Original paper

Tissue heterogeneity contributes to suboptimal precision of WHO 2010 scoring criteria for Ki67 labeling index in a subset of neuroendocrine neoplasms of the pancreas

Łukasz Liszka

Department of Pathomorphology, Medical University of Silesia, Katowice, Poland

Reporting of Ki67 labeling index (LI) is a routine in diagnostics of neuroendocrine neoplasms of the pancreas. The aim of the study was to examine whether heterogeneity of Ki67 LI distribution in primary tumoral tissue influences precision of reporting of Ki67 LI and Ki67-LI-based grade, both established in adherence to WHO 2010 guidelines. Seventy-one samples of neuroendocrine tumours (NET) and 6 samples of neuroendocrine carcinomas (NEC) of the pancreas were taken for manual counting of Ki67 LI in 25 portions of 100 cells (2500 cells in total) in 3 hot spots an in a single area of lower proliferation rate (cold spot) in each case. Both NET and NEC showed Ki67 LI heterogeneity within primary tumour. Almost 20% of NET showed higher grade when 500 cells rather than 2000 cells were counted in hot spot area. Suboptimal choice of hot spot resulted in under-grading of approximately 20% of NET. Cold spots were constantly present in NET. Heterogeneity of Ki67 LI was also present in NEC, but it virtually never resulted in under-grading. Concept and methodology of Ki67 LI counting in neuroendocrine neoplasms of the pancreas requires clarification. Efforts aiming to improve precision of assessment of Ki67 LI are needed.

Key words: Ki-67 antigen, neoplasm grading, neuroendocrine tumors, neuroendocrine carcinoma, pancreas, pancreatic neoplasms.

Introduction

Neuroendocrine neoplasms (NEN) constitute from 1 to 2% of neoplasms of the pancreas. The clinical course of pancreatic NEN is broad, ranging from "clinically benign" tumours effectively treated with enucleation to aggressive carcinomas diagnosed usually at the stage of locally advanced or metastatic disease [1]. Identification of reliable prognosticators as well as accurate predictors of effectiveness of medical treatment in NEN is necessary [2].

Ki67 protein expression has been utilized as an immunohistochemical (IHC) marker of proliferation across diversity of neoplastic and non-neoplastic diseases [3]. Assessment of Ki67 labeling index (LI) is now a routine procedure for pancreatic NEN [1, 4, 5]. Ki67 LI is the most important molecular biomarker used for optimizing clinical management of patients with NEN of the pancreas [6, 7].

Pancreatic NEN may show significant heterogeneity not only between patients, but also *within* tumoral tissue. Recognition of phenotypic tumour heterogeneity is important for adequate histopathological diagnostics and for understanding of ineffectiveness of therapy for cancer [8].

Assessment of Ki67 LI (in the context of conventional hematoxylin-eosin picture) is an elegant way to examine tissue heterogeneity in pancreatic NEN. Ki67 LI heterogeneity was observed not only within primary tumour tissue [9, 10, 11, 12], but also between primary and metastatic lesions [13, 14, 15] and within and between metastases [16, 17].

Currently we do not have any proliferation marker better than Ki67 in terms of prognostic significance in diagnostic histopathology of NEN [2, 11]. However, this does not necessarily indicate that Ki67 LI assessment is reliable (i.e. accurate and precise). Analytical validity is a *sine qua non* condition for any tumour biomarker which is intended to be used in clinical practice and to influence management of the patients [18]. Selected aspects of analytical validity of Ki67 LI in NEN were previously studied [19, 20, 21, 22, 23, 24, 25, 26, 27, 28].

Guidelines for Ki67 LI scoring in gastro-entero-pancreatic NEN were published (as detailed in Supplementary Information (SI) – Table S1) [1, 4, 5, 6, 11, 29-31]. According to the World Health Organization (WHO) 2010 guidelines, grading of gastro-entero-pancreatic neuroendocrine neoplasms includes documentation of mitotic count and Ki67 LI [1]. The latter should be reported as percentage of immuno-positive cells identified among from 500 to 2000 cells in areas of "strongest nuclear labeling" [1]. European Neuroendocrine Tumor Society (ENETS) 2006 guidelines stated that 2000 cells should be counted for Ki67 LI [4]. ENETS 2009 guidelines proposed to count 100 cells, and to evaluate several tumour areas in cases with uneven stain distribution [29]. However, the precision of scoring Ki67 LI in relation to the number of examined tumoral cells has not been extensively studied. Some investigators showed that grading of pancreatic NEN in cytological samples was less reliable (i.e. showed lower percentage of agreement with reference grade established in subsequent resection specimens) if number of examined cells was small [12, 32].

The aim of the present study was to report precision of manual assessment of Ki67 LI and Ki67-LI-based grade in histopathological samples of pancreatic NEN, in relation to the tumour area selected for scoring and to the number of counted cells, in close adherence to WHO 2010 guidelines.

Material and methods

Study samples

This study was based on examination of Ki67 LI in the samples of NEN of the pancreas diagnosed in author's institution and gathered prospectively between November 2010 and May 2016. None of the samples was included in the previous study on staging of NEN [33]. Some cases were included in the study on grossing technique of pancreaticoduodenectomy specimens [34]. The inclusion criterion for the study was histopathological diagnosis of NEN of the pancreas, made in resection specimen or in incisional biopsy of primary tumour. NEN diagnosed in fine needle aspirates or in core biopsies were not included – cytological specimens did not allow examination of distribution of Ki67 LI immunoreactivity, and core needle samples contained usually not enough tumoral cells for the purpose of the study. The exclusion criteria were: (1) material not stained for Ki67, (2) slides not available for re-review.

Histopathological diagnoses

Each case was diagnosed as well-differentiated neuroendocrine tumour (NET) or poorly differentiated neuroendocrine carcinoma (NEC), based on histopathological criteria [35, 36, 37], but importantly without taking into account results of mitotic count and Ki67 LI. Details on tissue processing procedure and diagnostics are provided in SI – Methods.

Ki67 immunostain

There are no guidelines describing how to select optimal tissue block for examination of Ki67 LI in pancreatic NEN. In this study, tissue blocks were selected based on adequate amount of tissue in block, inclusion of tumour areas with heterogeneity in architecture or cytology, as well as areas with detectable mitotic figures based on slide review [4]. A single slide from each case stained for Ki67 was used in this study. Details on Ki67 staining procedure and slide digitization are described in SI – Methods.

