Ki-67 Proliferation Index in Neuroendocrine Tumors: Can Augmented Reality Microscopy With Image Analysis Improve Scoring?

Swati P. Satturwar, MD [ID] [1]; Joshua L. Pantanowitz, BS[2]; Christopher D. Manko, BS[2]; Lindsey Seigh, BS[1]; Sara E. Monaco, MD [ID] [1]; and Liron X. Pantanowitz, MD [ID] [1]

BACKGROUND: The Ki-67 index is important for grading neuroendocrine tumors (NETs) in cytology. However, different counting methods exist. Recently, augmented reality microscopy (ARM) has enabled real-time image analysis using glass slides. The objective of the current study was to compare different traditional Ki-67 scoring methods in cell block material with newer methods such as ARM. **METHODS:** Ki-67 immunostained slides from 50 NETs of varying grades were retrieved (39 from the pancreas and 11 metastases). Methods with which to quantify the Ki-67 index in up to 3 hot spots included: 1) "eyeball" estimation (EE); 2) printed image manual counting (PIMC); 3) ARM with live image analysis; and 4) image analysis using whole-slide images (WSI) (field of view [FOV] and the entire slide). **RESULTS:** The Ki-67 index obtained using the different methods varied. The pairwise kappa results varied from no agreement for image analysis using digital image analysis WSI (FOV) and histology to near-perfect agreement for ARM and PIMC. Using surgical pathology as the gold standard, the EE method was found to have the highest concordance rate (84.2%), followed by WSI analysis of the entire slide (73.7%) and then both the ARM and PIMC methods (63.2% for both). The PIMC method was the most time-consuming whereas image analysis using WSI (FOV) was the fastest method followed by ARM. **CONCLUSIONS:** The Ki-67 index for NETs in cell block material varied by the method used for scoring, which may affect grade. PIMC was the most time-consuming method, and EE had the highest concordance rate. Although real-time automated counting using image analysis demonstrated inaccuracies, ARM streamlined and hastened the task of Ki-67 quantification in NETs. *Cancer Cytopathol* **2020;0:1-10**. © *2020 American Cancer Society*.

KEY WORDS: augmented reality microscope; cell block; digital image analysis; Ki-67 quantification; manual count; neuroendocrine tumors.

INTRODUCTION

Grading of gastrointestinal neuroendocrine tumors (NETs) is performed by quantifying their mitotic activity and/or Ki-67 index.[1] The grade is critical in NET prognostication and treatment decisions, along with other parameters such as tumor size, lymphovascular invasion, and stage of disease. The World Health Organization (WHO) classifies neuroendocrine neoplasms as well-differentiated NETs, poorly differentiated neuroendocrine carcinoma (NEC), and mixed neuroendocrine-nonneuroendocrine neoplasms. WHO categorizes NETs as grade 1 (those with a Ki-67 index <3% [mitotic count <2 per 10 high-power fields (HPF)]), grade 2 (Ki-67 index of 3%-20% [mitotic count of 2-20 per 10 HPF]), and grade 3 (Ki-67 index >20% [mitotic count >20 per 10 HPF]); NECs are grade 3 by definition.[1-3] The Ki-67 index is determined by counting the percentage of

positively stained tumor cells in an area of the tumor with the highest nuclear labeling (ie, a hot spot) based on 500 to 2000 tumor cells.

NETs involving the pancreas and their metastases to visceral sites (eg, the liver) frequently are diagnosed using fine-needle aspiration (FNA).[4-7] Often, these FNA specimens represent the primary diagnostic material for patients. Hence, the preoperative grading of NETs on such cytology material typically is required to guide treatment decisions.[8-11] Currently, to our knowledge, it remains questionable whether a Ki-67 index obtained on cytology material, such as a cell block, is truly representative of the entire tumor compared with the surgical resection specimen.[12] Several studies also have attempted to establish the best counting method for determining the Ki-67 index.[13-18] However, to the best of our knowledge, there is no accepted standard as yet and the best scoring method still is debatable. Different methods for Ki-67 index quantification include "eyeball" estimation (EE) performed manually by a pathologist counting in real time through their microscope eyepiece, manual counting using a printed image that was captured using a camera mounted to a microscope at ×20 magnification,[19] and computer-assisted quantification using digital image analysis (DIA).[15,20,21] Manual methods suffer from low reproducibility and high interreader variability among pathologists, especially for low-grade NETs. Moreover, the manual methods, especially if they involve the printing of images, are impractical in certain clinical practice settings and are labor-intensive. To overcome the interobserver variabilities inherent in human counting, researchers accordingly have explored automated counting methods using digital images and a variety of image analysis algorithms. Although faster and more precise, DIA systems can be costly because they may require an expensive whole-slide scanner to digitize slides and information technology support.

Recently, augmented reality microscopy (ARM) has become available that enables real-time image analysis to be conducted using a traditional light microscope and glass slides,[22] which avoids the need to first photograph or digitize slides. An augmented reality microscope is a modified ("smart") microscope that includes a small computer unit. This computer unit can be attached to the side of any microscope or it can be inserted between the microscope's objective lenses and eyepiece unit. It incorporates a built-in camera to capture high-quality images. The images acquired by this unit occur in real time and can be displayed on an attached computer monitor. In addition, if the end user looks through the eyepieces of the microscope, they can see computer-generated output, such as annotations, being projected as an overlay directly on the glass slide. ARM thus permits real-time image analysis to be performed on glass slides with the output of the algorithm superimposed on the slide, thereby generating a composite field of view (FOV) that can be used to supervise data collection. This can be used to perform simple measurements (eg, depth of tumor invasion) and, when coupled with image analysis software, to quantify immunohistochemical stains (eg, Ki-67 proliferation index) or it can be combined with more sophisticated deep learning algorithms (eg, to rapidly detect metastases in lymph nodes).[23] As opposed to DIA applied to whole-slide imaging (WSI), ARM is quicker to use, provides for the real-time and seamless integration of algorithms without the need to first acquire digital images, may be cheaper than buying a whole-slide scanner, and does not require special technical skills to operate. Based on our calculation of direct equipment costs, the ARM device is approximately 20 times less expensive than a high-capacity whole-slide scanner.

