 2nd ed. New York: Wiley; 1981. p. 191–230.
16. Singson RPC, Natarajan S, Greenson J, *et al.* Virtual microscopy and the Internet as telepathology consultation tools. A study of gastrointestinal biopsy specimen. Am J Clin Pathol 1999;111:792–5.
17. Riddell RH. Grading of dysplasia. Eur J Cancer 1995;31:1169–70.
18. Melville DM, Jass JR, Morson BC, *et al.* Observer study on the grading of dysplasia in ulcerative colitis; comparison with clinical outcome. Hum Pathol 1990;20:1008–14.
19. Dundas SAC, Kay R, Beck S, et al. Can histopathologists reliably assess dysplasia in chronic inflammatory bowel disease? J Clin Pathol 1987;40:1282–6.
20. Rosenstock E, Farmer RG, Petras R, et al. Surveillance for colonic carcinoma in ulcerative colitis. Gastroenterology 1985;89:1342–6.
21. Weinberg DS, Allaert FA, Dusserre P, et al. Telepathology diagnosis by means of digital still images. An international validation study. Hum Pathol 1996;27:111–8.

LIPPINCOTT WILLIAMS & WILKINS Unauthorized Use Prohibited