Ki67 LI scoring

Three "hotspots" (HS) were identified in each Ki67 section. HS were defined as areas with the highest nuclear labeling, in concordance with WHO 2010 [1], ENETS 2006 [4], and National Comprehensive Cancer Network 2016 guidelines [30]. This was based on careful but subjective examination of digitized slides. Additionally, in each case area with lowest Ki67 expression was found ("cold spot", CS). HS and CS were captured and printed in color on A3 size paper. Prints were coded and randomly used for manual counting of Ki67 LI. Counting was performed by manual marking of tumoral cells with and without Ki67 expression with color pens [21, 38]. Of note, this approach is a reference technique of Ki67 LI assessment in pancreatic NEN [21]. Cells were considered Ki67-positive if they showed immunoreactivity in nucleus or in a mitotic figure, irrespective of the stain intensity [22, 39], although it was recognized that some investigators ignored slightly stained nuclei for Ki67 LI assessment in NEN [21, 40]. In each HS and CS, 2500 cells were counted in 25 por-

tions of 100 cells each. In the first step, small area with subjectively maximal density of Ki67 positive tumoral cells in every HS was found. Then counts were extended peripherally, in every step aiming (subjectively) to maximize number of Ki67 positive cells. For CS, areas with subjectively minimal density of Ki67 positive cells were found. Then counting was performed peripherally aiming to preserve Ki67 LI as low as possible. HS with highest Ki67 index as measured in 2000 cells was named "A" hotspot (HS-A), 2 other HS were coded as "B" and "C" in random order. Importantly, non-neoplastic cells (lymphocytes, stromal cells, endothelia) were not counted [26]. During the process of counting, the investigator was blinded to results obtained previously in other HS or in CS in particular case, and also to clinico-pathological data and Ki67 score in original pathology report. Ki67 LI in pathology report and mitotic counts were not taken into account for grading purposes in this study.

Statistical analysis

Ki67 LI was defined as a percentage of Ki67-positive tumoral cells among all examined tumoral cells [39]. Ki67 counts were analyzed both as continuous and ordinal (Ki67-LI-based grade) variables. Cases with Ki67 LI below 3% were recorded as G1, cases with Ki67 LI between 3% and 20% were coded as G2, and cases with Ki67 LI above 20% were coded as G3 [5, 30]. Details on statistical analysis are described in SI – Methods.

Working hypotheses

See SI – Working hypotheses.

Ethics and guidelines

Institutional Review Board agreed to perform this non-interventional study without full review necessary for experimental studies. Histopathology-adjusted REMARK guidelines were followed [41].

Results

Seventy-seven samples were included in the study (71 NET and 6 NEC). Clinico-pathological data of the study cases are described in SI – Table S2. Flowchart is available as SI – Fig. S1.

Ki67 LI decrease with number of examined of cells in hot spots

As expected, Ki67 LI values in NET were lower than in NEC (Table I). Both NET and NEC showed heterogeneity of Ki67 LI within HS. Dilution effect [42] was evident, as mean Ki67 LI values decreased with the number of counted cells. Figure 1A and B display geometric means of Ki67 LI in relationship to the number of counted cells in NET and in NEC, respectively. Dilution effect of Ki67 LI in NET and NEC was seen in HS-A, HS-B, and HS-C.

Concordance rates between Ki67 LI obtained when counting different number of cells in HS-A and between HS and CS in NET are described in Table II. Paired *t*-tests showed statistical significance, indicating bias between Ki67 LI values obtained in NET in less than 2000 cells (100, 500, 1000 cells) in HS-A or in suboptimal tumour areas (HS-B, HS-C, CS) vs. Ki67 LI in 2000 cells in HS-A. There was just moderate agreement between Ki67 scores obtained in 500 cells vs. 2000 cells in NET.

Similar calculations were performed for NEC (SI – Table S3). As number of NEC samples in this study was small, some statistical analyses were probably unpowered. Paired *t*-tests did not show bias between Ki67 LI obtained by counting 100 or 500 cells vs. 2000 cells in HS-A. There was substantial agreement between Ki67 scores obtained in 500 cells vs. 2000

Table I. Summary statistics of Ki67 LI in neuroendocrine tumours and in neuroendocrine carcinomas of the pancreas[1]

	Neuroendocrine tumours	Neuroendocrine carcinomas
Hot spot A		
Ki67 LI in 100 cells	9.8 (1-50)	88.3 (75-97)
Ki67 LI in 500 cells	5.6 (0.6-50)	84.0 (68.4-94.4)
Ki67 LI in 2000 cells	3.9 (0.4-43.4)	82.3 (64.2-93.8)
Hot spot B		
Ki67 LI in 2000 cells	2.7 (0.3-42.2)	66.4 (21.6-92.4)
Hot spot C		
Ki67 LI in 2000 cells	2.5 (0.2-37.6)	76.6 (58.8-90.9)
Cold spot		
Ki67 LI in 2000 cells	0.6 (0-6.6)	61.4 (34.2-88.8)

[1] Values reported as geometric means and ranges, in percentages
LI – labeling index

