Comparison of a Medical-Grade Monitor vs Commercial Off-the-Shelf Display for Mitotic Figure Enumeration and Small Object (*Helicobacter pylori*) Detection

Andrew P. Norgan, MD, PhD,[1] *Vera J. Suman, PhD,*[2] *Charlene L. Brown,*[1] *Thomas J. Flotte, MD,*[1] *and Taofic Mounajjed, MD*[1]

From the Departments of [1]Laboratory Medicine and Pathology and [2]Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN.

Key Words: Informatics; Monitor; Calibration; Mitotic figure; *Helicobacter*; Dermatopathology; Gastrointestinal

Am J Clin Pathol February 2018;149:181-185

DOI: 10.1093/AJCP/AQX154

ABSTRACT

Objectives: *To examine the performance of a commercial off-the-shelf (COTS) monitor vs a medical-grade (MG) monitor for small object enumeration in standardized digital pathology images.*

Methods: *Pathologists reviewed 35 melanoma or 35 gastric biopsy images using the MG and COTS displays, with a 2-week washout period. Mitotic figure or* Helicobacter pylori *burden enumerations were compared with reference values reported by an expert subspecialist pathologist using a light microscope. Subjective evaluations of image color, brightness, and overall quality were also obtained.*

Results: *There was substantial agreement between the mitotic counts obtained by the evaluating pathologists between monitors and the reference mitotic figure or* H pylori *burden assessments. Six of the nine evaluating pathologists subjectively evaluated the monitors as substantially similar.*

Conclusions: *These findings are consistent with previous studies demonstrating that color calibration has limited impact on diagnostic accuracy and suggest that noncalibrated displays could be considered for fine assessment tasks.*

Image standardization efforts in digital pathology are challenging in that they must encompass software and hardware involved in image acquisition (scanning), manipulation (processing), and display, in addition to the significant variability in the source material that results from histology and staining processes. Recent reviews of end-to-end image standardization efforts[1] highlight significant progress in developing standards for digital slide scanning, but comparatively less work addresses display device requirements. As the pathology transitions to digital workflows, it will be vital to understand if (and how) display quality and calibration affect pathologic diagnosis. Diagnostic radiologists faced similar issues when transitioning from film to computer monitors. As studies revealed that display characteristics have a clear impact on radiologic diagnosis, digital radiology practice was standardized around the DICOM Grayscale Standard Display Function (GSDF)–calibrated medical-grade (MG) monochrome display.[2] The GSDF standard was developed to maximize the diagnostic information ("just noticeable differences") presented to radiologists and to ensure consistency of images viewed at different times, in different places, or with heterogeneous display hardware. The rapidly expanding use of color images throughout medical practice (eg, within radiology, ophthalmology, dermatology, and gastroenterology) has now driven similar initiatives to calibrate displays for interpretation of color data.[1,3-5]

Diagnostic and workflow benefits associated with the use of calibrated MG displays in radiology have been demonstrated for monochrome and MG color monitors.[6,7] However, the use of color in diagnostic radiology differs

from that in diagnostic pathology. Two studies examining color calibration of monitors for digital pathology demonstrated that calibration does not affect diagnostic accuracy of breast biopsy specimens.[8,9] Indeed, a third study, also on breast biopsy specimens, showed that a monochrome display was largely sufficient for an accurate diagnosis.[10] These studies do not conclude that color is an irrelevant image characteristic, but they do raise questions as to the strictness of color representation needed to enable accurate pathologic diagnoses. In contrast, color calibration and MG displays do appear to improve the speed of diagnosis.[9] One study has demonstrated a significant decrease in both time to identification of first diagnostic feature and total diagnostic time when using a composite display with twice the spatial resolution of a baseline display.[11]

The primary end point of interest in most digital pathology studies has been overall diagnostic accuracy. By contrast, this study was designed to assess whether the ability to identify and quantify small but clinically relevant diagnostic features differs with respect to whether a calibrated MG or commercial-grade monitor is used. Two features for which accurate quantitation or detection has a significant diagnostic and/or prognostic importance, mitotic figures in malignant melanoma and *Helicobacter pylori* detection in stomach biopsy specimens, were chosen as representative subjects for this study.

