Endocrine Pathology (2019) 30:17–23
https://doi.org/10.1007/s12022-018-9554-3

Comparison of Monitor-Image and Printout-Image Methods in Ki-67 Scoring of Gastroenteropancreatic Neuroendocrine Tumors

Fatih Mert Dogukan[1] · Banu Yilmaz Ozguven[2] · Rabia Dogukan[1] · Fevziye Kabukcuoglu[2]

Published online: 26 October 2018

Abstract

Gastroenteropancreatic neuroendocrine tumors (GEP-NET) are classified according to tumor grade. Ki-67 and mitotic count are the two determinants of this classification. Therefore, Ki-67 scoring becomes very important in classifying the patients accurately. Eye-balling, counting of cells through the microscope, automated image analysis systems, and manual counting of printed image are the four major scoring methods in use. The aim of this study is to show the agreement between monitor-image method (MIM) and printout-image method (PIM) of Ki-67 scoring. In our study, 120 GEP-NETs from 85 patients diagnosed between January 2005 and July 2017 were evaluated. Thirty-seven cases with either polypectomy or resection material were selected. Seven different scoring methods using either a monitor-image or a printout-image were applied for Ki-67 scoring. They are as follows: whole-PIM, 1/9-PIM, whole-MIM, 1/4-MIM, 1/6-MIM, 1/9-MIM, and 1/12-MIM. In the comparison of Ki-67 scoring methods, intraclass correlation coefficients ranging from 0.951 to 0.999 were found. The Bland-Altman analysis showed near-perfect agreement between whole-MIM and whole-PIM as well as 1/9-MIM and 1/9-PIM. The level of agreements among the other methods were sufficient too, but there was a relative decrease in the level of agreement as the area of counting becomes smaller. The average application time decreased from 373.7 to 41.7 s gradually as the scoring area becomes smaller. Our study shows that there is a remarkable agreement between the MIM and PIM used in Ki-67 scoring.

Keywords Neuroendocrine tumors · Ki-67 · Gastrointestinal neoplasms · Pancreatic neoplasms · Tumor grading

Introduction

Gastroenteropancreatic neuroendocrine neoplasms (GEP-NEN) represent a heterogeneous group of tumors arising from the neuroendocrine cells of the gastrointestinal tract and pancreas [1]. Although all GEP-NENs are defined as malignant, their prognosis is widely variable [1]. Taking this prognostic diversity into account, a neuroendocrine neoplasm classification has been established which is based on both the differentiation and the proliferation rates of the tumors [1–4]. Accordingly, poorly differentiated lesions are diagnosed as neuroendocrine carcinoma and well-differentiated neuroendocrine tumors are categorized into three grades by their mitotic rates and Ki-67 scores [1–3]. Grade 1 tumors display a mitotic rate of < 2 per 2 mm^2 or a Ki-67 score of < 3%; grade 2 tumors have a mitotic rate of 2–20 per 2 mm^2 or a Ki-67 score of 3–20%, and the tumors with mitotic rates or Ki-67 values above these upper thresholds are called grade 3 [1, 2, 5].

Ki-67 is an immunohistochemical marker expressed by the proliferating cells during the S, G2, or M phases of the cell cycle [1]. World Health Organization (WHO) classifications of neuroendocrine neoplasms, standard reporting protocols of College of American Pathologists (CAP), and European Neuroendocrine Tumor Society (ENETS) describe Ki-67 as an independent prognostic parameter and a major determinant of the tumor grade [1–5]. Therefore, the standardization of Ki-67 staining and scoring procedures becomes essential to obtain more accurate and significant results.

In Ki-67 scoring, at least 500–2000 cells must be counted and if the total number of cells is less than 500, this must be specified in the report [1, 5]. In addition, approaches like synaptophysin/Ki-67 double-staining may be helpful in eliminating false-positive signals [6].