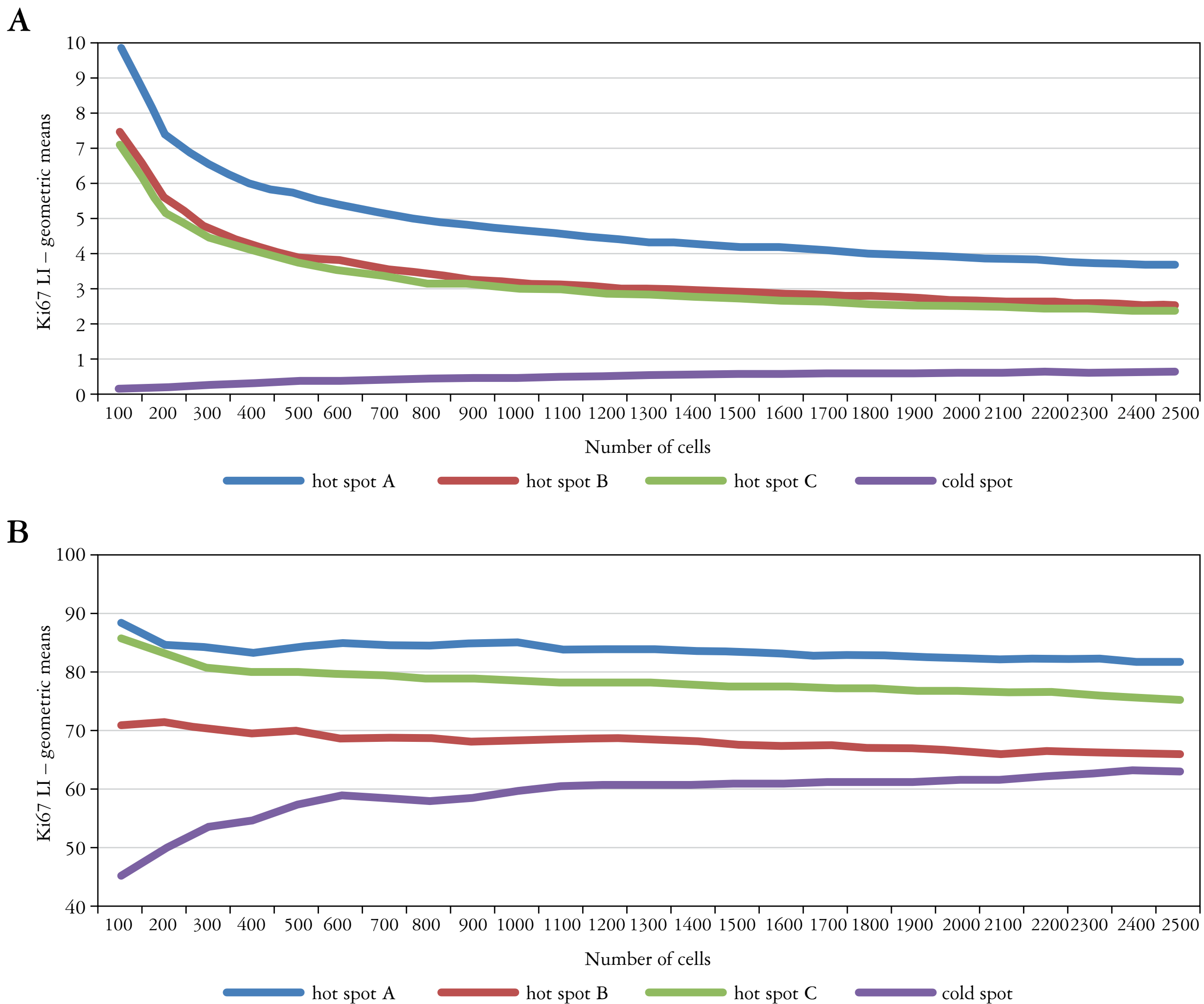

Fig. 1. Geometric means of Ki67 LI in relationship to the number of counted cells in neuroendocrine tumours (A) and in neuroendocrine carcinomas (B)

cells in NEC. However, concordance rates between values obtained in HS-B or HS-C vs. HS-A in NEC were not perfect. Figure S2A, B (SI) display the relationship between raw Ki67 LI scores in 500 cells vs. 2000 cells in HS-A in NET and in NEC, respectively.

Inspection of Bland-Altman plot suggested both proportional and fixed bias between Ki67 LI obtained in 500 vs. 2000 cells in NET in HS-A (SI – Fig. S3A). The mean difference between Ki67 LI in 2000 and in 500 cells (raw data) was –1.94. Geometric mean of the ratios was 0.696, indicating that Ki67 LI in 2000 cells in HS-A were on average 30% lower than Ki67 LI in 500 cells in HS-A. In contrast, fixed but not proportional bias was seen between Ki67 LI obtained in 2000 cells in HS-A vs. 2000 cells in HS-B, and vs. 2000 cells in HS-C (SI – Fig. S3B, C).

Grade proportions among NET are related to the number of examined cells in hot spot

The relationship between Ki67-LI-based grade proportions and number of examined cells in HS and CS in the entire study population is presented in SI – Fig. S4. Figure 2 and SI – Fig. S5A summarize data for NET in HS-A. As expected, proportion of grades in tumour samples changed with the number of counted cells. Only 2/71 (2.8%) NET cases were scored as G1 tumours based on counting of 100 cells in HS-A, but proportion of G1 cases increased to 19/71 (26.8%) and 30/71 (42.2%) when 500 and 2000 cells were counted, respectively. Among 45 NET cases scored as G2 in 500 cells (HS-A), 11 (24.4%) were moved to G1 category when 2000 cells were counted. Among 7 NET cases scored as G3 in 500 cells (HS-A), two were moved to G2 category when 2000 cells were counted. Of note, no case diagnosed as G3 when counting 100 cells was moved to G1 category when 2000 cells in HS-A were evaluated.

The concordance between grade proportions in relation to the number of examined cells is summarized in Table III. McNemar's tests showed systematic bias between grade proportions based on counting of

Table II. Concordance between Ki67 LI in hot spots and in cold spots in neuroendocrine tumours[1]

		HS-A (100 cells)	HS-A (500 cells)	HS-A (1000 cells)	HS-A (2500 cells)	HS-B (2000 cells)	HS-C (2000 cells)	CS (2000 cells)	HS-(A+B+C) – 500 cells in each (1500 cells in total)
HS-A (2000 cells)	*t*-test (*p*-value)	0.000[2]	0.000[2]	0.000[2]	0.000[2]	0.031[3]	0.011[3]	0.000[3]	0.000[2]
	Pearson's *R*	0.925 (p = 0.000)	0.978 (p = 0.000)	0.990 (p = 0.000)	0.998 (p = 0.000)	0.935 (p = 0.000)	0.933 (p = 0.000)	0.457 (p = 0.000)	0.979 (p = 0.000)
	Lin's CCC	0.576 (95% CI: 0.48-0.66)*	0.900 (95% CI: 0.86-0.93)**	0.973 (95% CI: 0.96-0.98)***	0.997 (95% CI: 0.99-1)****	0.875 (95% CI: 0.82-0.92)*	0.852 (95% CI: 0.79-0.90)*	0.161 (95% CI: 0.08-0.25)*	0.960 (95% CI: 0.94-0.97)***

[1] Calculations made on transformed data, [2] paired t-test, [3] unpaired t-test
CCC – concordance correlation coefficient; CI – confidence interval; CS – cold spot; HS – hot spot; LI – labeling index
** poor agreement; ** moderate agreement; *** substantial agreement; **** almost perfect agreement*

2000 cells vs. grade proportions based on counting of 100, 500 or 1000 cells. There was substantial but not perfect agreement between grades developed based on 500 and 2000 cells. Almost one fifth (13/71; 18.3%) of NET were attributed to different grade if 500 cells rather than 2000 cells were counted. Moreover, concordance between grades established in 100 cells vs. 2000 cells was unsatisfactory.

Some NET may be under-graded if suboptimal hot spots are selected for KI67 LI scoring

There was substantial but not perfect agreement between grade categories established in alternative HS (HS-B or HS-C) vs. HS-A. Again, approximately one fifth of NET were under-graded if HS-B or HS-C were selected for counting (percentages of disagreement for HS-B and HS-C vs. HS-A were 18.3% and 22.5%, respectively, details in SI – Table S4). According to some guidelines [11], counting cells in just a single HS is inadequate. In this study another method of Ki67 scoring was tested: Ki67 LI was calculated in 1500 cells – results in 500 cells were taken from each of 3 HS. Application of this approach resulted in applying of G2 category to 8/30 (26.7%) samples which were scored as G1 based on 2000 cells in HS-A. Details on concordance between grade proportion across HS and CS in NET are presented in SI – Table S4 and Fig. S5B–D.

Counting cells in cold spots very often leads to under-grading of NET but not NEC

In every NET areas of relatively lower proliferation (i.e. CS) were found (SI – Fig. S4). Scoring of CS resulted in very frequent under-grading of G2 and G3 tumours (SI – Table S4). Almost a half (35/71) of NET cases showed different grade when Ki67 LI examined in CS rather than in HS-A. Among 36 G2 NET cases (as scored in HS-A in 2000 cells), 30 (83.3%) were recognized as G1 based on CS examination (2000 cells). All 5 NET G3 cases were under-graded when CS was examined. Risk of under-grading of G2/G3 NET did not decrease with increasing number of examined cells – counting as many as 2000 cells within CS did not prevent under-grading (Fig. 1A and SI – Figs. S4 and S5D).

In contrary to NET, grade shift was not seen among NEC cases: all cases were diagnosed as G3 irrespective of number of examined cells in HS-A, HS-B, and HS-C (SI – Fig. S4). Similar results were obtained for counting in CS – all cases but one (5/6) were scored as G3 when 100 cells were counted. The only exception was a single case which showed lower Ki67 expression – but this could be caused by suboptimal fixation in a portion of cancer. Nevertheless, counting of just 100 cells was enough to grade NEC.