To the best of our knowledge, no studies to date have applied ARM to cytology. We anticipate that ARM may overcome many of the aforementioned issues related to Ki-67 quantification. Therefore, the objective of the current study was to compare ARM with different (manual and DIA with WSI) scoring methods used for determining the Ki-67 proliferation index of NETs in cell block material.

MATERIALS AND METHODS

Case Collection

After institutional review board approval, a total of 50 cytology NETs from 45 patients were retrieved (39 pancreatic tumors and 11 metastases to the liver) from January 2014 to July 2019. Data including patient age, sex, subsequent surgical resection, and American Joint Committee on Cancer clinical stage of disease were recorded. A total of 19 of the 39 pancreatic tumors (48.71%) underwent a subsequent surgical resection. The inclusion criteria for cases to be enrolled in this study were specimens for which a satisfactory cell block was available (ie, tumor cells were present), a NET proven by immunohistochemistry, the Ki-67 index was

available in the cytology report, and the original Ki-67 immunostained slide was available for further analysis. Exclusion criteria included cases with a cell block for which the tumor cellularity was <100 tumor cells. The cellularity in the cell block material varied (<500 tumor cells in 7 cases, 501-1000 tumor cells in 11 cases, 1001-2000 tumor cells in 10 cases, and >2000 tumor cells in 22 cases). The tumors enrolled in the current study were comprised of all grades of NET. Cytology cases were comprised of grade 1 NET in 26 cases, grade 2 NET in 13 cases, grade 3 NET in 3 cases, and grade 3 NEC in 8 cases. Surgical resection cases were comprised of grade 1 NET in 9 cases, grade 2 NET in 9 cases, grade 3 NET in no cases, and grade 3 NEC in 1 case.

Immunohistochemistry

Formalin-fixed, paraffin-embedded cell blocks were used for immunohistochemical staining. The Ki-67 (clone 30-9; Ventana Medical Systems, Oro Valley, Arizona) immunohistochemical analysis was performed on the Ventana Medical System, with appropriate positive controls run for all cases. Any staining intensity was considered to be positive.

Quantification Methods for the Ki-67 Proliferation Index

The Ki-67 proliferation index was recalculated using the different methods listed below in a blinded manner. Up to 3 hot-spot FOVs were evaluated based on the tumor cellularity of the cell block (3 FOVs in 30 cases, 2 FOVs in 11 cases, and 1 FOV in 9 cases). These hot spots were designated using a black permanent marker on the slides and the same hot spots were analyzed with each method. These black dots did not interfere with image acquisition or image analysis. The average of these hot spots was used to determine the overall tumor grade. The time spent determining the calculations for each method was recorded.

Eyeball estimation

EE data were obtained from the Ki-67 score provided in the original cytology report. Typically, cytopathologists on clinical service at the study institution assess the entire cell block glass slide at a low magnification (using Olympus microscopes [Olympus Corporation, Center Valley, Pennsylvania]) for the detection of hot-spot areas and then at an intermediate magnification to estimate the
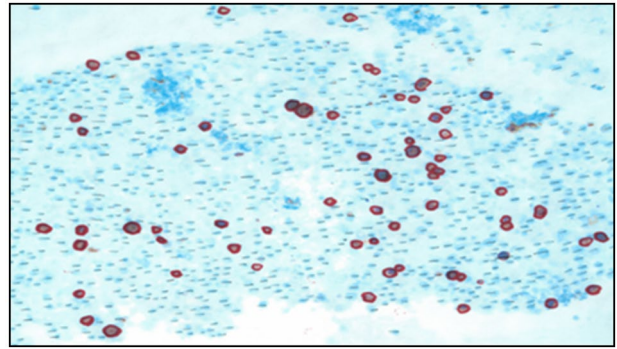

FIGURE 1. Example of a section of a neuroendocrine tumor cell block in which the Ki-67 index was scored using the printed image manual counting method. Positive tumor nuclei are circled in red and all negative tumor nuclei are crossed out with a gray pencil.

Ki-67 index without performing actual counts of individual cells. The cases enrolled in the current study were signed out by 3 board-certified cytopathologists. The time they spent determining these scores was not recorded because these estimations were performed prior to the study during routine clinical care.

Printed image with manual counting

The printed image with manual counting (PIMC) method was performed using a digital camera (Insight camera; SPOT Imaging Solutions, Sterling Heights, Michigan) attached to an Olympus light microscope (Olympus Corporation) to capture and then print out (on letter-size printer paper) static color images of up to 3 hot spot FOVs of each cell block at ×20 magnification. All Ki-67–positive tumor cell nuclei were circled in red and all negative nuclei were crossed out using a gray pencil (Fig. 1). The Ki-67 index was expressed as the positive tumor nuclei over the total tumor nuclei. PIMC was performed by only 1 of the authors (S.S.).