Materials and Methods

This study was reviewed and approved by the institutional review board. Skin biopsy specimens with a diagnosis of melanoma (received between January 2014 and July 2014) and random gastric biopsy specimens with and without *H pylori* bacteria (received between January 2013 and April 2014) were selected to ensure a range of mitotic figures (0-3) or a range of *H pylori* (none to many) by a board-certified dermatopathologist (T.F.) or board-certified gastrointestinal pathologist (T.M.), respectively **Table 1**. A representative slide with no folds in the tissues and high-quality H&E staining was chosen from each of the cases. A region of interest (ROI) (96,800 μm^2) with a desired mitotic or *H pylori* detection was then selected by the respective study pathologist by light microscopy and captured using an attached digital camera (Olympus DP71 [Olympus, Center Valley, PA]; TIFF format images; 24-bit color; approximately 4 MB in size each; 1,360 × 1,024 pixels). The study dermatopathologist enumerated the visible mitoses in the melanoma ROIs, and the study gastrointestinal pathologist determined the presence or absence of *H pylori* organisms and quantified the organisms when present (0 = none, 1 = few, 2 = moderate, and 3 = many) in the gastric ROIs.

Table 1
Description of the Study Case Features

Cohort/Feature Examined	No. (%) of Cases
Melanoma	
No. of mitoses (n = 35)	
0	10 (28.6)
1	14 (40.0)
2	7 (20.0)
3	4 (11.4)
Gastrointestinal	
Helicobacter pylori (n = 35)	
None	11 (31.4)
Few	11 (31.4)
Many	13 (37.1)

Nine reviewing pathologists (each with a minimum of 5 years of board-certified experience but no digital sign-out experience) participated in the study; seven pathologists (one dermatopathologist and six pathologists with subspecialty training in other areas) evaluated the melanoma cases, and two gastrointestinal pathologists evaluated the gastric cases. The reviewing pathologists were blinded to the evaluations of the study pathologists. ROIs were evaluated by the reviewing pathologists using a commercial off-the-shelf (COTS) Dell U3014 display (Dell, Round Rock, TX) with factory default calibration (gamma, 2.2; white point, 6,500K; screen size, 29.77 in.; resolution, 2,560 × 1,600 pixels; viewing angle [horizontal and vertical], 178 degrees; maximum luminance, 350 cd/m^2; contrast ratio, 1,000:1) and an MG BARCO Coronis Fusion 6MP LED display (Barco, Duluth, GA) with monitored calibration (gamma DICOM GSDF [grayscale display function]; white point, 8,000K; screen size, 30.4 in.; resolution, 3,280 × 2,048 pixels; viewing angle [horizontal and vertical], 178 degrees; luminance, 500 cd/m^2; contrast ratio, 1,000:1). The pathologists were randomized to a monitor for initial review, and there was a minimum 2-week washout period between reviews. In addition to evaluating the specific features of the ROIs, the pathologists provided a subjective evaluation of color, brightness, and overall quality for each monitor per case using a 5-point Likert-type scale (1 = very poor, 2 = poor, 3 = average, 4 = good, and 5 = very good).

The findings of the study pathologist and each of the reviewing pathologists and agreement between reviewing pathologists' findings using the COTS monitor and MG monitors were evaluated statistically using weighted κ.

Results

For both tasks of mitotic figure enumeration and *H pylori* burden classification, the level of agreement

DOI: 10.1093/ajcp/aqx154

Table 2
Weighted κ Coefficients and Cases in Agreement Between Each Evaluating Pathologist and the Expert Light Microscopy Assessment by Monitor Type for Both the Mitotic Count (Melanoma) Enumeration and *Helicobacter pylori* Burden Assessments

Evaluating Pathologist	No. of COTS Cases in Agreement	COTS (n = 35), κ (95% CI)	No. of MG Cases in Agreement	MG (n = 35), κ (95% CI)
Mitotic figure enumeration				
1	31	0.87 (0.74-0.99)	30	0.85 (0.72-0.98)
2	31	0.87 (0.74-0.99)	27	0.76 (0.61-0.92)
3	30	0.84 (0.71-0.98)	31	0.87 (0.75-0.99)
4	33	0.95 (0.87-0.99)	30	0.85 (0.72-0.98)
5	26	0.70 (0.51-0.88)	25	0.71 (0.56-0.86)
6	32	0.89 (0.77-0.99)	30	0.85 (0.71-0.98)
7	29	0.83 (0.69-0.96)	27	0.77 (0.62-0.91)
Helicobacter pylori assessment				
8	30	0.84 (0.70-0.97)	32	0.90 (0.80-0.99)
9	30	0.84 (0.70-0.97)	30	0.84 (0.70-0.97)