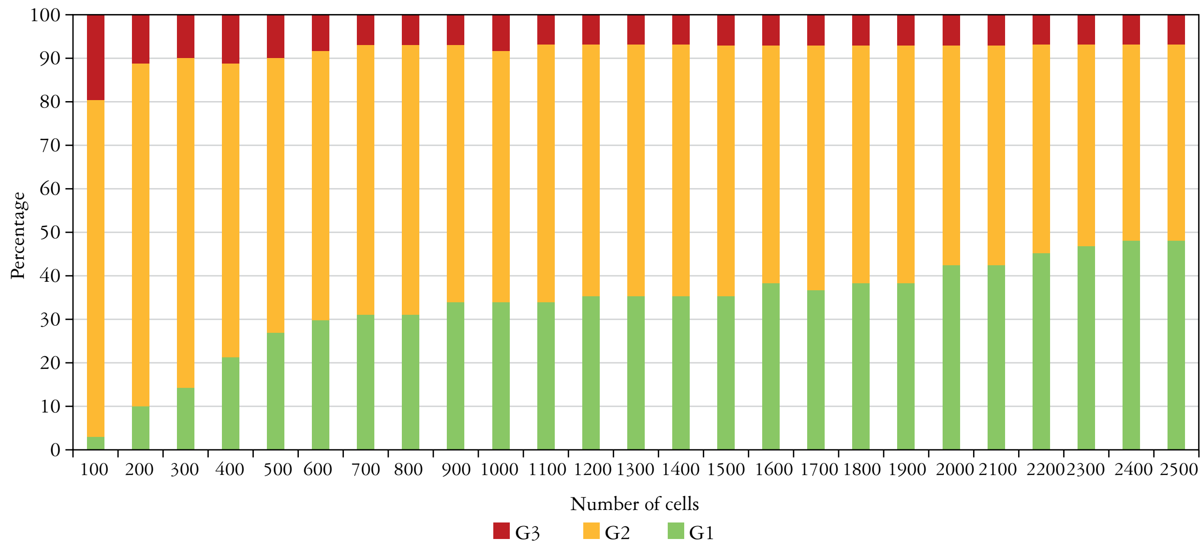

Fig. 2. The relationship between Ki67-LI-based grade proportions and number of examined cells in hot spot A in neuroendocrine tumours

Table III. Concordance of Ki67-LI-based grade in neuroendocrine tumours in hot spot A related to the number of counted cells

		HS-A (100 cells)			HS-A (500 cells)			HS-A (1000 cells)			HS-A (2500 cells)		
		G1	G2	G3	G1	G2	G3	G1	G2	G3	G1	G2	G3
HS-A (2000 cells)	G1	2	28	0	19	11	0	24	6	0	30	0	0
	G2	0	27	9	0	34	2	0	35	1	4	32	0
	G3	0	0	5	0	0	5	0	0	5	0	0	5
McNemar's test		p = 0.000			p = 0.001			p = 0.023			p = 0.134		
Percentage of agreement		34/71 (47.9%)			58/71 (81.7%)			64/71 (90.1%)			67/71 (94.4%)		
Weighted κ		κ = 0.22 (95% CI: 0.08-0.35)* (p = 0.000)			κ = 0.70 (95% CI 0.56-0.85)** (p = 0.000)			κ = 0.84 (95% CI: 0.73-0.95)*** (p = 0.000)			κ = 0.91 (95% CI: 0.82-1.00)*** (p = 0.000)		
Weighted κ significantly above 0.6		NS			p = 0.081			p = 0.000			p = 0.000		
Spearman's rho		R = 0.557 (p = 0.000)			R = 0.769 (p = 0.000)			R = 0.866 (p = 0.000)			R = 0.914 (p = 0.000)		

CI – confidence interval; HS – hot spot; LI – labeling index; NS – not significant
** fair agreement; ** substantial agreement; *** almost perfect agreement*

Some NET are scored as G2 in 500 cells but as G1 in 2000 cells within hot spot

Eleven NET showed G2 grade when 500 cells in HS-A were counted, but G1 when 2000 cells were counted (named here G1.5 subgroup). The statistical analyses aiming to compare clinico-pathological profiles of G1 vs. G1.5 vs. G2 subgroups were surely unpowered, but some tentative observations were made. Clinico-pathological profile of G1.5 tumours resembled G1 rather than G2 tumours. In particular, G2 tumours tended to be larger and more often diagnosed at higher stage than G1.5 tumours, and contained necrotic foci in a subset of cases, in contrary to G1.5 samples. These observations were not confirmed with statistical significance, however (details in SI – Methods and Table S5).

Ki67 LI is imperfect predictor of regional lymph node metastasis

In selected patients, NET may be treated with limited resection or even managed conservatively, for example due to co-morbidities or presumable low risk of progressive disease [6]. Tumour grade was previously considered as a supplementary tool to

predict presence of regional lymph node metastases [43, 44, 45].

In this study, G2/G3 vs. G1 grade showed at best moderate ability to detect pN1 stage. For example, positive and negative likelihood ratios of G2/G3 for detection pN1 (as scored in 2000 cells in HS-A) were 1.716 and 0.418, respectively (SI – Table S6). Ki67 LI examined as a continuous variable was also not fully accurate for pN1 detection, irrespective of number of scored cells. For example, areas under receiver-operatic characteristics graphs for 500 and 2000 cells (in HS-A) were 0.751 and 0.771, respectively (this indicated fair accuracy, SI – Table S7).

Five-percent Ki67 LI cut-off value for G1/G2 distinction do not improve grading precision

Some investigators proposed 5% Ki67 LI as optimal cut-off value for distinguishing G1 and G2 NET [46, 47, 48]. Such modification resulted previously in statistically significant differences in survival between G1 and G2 cases in multivariate analyses adjusted for stage, in contrary to original 2(3)% cut-off [47, 48]. This result was confirmed by meta-analysis [49], which showed better ability of 5% value for prognostication in NET.

In this study, application of 5% cut-off value did not improve agreement between scores based on different number of cells in HS nor scores based on different HS nor Ki67 LI ability to predict lymph node metastasis (SI – Tables S8, S9, and S10).