Several Ki-67 scoring methods have been proposed and eye-balling, counting of cells through the microscope, automated

✉ Fatih Mert Dogukan
fmdogukan@hotmail.com

[1] Department of Pathology, Mardin State Hospital, Mardin, Turkey

[2] Department of Pathology, University of Health Sciences Sisli Hamidiye Etfal Education and Research Center, Istanbul, Turkey

image analysis systems, and manual counting of printed images represent the four major scoring methods in use today [7, 8]. Automated systems are expensive and user-dependent methods that are not widely accessible [9]. The once-accepted eye-balling technique is not suggested anymore due to its inadequacy in tumor grading, especially for low-grade tumors [8, 10, 11]. Manual counting of printed images of tumor hot spots is currently the most recommended Ki-67 scoring method [7]. Although this method is quite practical and allows more accurate results, it has not been preferred enough by the practitioners due to its long application time [7, 11, 12]. A recent modification called "the acetate grid method" has significantly shortened the application period [12]. This modification involves a gridded acetate sheet placed on the A4-printed hot spot image and an estimated total cell number that is obtained by multiplying the total cell count of the most cellular grid zone with the total number of zones [12]. Thus, Ki-67 scoring by manual counting becomes less time-consuming without a significant decrease in accuracy [12]. It is also mentioned that both gridding and manual counting can be done by using a computer instead of a print-out [12]. However, to date, there is no published article considering this subject.

The aim of this study is to show that monitor-images and camera software can be used in Ki-67 scoring as well as camera captured/printed out images. In addition, Ki-67 scores will be calculated by using grids with different partition coefficients in order to investigate the possible effects of gridding on the ultimate Ki-67 score. Consequently, an alternative method of Ki-67 scoring which uses gridded monitor-images and does not require any printer or paper, will be documented for the first time.

Material and Methods

Case Selection

A retrospective review of GEP-NENs at University of Health Sciences Sisli Hamidiye Etfal Education and Research Center, Istanbul, Turkey was carried out. One hundred twenty tumors from 85 patients diagnosed between January 2005 and July 2017 were evaluated. After the exclusion of biopsy-only specimens, neuroendocrine carcinomas and appendiceal neuroendocrine tumors, 37 cases with resection/polypectomy material (23 stomach, 6 ileum, 5 pancreas, 2 colorectum, 1 distal biliary tract) were included. Appendiceal neuroendocrine neoplasms were excluded due to the highly inflamed background that could interfere with staining and scoring.

Ki-67 Scoring

Ki-67 slides (clone MIB-1, 1/100, Dako, Carpinteria, CA) were examined by using a DS-Fi2 camera and DS-L3 camera control unit-assisted Nikon Eclipse Ni-U microscope. Tumor components with the highest staining (hot spots) were identified at × 100 and counted at × 400 magnification. The image on the monitor of the control unit was a rectangle with an area of 0.073 mm^2 (0.234 mm × 0.310 mm). This area is smaller than the actual circular image encountered by the naked eye at the same magnification (0.238 mm^2). The monitor-image was divided into nine equally sized rectangular fields by inserting the calculated measures into "the grid mode" of the control unit and captured to print out later. After the counting of all Ki-67 positive and negative cells on the whole image by whole monitor-image method (W-MIM), the field with the highest cellularity on the 1/9 divided image was determined. The cell count of this field was then multiplied by 9 to have an estimated number of total cellularity (1/9-MIM). Similarly, this procedure was repeated to divide the monitor image into 4, 6, and 12 equally sized fields and to define the estimated total cell numbers for each approach (1/4, 1/6, and 1/12-MIMs, respectively) (Fig. 1).

The previously captured images of 1/9 grid mode were printed on A4 paper. In whole printout-image method (W-PIM), all positive and negative cells were counted manually by using a pen to mark each cell in order to prevent double-counting. The field with the highest cellularity was determined to calculate the estimated total cell number again (1/9-PIM). All counts were performed by the same person (FMD) in a case-blinded fashion.

An extra hot spot was defined and evaluated if the total cell count was lower than 500 cells in the first one. A single hot spot was sufficient to reach this threshold in 24 cases, whereas a second one was included in the remaining 13 cases.

At the end of the cell counting process, the total cell numbers with seven different counting models were reached and a unique Ki-67 score was calculated for each case by each modality (Table 1). The values obtained were rounded to the nearest two decimal places. In each case, the time spent for each method was measured by a chronometer and recorded in seconds.