CI, confidence interval; COTS, commercial off-the-shelf; MG, medical grade.

between the study pathologist's findings with light microscopy and the reviewing pathologists' findings using either the COTS monitor or MG monitor were substantially similar, with κ values of 0.75 or more (indicating substantial agreement) for all but one pathologist Table 2. Agreement between the pathologists' findings with the COTS and MG displays ranged from 0.70 to 0.95 (six of seven pathologists ranged from 0.8-0.95) for mitotic figure enumeration and was 0.80 for both the pathologists classifying *H pylori* burden Table 3. From a clinical perspective, detection (rather than enumeration) of *H pylori* is essential. No substantial difference in the ability of pathologists to identify *H pylori* using either display was observed Table 4.

Ratings for color, brightness, and overall quality were similar between COTS and MG displays for six of the nine pathologists, with 88% or more of the cases reviewed rated as having good or very good color, brightness, and overall quality. Of the three remaining pathologists, the overall quality ratings of the COTS display were higher than the MG display in most cases for two of the pathologists and lower for the third pathologist Table 5.

Table 3
Weighted κ Coefficients and Cases in Agreement Between the COTS and MG Monitors for Each Testing Pathologist for Melanoma Mitotic Figure Enumeration and *Helicobacter pylori* Burden Assessments

Testing Pathologist	No. of Cases in Agreement	κ (95% CI) (n = 35)
Mitotic figure enumeration		
1	34	0.87 (0.74-0.99)
2	31	0.87 (0.74-0.99)
3	32	0.84 (0.71-0.98)
4	31	0.95 (0.87-0.99)
5	29	0.70 (0.51-0.88)
6	31	0.89 (0.77-0.99)
7	31	0.83 (0.69-0.96)
Helicobacter pylori assessment		
8	29	0.80 (0.65-0.95)
9	29	0.79 (0.64-0.95)

CI, confidence interval; COTS, commercial off-the-shelf; MG, medical grade.

Discussion

For both the COTS and MG color-calibrated display, a high level of agreement between the study pathologist and each evaluating pathologist was observed for both mitotic count and *H pylori* burden assessments. These data suggest that there is no significant negative impact in using a noncalibrated display for either of these detailed assessment tasks. In addition to improved color accuracy, the MG display used in this study also had superior absolute and spatial resolution and luminance; however, these factors do not appear to have substantially improved pathologist performance for the two assessed tasks. These findings are consistent with those of previous reports demonstrating that color calibration has a limited impact on diagnostic pathology accuracy. A potential limitation of this study is that time to identification of first diagnostic feature and overall interpretation speeds were not assessed.

Mitotic figure counting is considered a difficult task, and there is significant variability in reported mitotic counts between observers.[12,13] The substantial interobserver and intraobserver agreement observed in this study is likely a result of using confined digital ROIs rather than whole slides or whole-slide images. Overall, the use of ROIs (vs whole slide images) in this study was justified in that ROIs enabled the analysis to be focused on differences inherent to the individuals' monitors by reducing the risk that confounding elements such as search strategy (which studies suggest is influenced to a greater degree by display resolution differences) would influence the results. Nevertheless,

DOI: 10.1093/ajcp/aqx154

Table 4
COTS and MG Monitor Performance in Identification of Presence or Absence of *Helicobacter pylori* in 35 Cases

Testing Pathologist	Monitor	Sensitivity, % (No./Total No.)	Specificity, % (No./Total No.)
8	COTS	96 (23/24)	100 (11/11)
	MG	100 (24/24)	100 (11/11)
9	COTS	96 (23/24)	91 (10/11)
	MG	96 (23/24)	91 (10/11)

COTS, commercial off-the-shelf; MG, medical grade.

use of ROIs is a clear limitation of this work's generalizability to whole-slide evaluation. In addition, it should be noted that no additional calibration of the COTS was performed (ie, the COTS monitor was used with factory settings "out of the box"). Color, contrast, white point, and luminance calibration of a consumer-grade monitor are possible and could potentially have influenced the subjective preference judgments. Finally, this study was focused on the tasks of mitotic counting and *H pylori* identification, tasks in which color is expected to play an important but limited role. Other display factors (in particular resolution) may be of greater significance in the performance of those tasks using digital slides.