Discussion

There were several findings of this study: (1) Pancreatic NEN showed Ki67 LI heterogeneity within primary tumour tissue. (2) Ki67 LI heterogeneity in NET resulted in imperfect precision of Ki67 scoring for grading purposes, in case the latter was performed in adherence to WHO 2010 guidelines. (3) In many pancreatic NET, number of cells which are to be counted in HS was critical for precise grading. (4) In almost one fifth of pancreatic NET, counting of just 500 cells in HS (as allowed in WHO 2010 guidelines) resulted in different (i.e. higher) grade than counting of 2000 cells (as recommended by ENETS 2006 guidelines). (5) Counting of just 100 cells (as described in ENETS 2009 guidelines) resulted with much higher number of G2 and G3 cases of pancreatic NET than counting of 500 or 2000 cells. (6) In approximately one fifth of NET, suboptimal choice of HS caused under-grading of tumour. (7) Areas of lower proliferation rate (CS) were constantly present in pancreatic NET, so counting of random cells (as in cytological preparations) may result in under-grading. (8) NET which showed G2 grade when 500 cells were counted, but G1 when 2000 cells were counted in HS had clinico-pathological profile resembling G1 rather than G2 tumours. (9) Ki67 LI and Ki67-LI-based grade were imperfect predictors of regional lymph node metastasis in pancreatic NET. (10) Utilization of 5% Ki67 LI as a cut-off value for G1/G2 distinction did not improve precision of grading of pancreatic NET. (11) All NEC cases were diagnosed as G3 irrespective of number of examined cells and HS selected for counting. (12) In adequately preserved samples of NEC, Ki67 LI in CS was high enough to establish G3 grade. (13) WHO and ENETS scoring criteria were not sufficiently detailed to ensure reliable reporting of grade in a subset of pancreatic NEN.

Analytical validation

Validation of a (bio)marker assay for clinical use requires evaluation of its analytical validity, clinical validity, and clinical utility [18, 50, 51, 52]. Every laboratory test must show acceptable level of analytical validity to have any clinical utility [50, 53]. Many parameters are used for analytical validation of biomarker assays, including accuracy, trueness, precision, reproducibility, repeatability, linearity, and robustness [52, 54]. Analytical validation of IHC tests has gained attention recently [55]. To analytically validate IHC test, it is necessary to examine its accuracy, analytical sensitivity, analytical specificity, and concordance with a gold standard, as well as its robustness and precision [55]. As there is no reference gold standard test for almost all IHC assays [55, 56], therefore documentation of accuracy of IHC test may be problematic. Validation of IHC assay needs to deal with documentation of pre-analytical, analytical, and post-analytical phases of the test [39, 55, 57, 58].

Pre-analytical and analytical phase of KI67 LI assessment

Ki67 LI requires analytical validation as other IHC stains [39, 55]. Ki67 LI need be to reproducible and consistent if it is to be a clinically useful test [39]. This is critically important issue, as Ki67 LI has some features of "class II" IHC assay. This indicates that it may potentially serve as an independent test, disconnected operationally from other parts of pathology report [55]: it may be used not only as a prognostic marker, but also for guiding therapy. There is no gold standard test available in surgical pathology laboratory to which Ki67 LI could be compared. Probably the "true" KI67 LI is that one, which has the strongest prognostic or predictive value [59].

Majority of data on validation of Ki67 LI came from studies on breast cancer. In breast cancer, Ki67 scores are influenced by many pre-analytical and analytical factors, for example type of fixative, delay of fixation, prolonged fixation, cutting of the fresh tumour tissue aiming to improve fixation, storage of

cut sections, methods of antigen retrieval, antibody clone, and counterstaining [39, 57, 60]. Significant inter-laboratory variability of Ki67 LI reporting was seen in breast cancer [53], but also in NEN [22].

Post-analytical phase of Ki67 LI assessment

Similarly important are post-analytical (interpretative) aspects of Ki67 LI. Significant efforts were made to improve reliability of Ki67 reporting in breast cancer [39, 53, 61]. In their early multi-institutional study on Ki67 reproducibility in breast cancer, Polley *et al.* showed that analytical validity of Ki67 LI was "unacceptably poor" [53]. Calibration using web-based tool allowed to increase concordance in reporting of Ki67 LI [61]. Despite these efforts, it is probably still too early to recommend usage of Ki67 LI for management of patients with breast cancer [53, 61, 62]. Reproducibility of Ki67 scoring in breast cancer seems to be not enough to ensure consistent reporting [51, 53, 62, 63]. Making therapeutic decisions for patients with breast cancer based on Ki67 scores should be made in caution [51, 53, 62, 63]. Interpretation the KI67 LI obtained in particular breast cancer sample should always be done in the context of local laboratory values [64]. Detailed guidelines for Ki67 LI assessment in breast cancer (including selection of the specimen, its fixation, technical aspects of IHC procedure, methodology of stain interpretation, reporting, and statistical analysis for clinical trials) were published [39]. Similar initiatives aiming to document and if necessary to improve reproducibility and precision of Ki67 LI in NEN are lacking.

Significance of Ki67 LI in pancreatic NEN

Firstly, Ki67 LI is used for diagnosis of gastro-entero-pancreatic NEN [1]. Together with assessment of mitotic index, Ki67 LI is used for grading purposes, i.e. distinction G1, G2, and G3 neoplasms. What is more important, it is critical for establishing the diagnosis of pancreatic NEC, which in current WHO guidelines is based on assessment of proliferation rate in NEN and independent largely from consideration of morphological picture. As clinical management of NET and NEC is different [6, 65], to distinguish between these entities is the most important aspect in diagnostics of neuroendocrine neoplasms [11].

Secondly, Ki67 LI is a prognostic factor in NEN (as reviewed in [49, 66]). Ki67 LI is a risk factor of progression in advanced pancreatic NEN [67]. Its prognostic value in pancreatic NET is additive to prognostic value of mitotic index [26]. However, interpretation of studies on prognostic value of Ki67 LI in NEN is difficult, as they frequently include patients enrolled in long time periods, who were treated with different methods, and diagnosed with NEN with subsequent assessment of Ki67 LI in different stages of their disease. Some older studies did not distinguish NET and NEC. Moreover, different methods of scoring and different cut-off values, problems of multivariable models (for example non-uniform across studies adjustment for clinical data, including stage; lack of adjustment of important histopathological variables, like perineural invasion or lymph-vascular invasion; overfitting of the models) hamper considerably comprehensive interpretation of available data. It is not clear if distinction between G1 and G2 based on WHO 2(3)% cut-off value is prognostically significant in patients with pancreatic NET if tumour stage of the disease in taken into account. Such an observation was reported in a single study [68]. In some studies stage-independent prognostic value of Ki67 LI was proven only when cut-off value of 4.8% [46] or 5% [47, 48] was applied, instead of 2(3)%, or alternatively, when study population included also gastric and duodenal neoplasms [69].

Thirdly, Ki67 LI may guide therapy in NEN. Ki67 LI may be used for selection of particular types of treatment and for surveillance programmes [6, 7], for example for a choice – surgery vs. surveillance for small non-functioning NET [6], treatment selection in advanced disease [7, 70], or length of intervals between assessments during follow-up [6]. Ki67 LI may also serve as a potential predictor of response to chemotherapy [71], radioembolization or chemoembolization [72] in non-operable NEN.

KI67 LI scoring criteria in gastro-entero-pancreatic NEN

Ki67 LI scoring criteria provided by authoritative sources differ in some aspects (SI – Table S1). 2016 guidelines by National Comprehensive Cancer Network [30] still allow eyeballed estimation of KI67 LI. In contrast, guidelines by WHO 2010 [1] and by ENETS 2006 [4] suggests that formal counting of cells should be performed. Virtually all these sources suggest that Ki67 should be performed in HS [1, 4, 11, 29, 30, 31]. As mentioned above, number of cells which should examined for Ki67 LI differ between some guidelines.