Statistical Analysis

Statistical analysis was carried out by using the Statistical Package for Social Sciences software version 15.0 (SPSS; Chicago, IL, USA). Values were submitted as mean ± standard deviation. The level of agreement among the Ki-67 scoring methods was tested by the intraclass correlation coefficient (ICC) and the Bland-Altman plot [13, 14]. McGraw and Wong's "two-way mixed effects, absolute agreement, single rater/measurement" form of ICC was used [13]. Counting durations of scoring methods were compared using the paired *t* test. *P* values lower than < 0.05 were considered statistically significant.

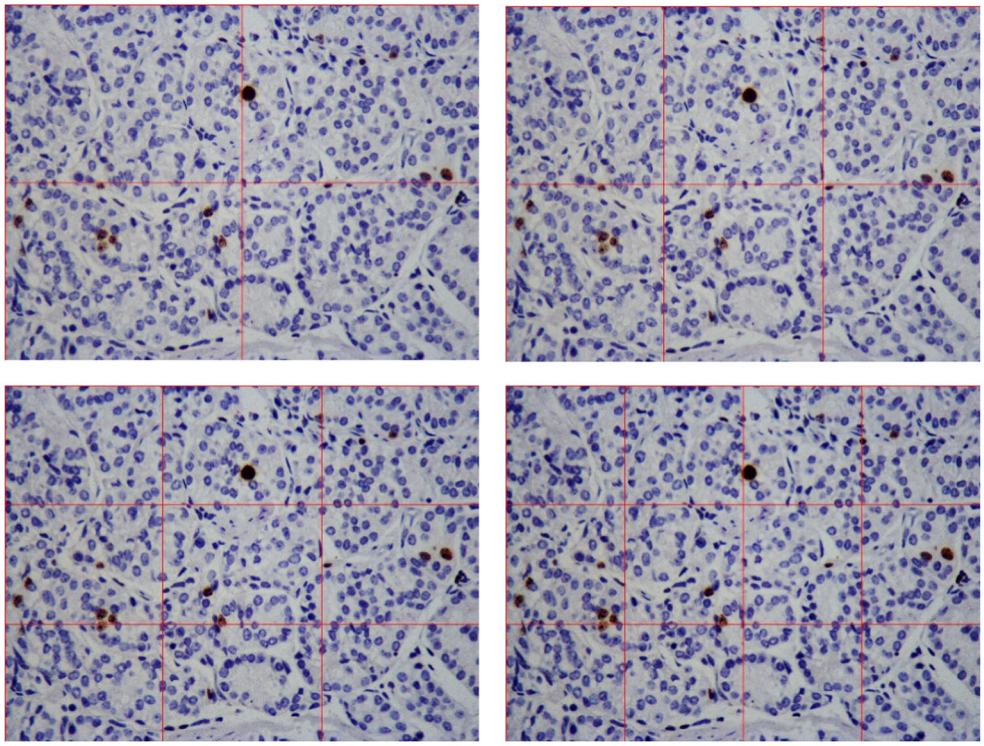

Fig. 1 Samples of monitor-images of a single GEP-NEN case used to calculate Ki-67 scores; 1/4, 1/6, 1/9, 1/12 grid modes (grid lines are in red)

Results

A group of 37 GEP-NEN cases meeting the criteria described above were included in the study. The mean age was 54.2 ± 13.2 years. Nineteen (51.4%) of the cases were female and 18 (48.6%) were male. In 24 cases (64.9%), the lowest threshold value (500 cells) set for the Ki-67 count was exceeded by counting only one hot spot, whereas a second hot spot had to be counted in 13 cases (35.1%). Thirty-one cases (83.8%) were grade 1, five cases were grade 2, and only one case was grade 3.[1]

Ki-67 scores of the 37 cases were calculated in seven distinct scoring models. The real and the estimated total cell count averages were 755.8 and 992.6 by the W-PIM and 1/9-PIM respectively. The mean Ki-67 scores calculated by the W-PIM and 1/9-PIM were 1.96 and 1.50%. Starting from the W-MIM, the average total cell numbers obtained with 1/4, 1/6, 1/9, and 1/12-MIMs were 715.2, 796.5, 871.8, 947.0, and 1033.6 respectively. Accordingly, the mean Ki-67 scores were 1.99%, 1.82%, 1.68%, 1.53%, and 1.44% in the same order (Table 2). There was an increasing trend in the total cell counts as the area to be counted became smaller. Since there is an inverse relationship between the total cell count and the Ki-67 score,[2] this resulted in a proportional decrease in the Ki-67 scores (Fig. 2).