Six of the nine evaluating pathologists found that the overall quality of the monitors was good to very good in a similar proportion of cases.[14,15] Both displays were rated highly in this study, and the differences between them appear to be of a lesser magnitude than in some previous studies, although direct comparison is difficult due to the use of different evaluation metrics and scales. Pathologists have been long accustomed to variability in stain quality and microscope performance and therefore differences in color, contrast, luminance, and resolution. Adaptability borne out of conventional pathology practice may explain why studies have generally shown that variation in color or resolution has only negligible impacts on diagnostic accuracy. If the performance of consumer- and prosumer-grade "off-the-shelf" color displays is sufficient for all or some diagnostic pathology, the lower costs of consumer displays ($500-$2,000 for COTS displays vs $10,000-$17,000 typical of MG monitors used for diagnostic tasks) are likely to generate significant interest in their use, as it has in other areas of medicine.[16-18] However, it will be important to balance a lower initial purchase cost against diagnostic performance over time. MG displays typically have ancillary monitoring and calibration services available that can maintain and verify performance over time, albeit with additional expense; such services are not usually available for consumer-level displays.

Table 5
Pathologists' Assessment of Display Quality, Brightness, and Color (n = 35)[a]

Testing Pathologist	COTS, No. (%)		MG, No. (%)	
	2-3	4-5	2-3	4-5
1				
Quality	4 (11.4)	31 (88.6)	2 (5.7)	33 (94.3)
Brightness	2 (5.7)	33 (94.3)	1 (2.9)	34 (97.1)
Color	2 (5.7)	33 (94.3)	2 (5.7)	33 (94.3)
2				
Quality	17 (48.6)	18 (52.4)	12 (34.3)	23 (65.7)
Brightness	11 (31.4)	24 (68.6)	6 (17.1)	29 (82.9)
Color	12 (34.3)	13 (65.7)	9 (25.7)	26 (74.3)
3				
Quality	0	35 (100)	0	35 (100)
Brightness	0	35 (100)	0	35 (100)
Color	0	35 (100)	0	35 (100)
4				
Quality	2 (5.7)	33 (94.3)	4 (11.4)	31 (88.6)
Brightness	0	35 (100)	0	35 (100)
Color	0	35 (100)	0	35 (100)
5				
Quality	4 (11.4)	31 (88.6)	3 (8.6)	32 (91.4)
Brightness	1 (2.9)	34 (97.1)	0	35 (100)
Color	0	35 (100)	0	35 (100)
6				
Quality	1 (2.9)	34 (97.1)	1 (2.9)	34 (97.1)
Brightness	0	35 (100)	0	35 (100)
Color	0	35 (100)	0	35 (100)
7				
Quality	1 (2.9)	34 (97.1)	6 (17.1)	29 (82.9)
Brightness	1 (2.9)	34 (97.1)	4 (11.4)	31 (88.6)
Color	1 (2.9)	34 (97.1)	6 (17.1)	29 (82.9)
8				
Quality	1 (2.9)	34 (97.1)	1 (2.9)	34 (97.1)
Brightness	0	35 (100)	0	35 (100)
Color	0	35 (100)	0	35 (100)
9				
Quality	11 (31.4)	24 (68.6)	2 (5.7)	33 (94.3)
Brightness	5 (14.3)	30 (85.7)	0	35 (100)
Color	5 (14.3)	30 (85.7)	0	35 (100)

COTS, commercial off-the-shelf; MG, medical grade.
[a]1 = poor to 5 = excellent.

At present, the question of which display to use for diagnostic pathology is open only for secondary diagnosis. In the recent approval of a digital system for primary diagnosis in pathology, the monitor was included within the US Food and Drug Administration definition of the digital pathology "device," and therefore a vendor-specified display is required for primary diagnostic use. To our knowledge, no scientific evidence supports the necessity or superiority of a vendor-supplied display over another of comparable quality. This precedent is concerning, as it may limit competition within the digital pathology space to those vendors able to offer a complete end-to-end digital pathology solution, potentially reducing the rate of innovation and increasing costs. As pathology moves toward a digital practice, it will

DOI: 10.1093/ajcp/aqx154

be to the benefit of the field for pathology systems to more closely resemble those used in radiology—a series of interchangeable components from different vendors that have standards-based interoperability.