Hot spots

Important aspect of Ki67 scoring is selection of tumour regions for counting [42, 53, 73, 74, 75, 76, 77, 78]. Current counting methodology in NEN is based on recognition of HS. This comes from assumption that results obtained in HS rather than in CS or in "average" areas are critical for prognosis [59]. HS is somewhat poorly defined term [74, 78]. It is usually defined as a high-proliferation-area [78]. HS may differ in size, shape and gradient of stain from HS to surrounding areas [42, 74, 78]. Density

of cells in particular tumor area may also influence selection of this area as HS [75]. HS selection is done subjectively and results of these choices (at least in breast cancer) do not necessarily correlated with HS recognized using digital image analysis [73, 74]. Automated HS detection may be an alternative [74, 76, 77, 79, 80]. It is not fully clear whether single HS is enough for Ki67 LI assessment or whether several HS should be counted [81]. It is also not known if separate Ki67 LI should be simultaneously reported for primary tumours and their regional or distant metastases [2, 81].

One of the finding of this study was observation that HS selection in many NET was important for reliable grading. Selection of "suboptimal" HS resulted in under-grading of about 20% NET. In contrast, all NEC in this study were scored as G3 neoplasms, irrespective of selected HS. As mentioned, HS may differ in size and shape [74, 78]. Problem of HS configuration was experienced by the investigator during progress of counting more and more cells in particular HS. Decisions on choice of HS were made subjectively, so errors were unavoidable.

Number of cells for counting Ki67 LI

It is not fully clear how many cells and in how many HS should be counted for grading purposes. Counting more cells is believed to be more reliable than counting less – however this may not necessarily be true, as examined cells (sample) are selected in non-random way from non-homogenous (in statistical sense) population of tumoral cells [42]. In fact, the relationship between number of cells examined and precision of Ki67 scores is not fully described [75]. Christgen *et al.* in their study on breast cancer showed that values of Ki67 LI were related to number of examined cells, as counting more cells was associated with extension from areas of high proliferation to low-proliferation periphery [78]. This resulted in change of proportion of breast cancers with low proliferation (defined as below 20%) from 2% to 56% (!) if number of examined cells increased from 50 to 10000 [78]. Observations on the dilution effect in Ki67 scoring in breast cancer were also made by Romero *et al.* [42].

In contrast, the relationship between Ki67 LI and the number of examined cells in histopathological sections of NEN is rather unexplored. Similarly to reports of Christgen *et al.* and Romero *et al.* in breast cancer, one finding of this study was that number of examined cells within HS was critically important for grading of pancreatic NET. Dilution effect was clearly seen in many NET. Counting of just 100 cells (as allowed in 2009 ENETS guidelines) caused almost total disappearance of G1 NET. Moreover, many cases would had been classified as NET G3, just recently recognized entity [82]. Counting of 500 cells in a subset of cases was not equal to counting of 2000 cells. Prevalence of NET G3 among NET differed if 500 rather than 2000 cells were counted for Ki-67 LI. Some NET were graded as G2 and G3 based on counting of 2000 and 500 cells, respectively. Eleven NET in this study were graded as G2 in 500 cells, but as G1 in 2000 cells. Their pathological profile was similar to G1 tumours, but this analysis was unpowered. In contrast to NET, NEC examined in this study were scored as G3, irrespective of number of examined cells in HS.

Ki67 LI heterogeneity in the context of CS

Approximately 90% (64/71) NET examined in this study would had been graded as G1 if CS areas were chosen for sampling. Similar observations were made by Hasegawa *et al.*, who showed that majority of volume of G2 tumours was in fact composed of G1 areas [12].

Ki67 LI heterogeneity is very important for reliability of grading when a limited portion of tissue is available for examination. Li *et al.* in their meta-analysis showed that Ki67 LI assessment based on cytological samples had a pooled sensitivity of 64% for detection of G2/G3 vs. G1 tumours [19]. Carlinfante *et al.* showed that Ki67 LI in cytological sample had poor sensitivity for detection of G2 NET, but allowed perfect distinction of NET and NEC [83]. Change of grade of NET is not rare at the stage of synchronous and especially metachronous metastatic disease [15]. Ki67 LI heterogeneity of different extent may be also found in some metastases of pancreatic NET [16, 17].

Results of this one and previous studies indicate that recognition of G1 in a limited "random" (not HS) sample of tumoral cells gives no guarantee that some G2 areas are not missed for scoring. Counting of 2000 cells in CS did not prevent under-grading in this study. It is possible that relying on reports based on random sample of tumoral cells may result in under-treatment of some patients with pancreatic NET.

Ki67 LI heterogeneity beyond single HS or CS

Heterogeneity of Ki67 LI is well recognized and best described in breast cancer [74, 78, 84]. This is critical for making clinical decisions, as assessment of non-representative or suboptimal areas for Ki67 LI may result in inadequate selection of chemotherapy protocols in breast cancer [84].

Comparatively, little attention has been paid to heterogeneity of Ki67 LI in pancreatic NEN. As mentioned above, Hasegawa *et al.* documented that NET G2 consist in majority from G1 areas [12]. Goodell *et al.* distinguished 3 groups of pancreatic NET based on Ki67 LI in a single-field HS and 10 consecutive fields average: 13/45 (28.9%) and 6/45 (13.3%) were

scored as G1 and G2 based on both scoring techniques (respectively), but as many as 26/45 (57.8%) were scored as G2 based on a single HS, but as G1 based on 10 consecutive fields average [24]. Grillo *et al*. reported that 37% (22/60) of gastro-entero-pancreatic NEN showed different Ki67 LI between tissue blocks of primary tumour (2-5 paraffin block were examined per case). However, only in 5% (3/60) of cases this resulted in change of grade from G1 to G2 [15]. This value could be underestimated, since higher grade component may form a minor portion of the tumour [9]. Valous *et al.* showed that lacunarity measurement allows categorization of samples of pancreatic NEN to groups with homogenous or with heterogenous spatial Ki-67 LI distribution across whole virtual sections [10]. Interestingly, group of neoplasm with homogenous (uniform) proliferation in that study was enriched in G3 samples (6/6) and G1 samples (17/23) rather than in G2 samples (7/14) [10]. It is now clear that automated methods of comprehensive assessment of tissue stain distribution in the whole slides give opportunity to examine and report phenotypic tumour heterogeneity in a systematic fashion [74].

In this study, only 34/71 (47.9%) NET showed the same grade when 100, 500, and 2000 cells were counted in HS-A. However, the same grade in 100, 500, and in 2000 cells in all 3 HS was seen only in 14/71 (19.7%) NET (SI – Table S11). Some NET indeed showed homogenous distribution of Ki67-positive cells (at least in a single tumour section).