Augmented Reality Microscopy

To perform ARM, an Augmentiqs device (Augmentiqs, Haifa, Israel) was attached to an Olympus light microscope between the microscope's objective lenses and eyepiece unit (Fig. 2 Top). The Augmentiqs device also was attached to an adjacent computer workstation with a monitor. The embedded camera on the device maintained normal viewing of the optical plane of the microscope while simultaneously allowing the inbuilt projection system to generate live augmented reality
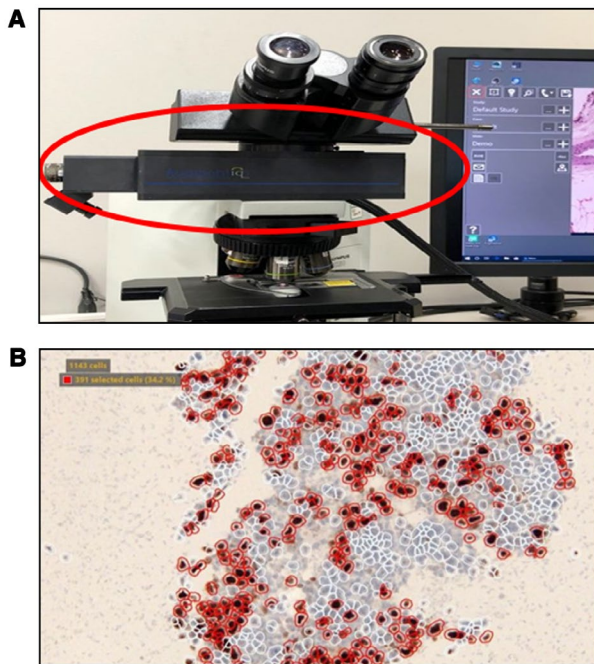
FIGURE 2. (*Top*) Light microscope with the Augmentiqs device (red circle) fitted between the objective lenses and the eyepiece. (*Bottom*) Screenshot showing a field of view using the augmented reality microscopy counting method in which the image analysis segments individual nuclei (white circles) and the end user manually selects positive nuclei (red circles).

composite images within the microscope eyepiece, and simultaneously displayed this composite image on the attached computer monitor. For quantitative analytics, this ARM setup was coupled with image analysis software modified from QuPath software (Queen's University Belfast, Belfast, Northern Ireland; https://qupath.github.io) for the detection, segmentation, and scoring of nuclear stains (Fig. 2 Bottom). For all NET cases, the same hot spots at ×20 magnification were analyzed. After image analysis results were superimposed on glass slides, 2 of the authors (J.P. and C.M.) manually selected all stained nuclei. These selected stained cells then were divided by the total number of cells that were automatically detected by the software to calculate the Ki-67 index for each FOV. Hence, the actual quantitative assessment component still required human involvement and was not automated.

Digital image analysis using whole-slide images

Ki-67–stained slides were scanned using an Aperio AT2 scanner (Leica Biosystems, Wetzlar, Germany) using ×40

magnification and 1 Z-plane. DIA then was performed using the Aperio immunohistochemistry color convoluted, nuclearV9 quantitative image analysis algorithm (Leica Biosystems). This allowed for the percentage of positive over total tumor nuclei (Ki-67 index) to be calculated. DIA-WSI was performed for each hot-spot FOV (FOV analysis) and also for the entire cell block (slide-level analysis).

Histology analysis

These data were obtained from the 19 surgical resection pathology reports. The Ki-67 index had been calculated by 7 different surgical pathologists with expertise in gastrointestinal pathology, using manual counting of printed images of at least 500 cells.

Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics 22 (IBM Corporation, Armonk, New York), and statistical significance was assumed at $P \leq .05$. The Cohen kappa was used to determine interrater reliability for tumor grade assessment (grades 1-3) for pairwise comparison of Ki-67 quantification methods. The normality of the distribution of continuous variables (number of minutes and percentage of Ki-67) was examined using the Shapiro-Wilk normality test. Because the data were not normally distributed, nonparametric statistical tests were used. The Kendall tau-b correlation coefficient was used to determine the relationship between the amount of time spent in assessing the presence of Ki-67 and the amount of Ki-67 identified. The interclass correlation coefficient (ICC) was calculated to determine interrater reliability among all cytological methods, and when available with and without subsequent surgical resection procedures.

RESULTS

The clinicopathologic characteristics of the NETs included in the current study are summarized in Table 1. The 50 cases were obtained from 45 patients with an average age of 62 years and demonstrated a male-to-female ratio of 1:1. The most common location for pancreatic tumors was the body, followed by the tail. Pancreatic NETs measured on average 3.12 cm in size. Of the cases that had subsequent surgical resection specimens, >50% of these patients had lymph node metastases and 10%

TABLE 1. Clinicopathologic Characteristics of NETs Studied (50 by Cytology and 19 by Histology

Characteristic	Value
Average patient age (range), y	62 (35-79)
Sex	
Male	22 (49%)
Female	23 (51%)
Average tumor size (range), cm	3.12 (1.5-8.5)
Location of tumor, no. (%)	
Pancreatic head	9 (23%)
Pancreatic body	14 (36%)
Pancreatic tail	13 (33.33%)
Pancreatic body-tail	3 (7.69%)
Liver (metastasis)	11 (22%)
Histologic grade (n = 19)	
1	9 (47%)
2	9 (47%)
3	1 (5%)
T classification, no. (%)	
T1	3 (15.7%)
T2	10 (52.63%)
T3	6 (31.57%)
T4	0
N classification, no. (%)	
N0	7 (36.84%)
N1	10 (52.63%)
NX	2 (10.52%)
M classification, no. (%)	
M0	17 (89.47%)
M1	2 (10.52%)
AJCC stage, no. (%)	
I	4 (21%)
II	6 (31.57%)
III	7 (36.84%)
IV	2 (10.52%)

Abbreviation: AJCC, American Joint Committee on Cancer eighth edition; NETs, neuroendocrine tumors.

had other distant metastases. Of the 11 cases of liver metastases, 6 (54.55%) originated from the pancreas and 5 (45.45%) were metastases from the lung or intestinal NETs.