In all the PIMs and MIMs, the average time spent on the Ki-67 scoring was reduced as the area to be counted narrowed. As expected, the average counting time was gradually reduced from 373.7 to 53.6 s for the PIMs and from 357.9 to 41.7 s for the MIMs (Table 2). The difference of the time spent was statistically significant for each comparison when the image partitioning coefficients of the comparative methods were different from each other ($p < 0.001$).

The kappa values and the confidence intervals of ICCs displayed a high to near-perfect agreement between any PIM or MIM in terms of Ki-67 scores in general. More specifically, the high to near-perfect agreement found between the previously validated W-PIM and 1/9-PIM (0.956 (0.883–0.981)) and the almost perfect agreement found between the W-MIM and W-PIM were notable (0.999 (0.999–1000)) (Table 3).

Bland-Altman plots were drawn to compare the results of the scoring methods based on the mean values and the differences of the Ki-67 scores for each case. In the plots of W-PIM/

Table 1 Ki-67 scoring methods

Printout-image methods	Monitor-image methods
Whole-image counting	Whole-image counting
1/9—image counting	1/4—image counting
	1/6—image counting
	1/9—image counting
	1/12—image counting

[1] Considering the Ki-67 values obtained by W-PIM

[2] The total cell count being the denominator in the calculation

Table 2 The mean values of total cell counts, Ki-67 scores and time spent for all Ki-67 scoring methods

	Total cell count (*n*)	Ki-67 (%)	Time (s)
W-PIM	755.8 ± 182.9	1.96 ± 3.40	373.7 ± 93.1
1/9-PIM	992.6 ± 261.4	1.50 ± 2.62	53.6 ± 14.1
W-MIM	715.2 ± 161.7	1.99 ± 3.31	357.9 ± 80.1
1/4-MIM	796.5 ± 177.0	1.82 ± 3.16	105.9 ± 22.9
1/6-MIM	871.8 ± 201.4	1.68 ± 2.89	74.6 ± 18.5
1/9-MIM	947.0 ± 229.9	1.53 ± 2.64	53.2 ± 14.2
1/12-MIM	1033.6 ± 261.8	1.44 ± 2.61	41.7 ± 9.8

W whole, *PIM* printout-image method, *MIM* monitor-image method

W-MIM and 1/9-PIM/1/9-MIM, the mean of differences (bias) were 0.03 and 0.02 and the limits of agreement values were from 0.25 to − 0.19 and from 0.26 to − 0.22 respectively. Accordingly, 97 and 95% of the cases were located among the limits of agreement (Fig. 3–4). In plots where W-MIM is compared with 1/4, 1/6, 1/9, and 1/12-MIMs, the means of differences (bias) were 0.17, 0.31, 0.5, and 0.6 respectively. The limits of agreement values were from 0.66 to − 0.31, from 1.19 to − 0.57, from 1.8 to 0.9 and from 2.0 to 0.9 in the same order. The plots of W-MIM/1/9-MIM and W-MIM/1/12-MIM showed a proportional error as well.

Discussion

Gastroenteropancreatic neuroendocrine neoplasms represent a heterogenous group of tumors featuring a high prognostic variance [1]. In GEP-NEN classification systems, tumor differentiation, mitotic rate, and Ki-67 score constitute the three decisive parameters [1, 2, 4]. Therefore, a standardized Ki-67 scoring is of paramount importance to classify and grade these tumors accurately [9].