Cut-off value for G1/G2 distinction

The selection of cut-off values for Ki67 LI is also a controversial issue [42, 85]. Many cut-offs values may work virtually equally well for prognostic purposes in a particular cohort. This does not necessarily indicate that they would equally well work in other population or when examined by other investigators or with different methodology [39, 85, 86]. Moreover, cut-off points should depend on clinical context, for example for prognostic purposed or for prediction of response of particular therapy protocol [39, 42], or for monitoring disease progression. This indicates that "the best" one and only Ki67 LI cut-off point for a particular disease does not exist [42]. At this moment there is no data supporting thesis that G1 and G2 NET (irrespective of 2%, 3%, or 5% cut-off value for their distinction) are biologically different entities. For that reason Ki67 LI cut-off value for distinction of G1 vs. G2 among pancreatic NET selection in purely artificial.

In this study, consideration of 5% Ki67 LI as a cut-off value for G1/G2 distinction did not result in better precision of the scoring, in relation to agreement between HS, to number of cells counted in HS, and to risk of under-grading based on CS assessment.

Ki67 LI as a risk factor of regional lymph node metastasis

In some patients with NET enucleation or limited resection rather than formal partial pancreatectomy with regional lymphadenectomy is performed. This gives lower morbidity but is associated with a risk of metastasis in regional lymph nodes which are not resected (as discussed in [43, 44, 45]). Risk of LN metastasis in G2 tumours is higher than in G1 [43, 44, 45], but LN metastasis is not rare in G1 tumours (7.5% [45] – 23.1% [43]).

In this study Ki67 LI was imperfect predictor of regional lymph node metastasis in pancreatic NET. This was not related to number of examined cells and area selected for counting.

Human factor

Human factor is critical for Ki67 LI assessment, not only for eyeballed estimates, but also for formal manual counting [39, 56, 86]. Important issue for manual counting is HS recognition, as discussed above. Another critical issue of Ki67 LI is a subjective threshold of stain intensity, which is enough to score a particular nucleus as positive [53, 73]. Interestingly, inter-observer variability in scoring breast cancer samples cannot be fully explained by technique of scoring (estimation vs. counting), HS selection, positivity threshold or IHC technique [56]. Related issue is psychological aspect of manual counting of Ki67 LI. According to some pathologists, counting cells for reporting Ki-67 LI is "waste of time" (as reported by Vörös *et al.* [87]). Counting of cells truly can be a very frustrating or boring task [88].

Toward optimal method of Ki67 LI assessment

It is not clear if counting cells (either manual or performed with digital image analysis) is truly "better" than eyeballed estimation of Ki67 LI. In breast cancer, some researchers reported that estimation may be adequate or at least not worse than counting [56, 87]. However, multi-institutional studies showed that formal counting was more reproducible than estimation in breast cancer [53, 75]. Similar observation were made in NEN: results of eyeballed estimation differed from results of counting [20, 21, 25] and were irreproducible [25], but some investigators showed that estimation was not as bad [27, 28]. Pathologists in general overestimate the Ki67 LI during eye-balled estimation in comparison to counting technique [25, 27].

As discussed by Varga *et al.* [56], it is general belief that formal counting is more reliable than estimation. However, there are many purely subjective steps even during formal counting cells for Ki67 LI: selection of tumour block (blocks?) for staining, selection of one or more HS on the slide, assessment of size of HS, se-

lection of area within HS to be captured for counting, number of cells to be counted (100-2000 or more, in 1 or more HS), or selection of direction of counting during incremental "moving" to areas of lower proliferation. All these steps may dramatically influence the final Ki67 score. The scoring procedure is as reliable as precise is the "weakest", most subjective step of the process. In this study it was found that Ki67 LI scores may significantly differ despite application of WHO scoring criteria. So reporting of exact Ki67 LI value obtained in limited number of cells seems to be falsely precise exercise.

This examination and previous studies [10, 12, 15, 24] showed that pancreatic NEN (in particular NET) may show heterogeneous distribution of Ki67 LI. Unfortunately, adherence to WHO 2010 scoring criteria did not allow providing precise reports of tumour grade in a subset of pancreatic NET, at least from a single patient perspective. Concept and methodology of Ki67 LI *counting* in gastro-entero-pancreatic assessment lacks full clarity, so standardization is difficult to achieve. It is clear that better operating definition of HS is necessary. Examination of just a single HS is believed to inadequate [11], however reasons of this statement are unknown (Does Ki67 LI in a single HS overestimate "true" score? Is examination of more than 1 HS is a correction factor for suboptimal choice of HS?). Number of cells which should be examined within HS also needs clarification. Unfortunately, it is possible that optimal number of cells for counting does not exist, similarly to optimal cut-off value for G1/G2 distinction. Application of automated stain analysis will not solve the problem of limited precision of grading if other aspects of Ki67 LI assessment (like pre-analytical and analytical factors, or selection of portion of tumour for counting) will still be non-standardized/subjective.

Limitations of the study

See SI – Study Limitations.

Conclusions

Counting of Ki67 LI and establishing of Ki67-LI-based grade performed in adherence to WHO 2010 guidelines index had suboptimal precision in relation to the number of evaluated cells in NEN of the pancreas. Up to 20% of NET showed higher grade when 500 cells rather than 2000 cells were counted in HS area. Similarly, suboptimal choice of HS for assessing Ki67 LI resulted in under-grading of approximately 20% of NET. Low-proliferation areas were present in virtually every NET, so counting cells for Ki67 LI without consideration of hot spots may result in under-grading. Heterogeneity of Ki67 LI was also present in pancreatic NEC, but it virtually never resulted in under-grading. Ki67 LI in many pancreatic NEN were heterogeneous in relation to area selected for counting and number of examined cells. Concept and methodology of Ki67 LI *counting* in gastro-entero-pancreatic NEN requires clarification. Systematic effort aiming to document and if necessary to improve precision of assessment of Ki67 LI in NEN of pancreas is needed.

The study was partially supported by internal grant of Medical University of Silesia (KNW-1-119/N/5/0).

The author declares no conflict of interest.

References

1. Rindi G, Arnold R, Bosman FT, et al. Nomenclature and classification of neuroendocrine neoplasms of the digestive system. In: WHO classification of tumours of the digestive system. Bosman FT, Carneiro F, Hruban RH, Theise ND (eds). 4th ed. International Agency for Research on Cancer, Lyon 2010; 13-14.
2. Adsay V. Ki67 labeling index in neuroendocrine tumors of the gastrointestinal and pancreatobiliary tract: to count or not to count is not the question, but rather how to count. Am J Surg Pathol 2012; 36: 1743-1746.
3. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. J Cell Physiol 2000; 182: 311-322.
4. Rindi G, Klöppel G, Alhman H, et al. TNM staging of foregut (neuro)endocrine tumors: a consensus proposal including a grading system. Virchows Arch 2006; 449: 395-401.
5. Kunz PL, Reidy-Lagunes D, Anthony LB, et al. North American Neuroendocrine Tumor Society. Consensus guidelines for the management and treatment of neuroendocrine tumors. Pancreas 2013; 42: 557-577.
6. Falconi M, Eriksson B, Kaltsas G, et al. ENETS consensus guidelines update for the management of patients with functional pancreatic neuroendocrine tumors and non-functional pancreatic neuroendocrine tumors. Neuroendocrinology 2016; 103: 153-171.
7. Pavel M, O'Toole D, Costa F, et al. ENETS consensus guidelines update for the management of distant metastatic disease of intestinal, pancreatic, bronchial neuroendocrine neoplasms (NEN) and NEN of unknown primary site. Neuroendocrinology 2016; 103: 172-185.
8. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer 2012; 12: 323-334.
9. Tang LH, Untch BR, Reidy DL, et al. Well-differentiated neuroendocrine tumors with a morphologically apparent high-grade component: A pathway distinct from poorly differentiated neuroendocrine carcinomas. Clin Cancer Res 2016; 22: 1011-1017.
10. Valous NA, Lahrmann B, Halama N, et al. Spatial intratumoral heterogeneity of proliferation in immunohistochemical images of solid tumors. Med Phys 2016; 43: 2936.
11. Klimstra DS, Modlin IR, Adsay NV, et al. Pathology reporting of neuroendocrine tumors: application of the Delphic consensus process to the development of a minimum pathology data set. Am J Surg Pathol 2010; 34: 300-313.
12. 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.
13. Adesoye T, Daleo MA, Loeffler AG, et al. Discordance of histologic grade between primary and metastatic neuroendocrine carcinomas. Ann Surg Oncol 2015; 22 Suppl 3: S817-S821.