The Ki-67 proliferation index based off the original cytopathology reports indicated that the average Ki-67 score for the EE method of grading cell blocks was 18.17% (range, 0.5%-95.0%) and that for the surgical resection specimens (19 cases) was 8.91% (range, 0.7%-95.0%). When averaging the 3 hot spots, the Ki-67 index calculated using the different methods for the current study all differed. For the PIMC method, the average Ki-67 score was 16.89% (range, 0.1%-93.43%); it was 12.35% (range, 0.1%-85.83%) for the ARM method and was 19.01% (range, 0.57%-94.0%) when using DIA-WSI (FOVs) and 11.48% (range, 0.31%-87.98%) when using DIA-WSI (entire slide). When using the ARM method, some inaccuracies in cell segmentation were noted with the image analysis algorithm, resulting in certain cells (eg, overlapping cells present in clusters,

very faint staining nuclei) not being counted individually. For the subset of cases with subsequent pancreatic resections (19 cases), a Wilcoxon signed rank test indicated that the Ki-67 index of these NETs in the surgical specimen procedure (median, 3.5%) was significantly higher than the index obtained using the traditional EE method (median, 1.5%) when applied to corresponding preoperative cytology samples (Z [Z-statistic] = −2.136; r [effect size] = .35 [P = .033]). This equates to a difference of tumor grade 1 as determined by the traditional EE manual method in cell blocks versus grade 2 in the follow-up surgical specimens. For NETs with surgical follow-up in which the Ki-67 index in cell blocks was assessed using PIMC, ARM, and DIA-WSI (entire slide), these scores similarly resulted in a median equivalent of grade 1, but the difference in the median Ki-67 index for these computer-assisted methods was not statistically significant.

Tumor grade assessment using WHO guidelines and shown by different methods is shown in Table 2. The percentage of cases scored using DIA-WSI (FOV) with grade 2 was higher at 50% compared with approximately 30% for the other methods being used to score cell blocks. The main reason for the difference in the number of grade 1 NETs versus grade 2 NETs using the different methods was the narrow cutoff value of the Ki-67 index. The DIA-WSI (FOV) method had inaccuracies (typically causing upgrading of the tumor grade) due to miscounting of nontumoral material (debris in cell blocks, tissue folds, and proliferating inflammatory cells). Of the grade 3 tumors, 8 of 11 tumors were grade 3 NECs and all were scored as grade 3 using different cytology methods. The interrater reliability for tumor grade by paired Ki-67 assessment method is shown in Table 3. The pairwise kappa results varied from no agreement (0.04 for DIA-WSI [FOV] and histology comparisons) to near-perfect agreement (0.81 for the ARM and PIMC methods) based on a scale of <0.2 indicating no agreement, 0.21 to 0.4 indicating fair agreement, 0.41 to 0.6 indicating moderate agreement, 0.61 to 0.8 indicating substantial agreement, and >0.81 indicating near-perfect agreement. Using the surgical resection specimens (19 cases) as the gold standard, Table 4 compares the concordance rate between cytology and histology cases. Compared with this histopathology gold standard, the concordance rate was highest for the EE method at 84.2% (kappa, 0.71) followed by the DIA-WSI (entire slide) method at 73.7% (kappa, 0.52). ARM and PIMC

TABLE 2. Tumor Grade Assessment by Different Ki-67 Quantification Methods

Tumor Grade	Ki-67 Index	EE No.	EE %	PIMC No.	PIMC %	ARM No.	ARM %	DIA-WSI (FOV) No.	DIA-WSI (FOV) %	DIA-WSI (Entire Slide) No.	DIA-WSI (Entire Slide) %	Histology No.	Histology %
1	<3%	26	52%	24	48%	25	50%	10	20%	24	48%	9	47%
2	3%-20%	13	26%	15	30%	15	30%	28	56%	18	36%	9	47%
3	>20%	11	22%	11	22%	10	20%	12	24%	8	16%	1	5%
Total		50	100%	50	100%	50	100%	50	100%	50	100%	19	100%

Abbreviations: ARM, augmented reality microscopy; DIA, digital image analysis; EE, eyeball estimation; FOV, field of view; PIMC, printed image manual counting; WSI, whole-slide imaging.

TABLE 3. Pairwise Kappa Calculations to Determine Interrater Reliability for Tumor Grade Assessment Among Ki-67 Calculation Methods

Scoring Method	EE	PIMC	ARM	DIA-WSI (FOV)	DIA-WSI (Entire Slide)	Histology
EE	—	0.74	0.68	0.46	0.61	0.71
PIMC	0.74	—	0.81	0.50	0.43	0.33
ARM	0.68	0.81	—	0.44	0.45	0.33
DIA-WSI (FOV)	0.46	0.50	0.44	—	0.37	0.04
DIA-WSI (entire slide)	0.61	0.43	0.45	0.37	—	0.52
Histology	0.71	0.33	0.33	0.04	0.52	—

Abbreviations: ARM, augmented reality microscopy; DIA, digital image analysis; EE, eyeball estimation; FOV, field of view; PIMC, printed image manual counting; WSI, whole-slide imaging.
Based on a scale of <0.2 indicating no agreement, 0.21 to 0.4 indicating fair agreement, 0.41 to 0.6 indicating moderate agreement, 0.61 to 0.8 indicating substantial agreement, and >0.81 indicating near-perfect agreement.