Table 3 Intraclass correlation coefficients of Ki-67 scoring methods

	ICC (95% CI)
W-MIM vs. W-PIM	0.999 (0.999–1000)
1/4-MIM vs. W-MIM	0.996 (0.984–0.999)
1/6-MIM vs. W-MIM	0.985 (0.945–0.994)
1/9-MIM vs. W-MIM	0.962 (0.875–0.984)
1/12-MIM vs. W-MIM	0.954 (0.820–0.982)
1/9-PIM vs. W-PIM	0.956 (0.883–0.981)
1/4-MIM vs. W-PIM	0.995 (0.988–0.998)
1/6-MIM vs. W-PIM	0.982 (0.956–0.992)
1/9-MIM vs. W-PIM	0.958 (0.892–0,981)
1/9-MIM vs. 1/9-PIM	0.999 (0.998–0.999)
1/12-MIM vs. W-PIM	0.951 (0.852–0.979)

W whole, *PIM* printout-image method, *MIM* monitor-image method, *ICC* intraclass correlation coefficient, *CI* confidence interval

A number of studies discussing Ki-67 scoring methods have been done in recent years, especially after the spotlighting of Ki-67 in the WHO 2010 classification for gastrointestinal neuroendocrine tumors [1]. Eye-balling, counting of cells seen through the microscope by naked eye, counting by automated image analysis systems, and manual counting of camera captured/printed images are the four major scoring methods in use [7, 8]. Although automated counting systems are promising and preferred by some, these are expensive and not yet widespread [8]. Softwares for this purpose need careful calibration and further work for standardization [5]. Manual counting using printed images stands out to be optimal in terms of both practicality and accuracy and specified as the primarily recommended technique in general [5, 7, 12]. Nevertheless, pathologists have been reluctant to use this method due to the long application time, printer requirement and the commitment of performers to the conventional "eye-balling" technique [7, 11, 12]. However, "the acetate grid

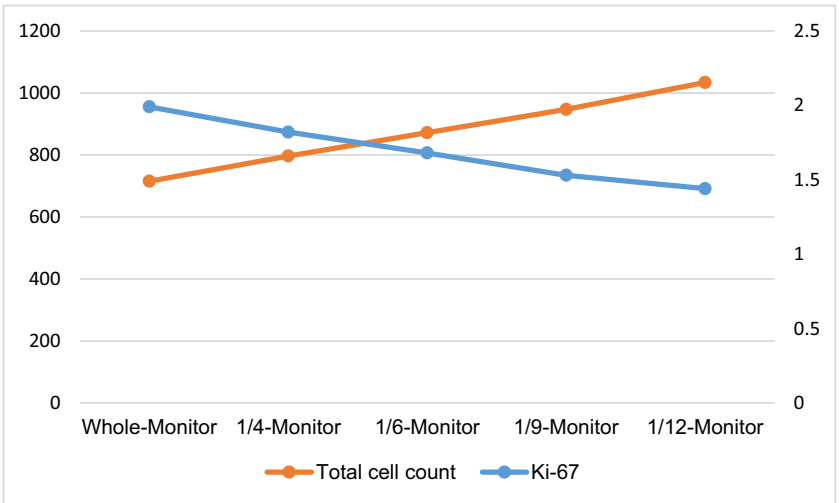

Fig. 2 Course of total cell counts and Ki-67 scores as the counting area becomes smaller

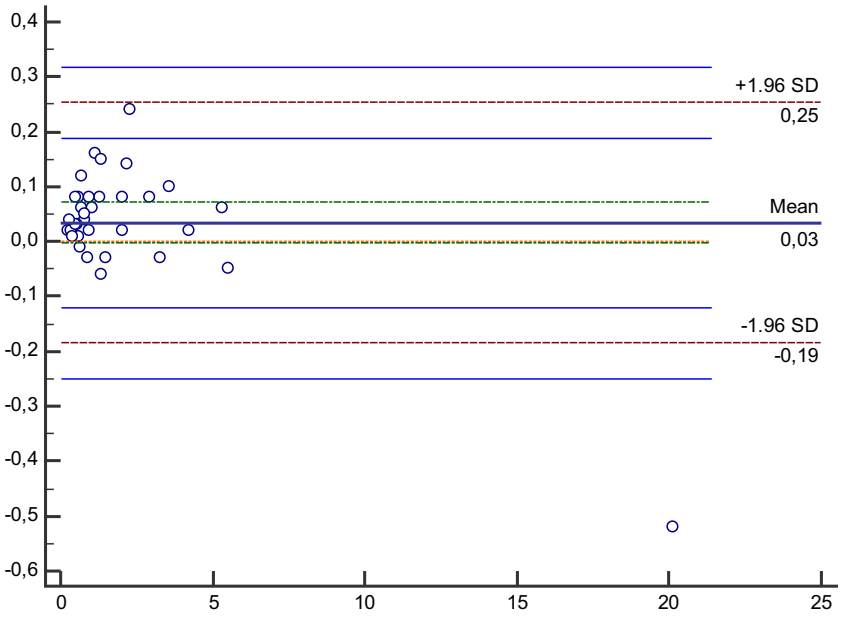

Fig. 3 Bland-Altman plot of Ki-67 scores obtained by whole monitor-image and whole printout-image scoring methods