Corresponding author: Andrew P. Norgan, Mayo Clinic, Laboratory Medicine and Pathology, 200 First St SW, Rochester, MN 55905; norgan.andrew@mayo.edu.

References

1. Badano A, Revie C, Casertano A, et al; Summit on Color in Medical Imaging. Consistency and standardization of color in medical imaging: a consensus report. *J Digit Imaging*. 2015;28:41-52.
2. Fetterly KA, Blume HR, Flynn MJ, et al. Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays. *J Digit Imaging*. 2008;21:193-207.
3. Bautista PA, Hashimoto N, Yagi Y. Color standardization in whole slide imaging using a color calibration slide. *J Pathol Inform*. 2014;5:4.
4. Revie WC, Shires M, Jackson P, et al. Color management in digital pathology. *Anal Cell Pathol* 2014;2014:1-2.
5. Kimpe T, Rostang J, Van Hoey G, et al. Color standard display function: a proposed extension of DICOM GSDF. *Med Phys*. 2016;43:5009.
6. Krupinski EA. Medical grade vs off-the-shelf color displays: influence on observer performance and visual search. *J Digit Imaging*. 2009;22:363-368.
7. Salazar AJ, Aguirre DA, Ocampo J, et al. DICOM gray-scale standard display function: clinical diagnostic accuracy of chest radiography in medical-grade gray-scale and consumer-grade color displays. *AJR Am J Roentgenol*. 2014;202:1272-1280.
8. Campbell WS, Hinrichs SH, Lele SM, et al. Whole slide imaging diagnostic concordance with light microscopy for breast needle biopsies. *Hum Pathol*. 2014;45:1713-1721.
9. Krupinski EA, Silverstein LD, Hashmi SF, et al. Observer performance using virtual pathology slides: impact of LCD color reproduction accuracy. *J Digit Imaging*. 2012;25:738-743.
10. Campbell WS, Talmon GA, Foster KW, et al. Sixty-five thousand shades of gray: importance of color in surgical pathology diagnoses. *Hum Pathol*. 2015;46:1945-1950.
11. Randell R, Ambepitiya T, Mello-Thoms C, et al. Effect of display resolution on time to diagnosis with virtual pathology slides in a systematic search task. *J Digit Imaging*. 2015;28:68-76.
12. Malon C, Brachtel E, Cosatto E, et al. Mitotic figure recognition: agreement among pathologists and computerized detector. *Anal Cell Pathol (Amst)*. 2012;35:97-100.
13. Yadav KS, Gonuguntla S, Ealla KK, et al. Assessment of interobserver variability in mitotic figure counting in different histological grades of oral squamous cell carcinoma. *J Contemp Dent Pract*. 2012;13:339-344.
14. Sharma G, Sharma G, Shah A, et al. Evaluation of different display modalities for whole slide images in pathology. *J Pathol Inf*. 2011;2:S44-S45.
15. D'Haene N, Maris C, Rorive S, et al. Comparison study of five different display modalities for whole slide images in surgical pathology and cytopathology in Europe. *Prog Biomed Opt Imaging Proc SPIE*. 2013;8676:1-10.
16. Kallio-Pulkkinen S, Haapea M, Liukkonen E, et al. Comparison between DICOM-calibrated and uncalibrated consumer grade and 6-MP displays under different lighting conditions in panoramic radiography. *Dentomaxillofac Radiol*. 2015;44:20140365.
17. Kallio-Pulkkinen S, Haapea M, Liukkonen E, et al. Comparison of consumer grade, tablet and 6MP-displays: observer performance in detection of anatomical and pathological structures in panoramic radiographs. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2014;118:135-141.
18. Farahani N, Post R, Duboy J, et al. Exploring virtual reality technology and the oculus rift for the examination of digital pathology slides. *J Pathol Inform*. 2016;7:22.

DOI: 10.1093/ajcp/aqx154