TABLE 4. Tumor WHO Grading Based on Ki-67 Quantification of Cell Block Material by Different Methods and Concurrent Surgical Resection Specimens

Scoring Method	Cytology Grades	Histology (Surgical Resection) Grade 1	Histology (Surgical Resection) Grade 2	Histology (Surgical Resection) Grade 3	Total All Grades	Percent Agreement[a]	Kappa
EE	1	9	3	0	12	84.2%	0.71
	2	0	6	0	6		
	3	0	0	1	1		
PIMC	1	6	4	0	10	63.2%	0.33
	2	3	5	0	8		
	3	0	0	1	1		
ARM	1	7	5	0	12	63.2%	0.33
	2	2	4	0	6		
	3	0	0	1	1		
DIA-WSI (FOV)	1	2	3	0	5	47.4%	0.04
	2	7	6	0	13		
	3	0	0	1	1		
DIA-WSI (entire slide)	1	8	4	0	12	73.7%	0.52
	2	1	5	0	6		
	3	0	0	1	1		

Abbreviations: ARM, augmented reality microscopy; DIA, digital image analysis; EE, eyeball estimation; FOV, field of view; PIMC, printed image manual counting; WHO, World Health Organization; WSI, whole-slide imaging.
[a] The percent agreement shows the correlation of the Ki-67 index between cytology cell block and subsequent surgical resection specimen assessments.

each had a concordance rate of 63.2% (kappa, 0.33). There was excellent reliability among all scoring methods when 3 hot spots (FOV) were evaluated (Table 5) compared with just counting 1 or 2 hot spots (ICC of 0.912 [$P < .001$] for 3 hot spots vs ICC of 0.706 [$P = .13$] for 1 hot spot and ICC of 0.606 [$P = .015$] for 2 hot spots).

The time spent, including the time required for the annotation of slides, as measured in minutes for the different methods for Ki-67 quantification varied markedly (Table 6). The traditional EE and manual histology methods were excluded for statistical analysis because the time spent by the pathologists was not recorded in the original reports. Generally, cases with higher percentages of Ki-67 immunostained cells were associated with taking longer times to score (Fig. 3). DIA-WSI scoring of only the hot spots was the fastest

TABLE 5. Interclass Correlation Coefficient Results for Tumor Grade Assessment Based on the Number of Hot Spot FOVs

Different methods (PIMC, ARM, DIA-WSI [FOV])		Interclass Correlation Coefficient	95% CI
1 FOV	Average measures	0.706	0.144-0.927
2 FOV		0.606	0.095-0.876
3 FOV		0.912	0.833-0.956

Abbreviations: ARM, augmented reality microscopy; DIA, digital image analysis; FOV, field of view; PIMC, printed image manual counting; WSI, whole-slide imaging.

method of calculating the Ki-67 proliferation index (average, 2.6 minutes per case), whereas the PIMC method was the most time-consuming (average, 15.3 minutes per case). Assessments using ARM took an average of 7.6 minutes per case, which was approximately one-half the time of the PIMC method. Using the Kendall tau-b correlation (τ), there was a strong, positive correlation noted between time spent scoring and the Ki-67 score for the ARM ($\tau = .536$; $P < .001$) and DIA-WSI (FOV) ($\tau = .408$; $P < .001$) methods. There also was a positive relationship observed between time to quantify and the Ki-67 index for manual assessments ($\tau = .221$; $P = .024$). However, there was no statistically significant correlation noted between the time involved and the Ki-67 score for the DIA-WSI (entire slide) assessment method ($\tau = .069$; $P = .447$).

DISCUSSION

The Ki-67 index clearly plays a critical role in the pathologic assessment of and clinical decisions related to the management of gastroenteropancreatic NETs.[24,25] The variabilities in different scoring methods, as well as interlaboratory and interobserver reproducibility, have been addressed in part through the use of standardized guidelines from the WHO and European Neuroendocrine Tumor Society (ENETS) for calculating the Ki-67 proliferation index in histopathology material.[1-4,24] These guidelines have been extrapolated to grading NET samples (eg, cell blocks) procured by FNA. However, several studies have shown that the grading of cell blocks using the Ki-67 proliferation index frequently results in undergrading of these tumors when compared with follow-up surgical resection specimens.[12] This fact was corroborated in the current study, in which the majority of NETs that were graded as 1 on cell blocks as per the original cytology reports subsequently were found to be grade 2 in the subset

of surgically resected tumors (19 tumors). Moreover, even when we used the PIMC or computer-assisted (ARM or DIA-WSI) methods for the grading of cell blocks, they still led to undergrading. Similarly, Tang et al reported that EE, manual counting, and even DIA all can provide incorrect Ki-67 assessments and therefore false grading of NETs.[15] This suggests that FNA does not lend itself to the accurate grading of NETs, which may be due to limited sampling and tumor heterogeneity.[26] For FNA samples that contain ≥1000 cells, one group of researchers demonstrated that counting hot spots instead of the complete cell block provided better correlation with surgical specimens.[17] We did analyze the same 3 hot spots using different methods, and found excellent reliability among these methods based on these hot spots versus analyzing just 1 or 2 of the FOVs. One limitation of the current study was the relatively small number of follow-up surgical resection specimens (19 specimens), which may have influenced effective correlation with cytology specimens.