method" proposed by Ozturk Sari et al. is a promising technical modification to make it preferable by shortening the application time [12]. Some previous reports considering this method have already mentioned that a computer screen may replace the printed image, but only in theory [7, 10, 12]. The present study becomes the first documentation of the monitor-assisted Ki-67 scoring method in the literature. In our study, a total of seven distinct Ki-67 scoring methods derived from the printed and monitor-image techniques were applied, and consequently, it has been shown that monitor images provide almost the same results as those of the printed images. Therefore, we concluded that using a monitor image in Ki-67 scoring is as preferable as using a printed image. In addition, the partition coefficients used to divide the image were compared with each other both in terms of time consumed and the agreement of the results to observe the effects of the gridding on the ultimate Ki-67 scores.

Increased partition coefficients yielded increasingly higher total cell counts. This is probably due to the fact that smaller areas are prone to represent compact hypercellular areas. In other words, cell intensity becomes higher. Considering this implication and the existing inverse relationship between the Ki-67 score and the total cell number, one can predict the inevitable decrease in Ki-67 scores as the counting area narrows. The level of consequent variations in the Ki-67 score will be decisive for the accuracy of the technique. Statistical

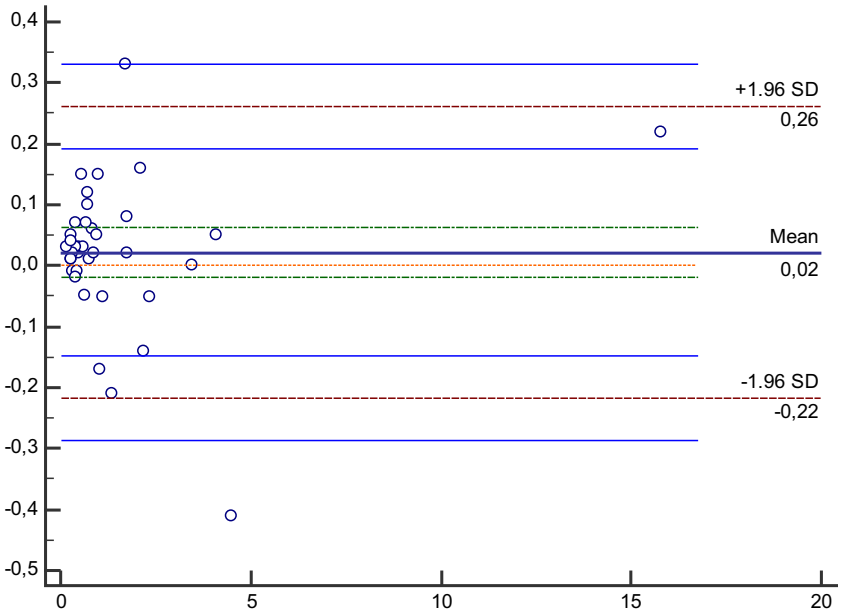

Fig. 4 Bland-Altman plot of Ki-67 scores obtained by 1/9 monitor-image and 1/9 printout-image scoring methods

methods like ICC and Bland-Altman plots were therefore needed to analyze these circumstances [13, 14].

The Ki-67 scoring methods in our study were highly correlated in terms of ICC values. There was a high-to-perfect agreement between the previously recommended W-PIM and 1/9-PIM (ICC = 0.956 (0.883–0.981)). The ICC values showed a perfect agreement in the comparisons of both W-PIM/W-MIM and 1/9-PIM/1/9-MIM as well (ICCs = 0.999 (0.999–1000) and 0.999 (0.998–0.999) respectively). So, the ICC values revealed a near-perfect agreement between the techniques used in the study of Ozturk Sari et al. and the MIMs that we proposed.

The Bland-Altman plot comparing the W-PIM and W-MIM showed a mean of differences (bias) of 0.03 and the agreement limits of 0.25 and − 0.19. Ninety-seven percent (97%) of the cases were in between the limits. Analyzing these statistical findings requires a clinical-based approach to show whether there is an agreement between the methods or the differences of the Ki-67 scores constitute a problem in clinical practice. In other words, Bland-Altman plot does not provide us the exact statistical values unlike ICCs; we instead interpret the clinical meaning of the results for the subject test. A mean difference of 0.03% in the Ki-67 scores implies that the disparities between the results of the two techniques are almost insignificant clinically. The second parameter we take is the limits of agreement which depict the range of variation that will be encountered when we use one method instead of the other. In the current Bland-Altman plot, the Ki-67 scores may become 0.25% higher or 0.19% lower when two methods are preferred interchangeably. At this point, it depends on the interpretation of the practicing pathologist again whether this variability will cause a problem in terms of the significance of the test. The last major determinant of the agreement is the distribution of the differences. The plots in which 95% or more of the cases are located in between the lower and upper limits of agreement, indicate a high level of agreement. In our study, all cases except one were among the limits (~97%) and that case had an extremely high Ki-67 score (20.12%) compared to the others. After all, it can be concluded that the W-PIM and W-MIM show a near-perfect agreement and the monitor image may be preferred as an alternative to the printed image in Ki-67 scoring.

Similarly, there was a high level of agreement between the scores rendered by the 1/9-PIM and 1/9-MIM. In their Bland-Altman plots, the mean of differences (bias) was 0.02 and the limits of agreement were 0.26 and − 0.22. Approximately, 95% of the cases were in the range of agreement. Therefore, the 1/9 gridding we have proposed as the equivalent of Ozturk Sari's new acetate grid method is shown to be applicable by using a monitor image as well [10].

Both the mean of differences (bias) and the limits of agreement tended to increase as the partition coefficients changed from 1/4 to 1/12 gradually. This implies that the level of agreement will decrease as we try to shorten counting times, regardless of the percentage of the cases staying between the limits of agreement. In order to prevent any clinically unwanted outcome, it is more important to know both the advantages and the disadvantages of each selected gridding option rather than to try to choose the best.

Another observation that may be useful for the practicing pathologist is the gradual increase of the mean Ki-67 scores from 1/12 to whole-image count. For example, a practitioner who selects 1/12 gridding technique and has a score of just over the threshold value (e.g., 3.1%) should know that the score would have been higher if it was scored with either 1/9 or 1/4 gridding techniques. However, this will not lead to any change in the tumor grade. In such a case, it may not be logical to reconsider the other scoring methods that take longer and are less practical. On the other hand, if the Ki-67 score is calculated by one of the 1/4, 1/6, 1/9, or 1/12 scoring methods and it falls short of reaching a threshold value by a small margin (e.g., 2.9%), the whole-image count should be carried out to ensure accuracy.

Conclusions

In our study, PIMs and MIMs for Ki-67 scoring were compared. Compatible Ki-67 scores were obtained both by the comparisons of whole image and 1/9 gridded versions of PIM and MIM. Accordingly, it has been shown that the Ki-67 scoring can be performed by using a monitor as well. However, it may provide more effective results to make these comparisons in larger series where the normal distribution of the data can be controlled. In addition, it is necessary to show the reproducibility of the MIMs by measuring the intra- and interobserver variability to validate the technique completely.

Using a grid makes the cell counting process significantly less time-consuming and more practical, but it is also shown that the smaller the count area, the lower the level of agreement becomes. Nevertheless, any method described above can be preferred under certain conditions for certain tumors. It is crucial for the pathologist to have a thorough understanding of the subtle effects each scoring method will have on the final Ki-67 score. All the methods described and compared herein need to be weighed and assessed for each case, just like any other diagnostic/prognostic test available to the pathologist.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Bosman FT, Carneiro F, Hruban RH, Theise ND (2010) WHO Classification of Tumours of the Digestive System, 4th edn. International Agency for Research on Cancer, Lyon.