As noted earlier, several methods have been proposed for determining the Ki-67 proliferation index in NETs. These include the manual EE method using a light microscope, the PIMC method that requires pathologists to print out photographs to count positive cells in hot spots, and DIA techniques. The data from the current study indicated that the level of reliability for these different methods when tested on cell block material was good to excellent for Ki-67 quantification. However, the average Ki-67 scores did differ with these various methods. It is interesting to note that the pairwise kappa result demonstrated near-perfect agreement for the PIMC and ARM methods. When using the Ki-67 score obtained from the 19 surgical pathology cases as the "gold standard," we found that the EE method of scoring cell blocks had the highest concordance rate. This result is surprising because EE is subjective and therefore prone to high interobserver variability. Indeed, prior studies have reported that EEs of pancreatic NETs were inaccurate.[27,28] To overcome this problem, Reid et al recommended the PIMC method.[19] These authors found in their experience that the manual counting of camera-captured and/or printed images was more reliable and had higher reproducibility than DIA, but took longer than EE. We also found the PIMC method to be the most time-consuming in the current study. However, Dogukan et al mastered this laborious task by using a monitor image instead of the printout image method for Ki-67 scoring.[29] These

TABLE 6. Amount of Time Spent to Quantify Ki-67 According to Different Scoring Methods

Scoring Method	Average Time, Minutes	Median Time, Minutes	Minimum Time, Minutes	Maximum Time, Minutes
PIMC	15.3	11.5	2.5	62.0
ARM	7.6	4.0	0.5	60.0
DIA-WSI (FOV)	2.6	3.2	0.3	3.5
DIA-WSI (entire slide)	14.2	9.4	0.7	46.6

Abbreviations: ARM, augmented reality microscopy; DIA, digital image analysis; FOV, field of view; PIMC, printed image manual counting; WSI, whole-slide imaging.
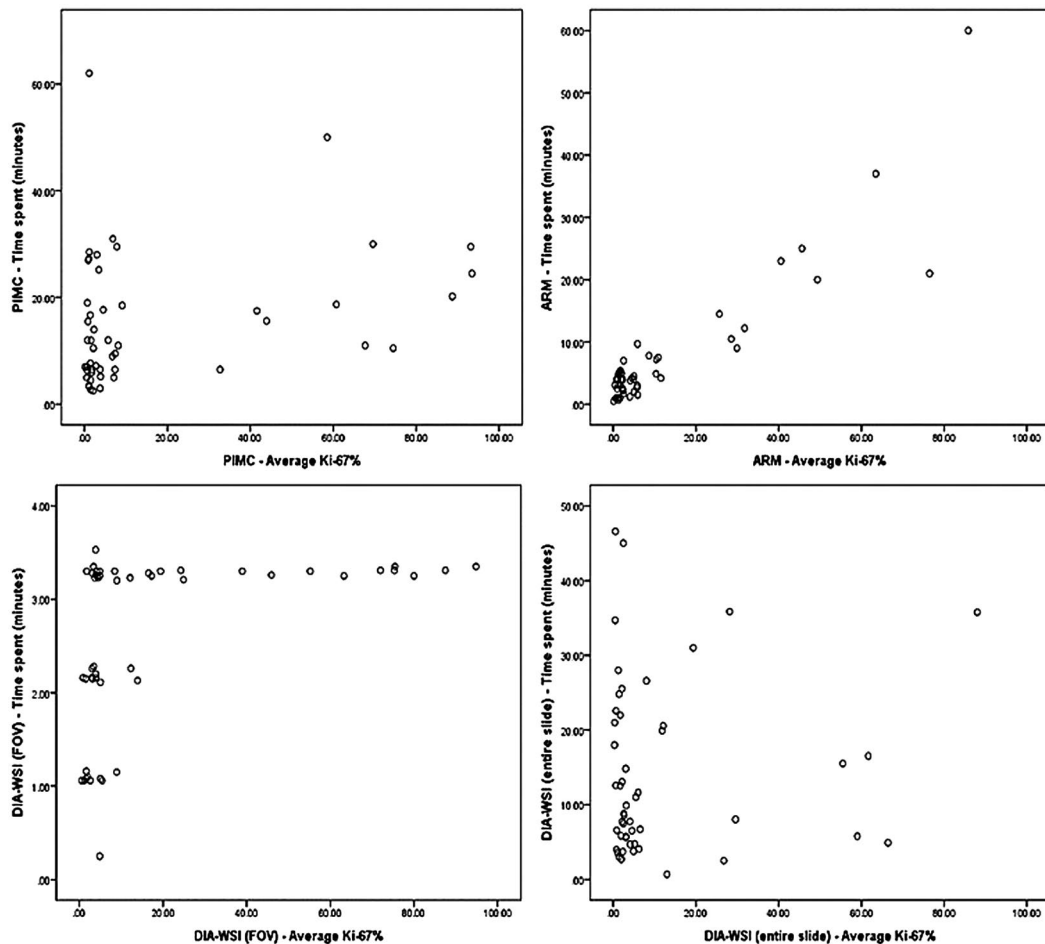

FIGURE 3. Scatterplots of time spent versus the Ki-67 index for different assessment methods. ARM indicates augmented reality microscopy; DIA, digital image analysis; FOV, field of view; PIMC, printed image manual counting; WSI, whole-slide imaging.

authors demonstrated that although monitor image and printout image methods for Ki-67 scoring of gastroenteropancreatic NETs resulted in comparable scores, they were able to save time performing this task by using a grid and a monitor.

For assessing biomarkers in breast cancer (eg, ER, PR, or HER2), DIA has been shown to outperform manual scoring.[30] Similar supportive findings using DIA have been reported for determining the Ki-67 labeling index in patients with breast cancer.[31] A limited number of publications to date have reported outcomes when using commercial and open-source (eg, ImmunoRatio http://wsiserver.jilab.fi/old-jvsmicroscope-software/) image analysis software to calculate the Ki-67 index in NETs.[16,19,32,33] However, for this task, the majority of these authors conveyed that manual counting was more accurate than DIA. Some of the reasons for this discordance with DIA included nontumor cell contamination (eg, proliferating

tumor–infiltrating lymphocytes) and background nonspecific staining. In the current study, we also noted inaccuracies in some cases when individual cells were not counted using the image analysis algorithm applied via ARM. The ARM vendor (Augmentiqs) in the current study has since modified various parameters of the image analysis algorithm for quantifying nuclear staining, and therefore improved results are expected in future studies. The majority of the prior studies using DIA to determine the Ki-67 index in NETs used static images of representative tumor regions. Very few researchers used WSI. Hasegawa et al did use WSI with DIA to evaluate the Ki-67 index in 58 patients and reported high concordance (90%) with surgically resected specimens, but only when their FNA samples had adequate cellularity (>2000 tumor cells).[9] WSI most likely aided their findings by allowing a greater percentage of the tumor to be analyzed, instead of just limiting their analyses to smaller FOVs (ie, snapshots). This also may explain why in the current study DIA-WSI of the entire slide demonstrated a better concordance with surgical resection specimens than the PIMC method.