2. Klöppel G, Couvelard A, Hruban RH, Klimstra DS, Komminoth P, Osamura RY, et al (2017) Neoplasms of the neuroendocrine pancreas. In: Lloyd RV, Osamura RY, Klöppel G, Rosai J (ed) WHO Classification of Tumours of Endocrine Organs, 4th edn. International Agency for Research on Cancer, Lyon.
3. Klöppel G, Rindi G, Perren A, Komminoth P, Klimstra D (2010) The ENETS and AJCC/UICC TNM classifications of the neuroendocrine tumors of the gastrointestinal tract and the pancreas: a statement. Virchows Arch 456(6):595–597. https://doi.org/10.1007/s00428-010-0924-6.
4. Rindi G, Klimstra DS, Abedi-Ardekani B, Asa SL, Bosman FT, Brambilla E, Busam KJ, de Krijger RR, Dietel M, el-Naggar AK, Fernandez-Cuesta L, Klöppel G, McCluggage WG, Moch H, Ohgaki H, Rakha EA, Reed NS, Rous BA, Sasano H, Scarpa A, Scoazec JY, Travis WD, Tallini G, Trouillas J, van Krieken JH, Cree IA (2018) A common framework for neuroendocrine neoplasms: an International Agency for Research on Cancer (IARC) and World Health Organization (WHO) expert consensus proposal. Mod Pathol. https://doi.org/10.1038/s41379-018-0110-y
5. Shi C, Adsay V, Bergsland EK, Berlin J, Branton PA, Fitzgibbons PL, et al (2017) Protocol for the examination of specimens from patients with neuroendocrine tumors (carcinoid tumors) of the stomach. College of American Pathologists Cancer Protocols. www.cap.org/cancerprotocols. Accessed 11 September 2018.
6. Matsukuma K, Olson KA, Gui D, Gandour-Edwards R, Li Y, Beckett L (2017) Synaptophysin-Ki67 double stain: a novel technique that improves interobserver agreement in the grading of well-differentiated gastrointestinal neuroendocrine tumors. Mod Pathol 30(4):620–629. https://doi.org/10.1038/modpathol.2016.225
7. Reid MD, Bagci P, Ohike N, Saka B, Erbarut Seven I, Dursun N, Balci S, Gucer H, Jang KT, Tajiri T, Basturk O, Kong SY, Goodman M, Akkas G, Adsay V (2015) Calculation of the Ki67 index in pancreatic neuroendocrine tumors: a comparative analysis of four counting methodologies. Mod Pathol 28(5):686–694. https://doi.org/10.1038/modpathol.2014.156.
8. Tang LH, Gonen M, Hedvat C, Modlin IM, Klimstra DS (2012) 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 36(12):1761–1770. https://doi.org/10.1097/PAS.0b013e318263207c
9. Remes SM, Tuominen VJ, Helin H, Isola J, Arola J (2012) Grading of neuroendocrine tumors with Ki-67 requires high-quality assessment practices. Am J Surg Pathol 36(9):1359–1363. https://doi.org/10.1097/PAS.0b013e3182632038.
10. Adsay V (2012) Ki67 labeling index in neuroendocrine tumors of the gastrointestinal and pancreatobiliary tract. Am J Surg Pathol 36(12):1743–1746. https://doi.org/10.1097/PAS.0b013e318272ff77.
11. Cottenden J, Filter ER, Cottreau J, Moore D, Bullock M, Huang WY, Arnason T (2018) Validation of a cytotechnologist manual counting service for the Ki67 index in neuroendocrine tumors of the pancreas and gastrointestinal tract. Arch Pathol Lab Med 142(3):402–407. https://doi.org/10.5858/arpa.2017-0203-OA.
12. Ozturk Sari S, Taskin OC, Yegen G, Ozluk Y, Gulluoglu M (2018) Ki67 scoring in pancreatic neuroendocrine tumors by a new method. Appl Immunohistochem Mol Morphol 26(4):283–287. https://doi.org/10.1097/PAI.0000000000000415.
13. McGraw KO, Wong SP (1996) Forming inferences about some intraclass correlation coefficients. Psychol Methods 1:30–46. https://doi.org/10.1037/1082-989X.1.1.30.
14. Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1:307–310.