To the best of our knowledge, the current study is the first to evaluate the use of ARM in cytology. Augmented reality (AR) refers to technology that combines reality and digital information.[34] This is created by superimposing a computer-generated digital image onto an object or user's view of the "real world." This differs from virtual reality, in which a complete digital or computer-generated environment is created. AR technology (eg, HoloLens; Microsoft Corporation, Redmond, Washington) has been applied successfully to anatomic pathology for unique applications such as 3-dimensional image viewing and real-time pathology-radiology correlation.[35] Recently, investigators introduced the novel AR microscope.[22] By attaching an AR unit to a conventional microscope, this accessory device converts the microscope into a digital pathology solution that now can be used to perform real-time telepathology and computer-assisted diagnostics (eg, image analysis and the application of artificial intelligence algorithms) in addition to AR.[22,23,36] In the current study, the AR microscope permitted us to rapidly execute image analysis to quantify Ki-67 directly from cytology glass slides while they were present on the microscope's stage. This allowed us to avoid having to first photograph these slides or digitize the entire slide with a whole-slide scanner. Moreover, because we were able to adjust the microscope's fine and coarse magnification

in real time during this undertaking, we overcame any focus issues that typically plague the digital imaging of cytology material. In summary, we found that ARM with DIA when directly supervised by a human streamlined and hastened the task of assessing the Ki-67 proliferation index in NETs. Given the versatility and cost benefit of microscope-based AR technology, we anticipate that there will be other innovative studies using ARM in the near future.

FUNDING SUPPORT

CONFLICT OF INTEREST DISCLOSURES

AUTHOR CONTRIBUTIONS
Swati P. Satturwar: Data collection, methodology (performed manual Ki-67 quantification), and writing–original draft. **Joshua L. Pantanowitz:** Methodology (performed Ki-67 quantification using an augmented reality microscope). **Christopher D. Manko:** Methodology (performed Ki-67 quantification using an augmented reality microscope). **Lindsey Seigh:** Digital analysis. **Sara E. Monaco:** Conceptualization, writing–original draft, and writing–review and editing. **Liron X. Pantanowitz:** Conceptualization, supervision, validation, and writing–review and editing.

REFERENCES
1. Fang JM, Shi J. A clinicopathologic and molecular update of pancreatic neuroendocrine neoplasms with a focus on the new World Health Organization classification. *Arch Pathol Lab Med.* 2019;143:1317-1326.
2. Guilmette JM, Nose V. Neoplasms of the neuroendocrine pancreas: an update in the classification, definition and molecular genetic advances. *Adv Anat Pathol.* 2019;26:13-30.
3. Klimstra DA, Kloppel G, La Rosa S, Rindi G. The WHO Classification of Tumors: Digestive System Tumors. 5th ed. IARC Press; 2019.
4. Sigel CS. Advances in the cytologic diagnosis of gastroenteropancreatic neuroendocrine neoplasms. *Cancer Cytopathol.* 2018;126:980-991.
5. Atiq M, Bhutani MS, Bektas M, et al. EUS-FNA for pancreatic neuroendocrine tumors: a tertiary cancer center experience. *Dig Dis Sci.* 2012;57:791-800.
6. Unno J, Kanno A, Masamune A, et al. The usefulness of endoscopic ultrasound–guided fine-needle aspiration for the diagnosis of pancreatic neuroendocrine tumors based on the World Health Organization classification. *Scand J Gastroenterol.* 2014;49:1367-1374.
7. Fujimori N, Osoegawa T, Lee L, et al. Efficacy of endoscopic ultrasonography and endoscopic ultrasonography–guided fine-needle aspiration for the diagnosis and grading of pancreatic neuroendocrine tumors. *Scand J Gastroenterol.* 2016;51:245-252.
8. Grosse C, Noack P, Silye R. Accuracy of grading pancreatic neuroendocrine neoplasms with Ki-67 index in fine-needle aspiration cell-block material. *Cytopathology.* 2019;30:187-193.
9. Hasegawa T, Yamao K, Hijioka S, et al. Evaluation of Ki-67 index in EUS-FNA specimens for the assessment of malignancy risk in pancreatic neuroendocrine tumors. *Endoscopy.* 2014;46:32-38.

10. Weynand B, Borbath I, Bernard V, et al. Pancreatic neuroendocrine tumour grading on endoscopic ultrasound–guided fine needle aspiration: high reproducibility and inter-observer agreement of the Ki-67 labelling index. *Cytopathology*. 2014;25:389-395.

11. Weiss VL, Kiernan C, Wright J, Merchant NB, Coogan AC, Shi C. Fine-needle aspiration–based grading of pancreatic neuroendocrine neoplasms using Ki-67: is accurate WHO grading possible on cytologic material? *J Am Soc Cytopathol*. 2018;7:154-459.

12. Farrell JM, Pang JC, Kim GE, Tabatabai ZL. Pancreatic neuroendocrine tumors: accurate grading with Ki-67 index on fine-needle aspiration specimens using the WHO 2010/ENETS criteria. *Cancer Cytopathol*. 2014;122:770-778.

13. Boutsen L, Jouret-Mourin A, Borbath I, van Maanen A, Weynand B. Accuracy of pancreatic neuroendocrine tumour grading by endoscopic ultrasound–guided fine needle aspiration: analysis of a large cohort and perspectives for improvement. *Neuroendocrinology*. 2018;106:158-166.

14. Goodell PP, Krasinskas AM, Davison JM, Hartman DJ. Comparison of methods for proliferative index analysis for grading pancreatic well-differentiated neuroendocrine tumors. *Am J Clin Pathol*. 2012;137:576-582.

15. Tang LH, Gonen M, Hedvat C, Modlin IM, Klimstra DS. Objective quantification of the Ki67 proliferative index in neuroendocrine tumors of the gastroenteropancreatic system: a comparison of digital image analysis with manual methods. *Am J Surg Pathol*. 2012;36:1761-1770.

16. Jin M, Roth R, Gayetsky V, Niederberger N, Lehman A, Wakely PE Jr. Grading pancreatic neuroendocrine neoplasms by Ki-67 staining on cytologic cell blocks: manual count and digital image analysis of 58 cases. *J Am Soc Cytopathol*. 2016;5:286-295.

17. Abi-Raad R, Lavik JP, Barbieri AL, Zhang X, Adeniran AJ, Cai G. Grading pancreatic neuroendocrine tumors by Ki-67 index evaluated on fine-needle aspiration cell block material. *Am J Clin Pathol*. 2020;153:74-81.

18. Niazi MKK, Tavolara TE, Arole V, Hartman DJ, Pantanowitz L, Gurcan MN. Identifying tumor in pancreatic neuroendocrine neoplasms from Ki67 images using transfer learning. *PLoS One*. 2018;13:e0195621.

19. Reid MD, Bagci P, Ohike N, et al. Calculation of the Ki67 index in pancreatic neuroendocrine tumors: a comparative analysis of four counting methodologies. *Mod Pathol*. 2015;28:686-694.

20. Van Velthuysen ML, Groen EJ, Sanders J, Prins FA, van der Noort V, Korse CM. Reliability of proliferation assessment by Ki-67 expression in neuroendocrine neoplasms: eyeballing or image analysis? *Neuroendocrinology*. 2014;100:288-292.

21. Volynskaya Z, Mete O, Pakbaz S, Al-Ghamdi D, Asa SL. Ki67 quantitative interpretation: insights using image analysis. *J Pathol Inform*. 2019;10:8.

22. Chen PC, Gadepalli K, MacDonald R. An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis. *Nat Med*. 2019;25:1453-1457.

23. Razavian N. Augmented reality microscopes for cancer histopathology. *Nat Med*. 2019;25:1334-1336.

24. O'Toole D, Kianmanesh R, Caplin M. ENETS 2016 Consensus Guidelines for the Management of Patients with Digestive Neuroendocrine Tumors: an update. *Neuroendocrinology*. 2016;103:117-118.

25. Sadot E, Reidy-Lagunes DL, Tang LH, et al. Observation versus resection for small asymptomatic pancreatic neuroendocrine tumors: a matched case-control study. *Ann Surg Oncol*. 2016;23:1361-1370.

26. Grillo F, Valle L, Ferone D, et al. KI-67 heterogeneity in well differentiated gastro-entero-pancreatic neuroendocrine tumors: when is biopsy reliable for grade assessment? *Endocrine*. 2017;57:494-502.

27. Young HT, Carr NJ, Green B, Tilley C, Bhargava V, Pearce N. Accuracy of visual assessments of proliferation indices in gastroenteropancreatic neuroendocrine tumours. *J Clin Pathol*. 2013;66:700-704.

28. Cottenden J, Filter ER, Cottreau J. Validation of a cytotechnologist manual counting service for the Ki67 index in neuroendocrine tumors of the pancreas and gastrointestinal tract. *Arch Pathol Lab Med*. 2018;142:402-407.

29. Dogukan FM, Yilmaz Ozguven B, Dogukan R, Kabukcuoglu F. Comparison of monitor-image and printout-image methods in Ki-67 scoring of gastroenteropancreatic neuroendocrine tumors. *Endocr Pathol*. 2019;30:17-23.

30. Stalhammar G, Fuentes Martinez N, Lippert M, et al. Digital image analysis outperforms manual biomarker assessment in breast cancer. *Mod Pathol*. 2016;29:318-329.

31. Maeda I, Abe K, Koizumi H, et al. Comparison between Ki67 labeling index determined using image analysis software with virtual slide system and that determined visually in breast cancer. *Breast Cancer*. 2016;23:745-751.

32. Hacking SM, Sajjan S, Lee L, et al. Potential pitfalls in diagnostic digital image analysis: experience with Ki-67 and PHH3 in gastrointestinal neuroendocrine tumors. *Pathol Res Pract*. 2020;216:152753.

33. Hwang HS, Kim Y, An S, et al. Grading by the Ki-67 labeling index of endoscopic ultrasound–guided fine needle aspiration biopsy specimens of pancreatic neuroendocrine tumors can be underestimated. *Pancreas*. 2018;47:1296-1303.

34. Berryman DR. Augmented reality: a review. *Med Ref Serv Q*. 2012;31:212-218.

35. Hanna MG, Ahmed I, Nine J, Prajapati S, Pantanowitz L. Augmented reality technology using Microsoft HoloLens in anatomic pathology. *Arch Pathol Lab Med*. 2018;142:638-644.

36. Siegel G, Regelman D, Maronpot R, Rosenstock M, Hayashi SM, Nyska A. Utilizing novel telepathology system in preclinical studies and peer review. *J Toxicol Pathol*. 2018;31:315-319.