14. Zen Y, Heaton N. Elevated Ki-67 labeling index in 'synchronous liver metastases' of well differentiated enteropancreatic neuroendocrine tumor. Pathol Int 2013; 63: 532-538.
15. Grillo F, Albertelli M, Brisigotti MP, et al. Grade increases in gastroenteropancreatic neuroendocrine tumor metastases compared to the primary tumor. Neuroendocrinology 2016; 103: 452-459.
16. Couvelard A, Deschamps L, Ravaud P, et al. Heterogeneity of tumor prognostic markers: a reproducibility study applied to liver metastases of pancreatic endocrine tumors. Mod Pathol 2009; 22: 273-281.
17. Yang Z, Tang LH, Klimstra DS. Effect of tumor heterogeneity on the assessment of Ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification. Am J Surg Pathol 2011; 35: 853-860.
18. Hayes DF. Biomarker validation and testing. Mol Oncol 2015; 9: 960-966.
19. Li J, Lin JP, Shi LH, et al. How reliable is the Ki-67 cytological index in grading pancreatic neuroendocrine tumors? A meta-analysis. Dig Dis 2016; 17: 95-103.
20. Young HT, Carr NJ, Green B, et al. Accuracy of visual assessments of proliferation indices in gastroenteropancreatic neuroendocrine tumours. J Clin Pathol 2013; 66: 700-704.
21. 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.
22. Blank A, Wehweck L, Marinoni I, et al. Interlaboratory variability of MIB1 staining in well-differentiated pancreatic neuroendocrine tumors. Virchows Arch 2015; 467: 543-550.
23. Nadler A, Cukier M, Rowsell C, et al. Ki-67 is a reliable pathological grading marker for neuroendocrine tumors. Virchows Arch 2013; 462: 501-505.
24. Goodell PP, Krasinskas AM, Davison JM, et al. Comparison of methods for proliferative index analysis for grading pancreatic well-differentiated neuroendocrine tumors. Am J Clin Pathol 2012; 137: 5765-5782.
25. Tang LH, Gonen M, Hedvat C, et al. 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.
26. McCall CM, Shi C, Cornish TC, et al. Grading of well-differentiated pancreatic neuroendocrine tumors is improved by the inclusion of both Ki67 proliferative index and mitotic rate. Am J Surg Pathol 2013; 37: 1671-1677.
27. van Velthuysen ML, Groen EJ, Sanders J, et al. Reliability of proliferation assessment by Ki-67 expression in neuroendocrine neoplasms: eyeballing or image analysis? Neuroendocrinology 2014; 100: 288-292.
28. Kroneman TN, Voss JS, Lohse CM, et al. Comparison of three Ki-67 index quantification methods and clinical significance in pancreatic neuroendocrine tumors. Endocr Pathol 2015; 26: 255-262.
29. Klöppel G, Couvelard A, Perren A, et al. ENETS Consensus Guidelines for the Standards of Care in Neuroendocrine Tumors: towards a standardized approach to the diagnosis of gastroenteropancreatic neuroendocrine tumors and their prognostic stratification. Neuroendocrinology 2009; 90: 162-166.
30. National Comprehensive Cancer Network. NCCN Clinical Practice Guidelines in Oncology. Neuroendocrine Tumors. Version 2.2016. https://www.nccn.org/professionals/physician_gls/pdf/neuroendocrine.pdf (Assessed: 24th July 2016).
31. College of American Pathologists. Protocol for the Examination of Specimens From Patients With Tumors of the Endocrine Pancreas. Pancreas Endocrine 3.3.0.0. January 2016 Asessed: http://www.cap.org/ShowProperty?nodePath=/UCMCon/Contribution Folders/WebContent/pdf/cp-pancreasendo-16protocol-3300.pdf (30th July 2016).
32. 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.
33. Liszka Ł, Pająk J, Mrowiec S, et al. Discrepancies between two alternative staging systems (European Neuroendocrine Tumor Society 2006 and American Joint Committee on Cancer/Union for International Cancer Control 2010) of neuroendocrine neoplasms of the pancreas. A study of 50 cases. Pathol Res Pract 2011; 207: 220-224.
34. Liszka Ł, Mrowiec S, Kuśnierz K, et al. Standardized grossing protocol is useful for the pathology reporting of malignant neoplasms other than adenocarcinomas treated with pancreaticoduodenectomy. Histol Histopathol 2017; 32: 177-192.
35. Tang LH, Basturk O, Sue JJ, et al. A practical approach to the classification of WHO grade 3 (G3) well-differentiated neuroendocrine tumor (WD-NET) and poorly differentiated neuroendocrine carcinoma (PD-NEC) of the pancreas. Am J Surg Pathol 2016; 40: 1192-1202.
36. Basturk O, Tang L, Hruban RH, et al. Poorly differentiated neuroendocrine carcinomas of the pancreas: a clinicopathologic analysis of 44 cases. Am J Surg Pathol 2014; 38: 437-447.
37. Vélayoudom-Céphise FL, Duvillard P, Foucan L, et al. Are G3 ENETS neuroendocrine neoplasms heterogeneous? Endocr Relat Cancer 2013; 20: 649-657.
38. Basturk O, Yang Z, Tang LH, et al. The high-grade (WHO G3) pancreatic neuroendocrine tumor category is morphologically and biologically heterogenous and includes both well differentiated and poorly differentiated neoplasms. Am J Surg Pathol 2015; 39: 683-690.
39. Dowsett M, Nielsen TO, A'Hern R, et al. International Ki-67 in Breast Cancer Working Group. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst 2011; 103: 1656-1664.
40. Reid MD, Balci S, Saka B, et al. Neuroendocrine tumors of the pancreas: current concepts and controversies. Endocr Pathol 2014; 25: 65-79.
41. Knijn N, Nagtegaal ID. Guidelines for reporting histopathology studies. J Clin Pathol 2015; 68: 173-174.
42. Romero Q, Bendahl PO, Klintman M, et al. Ki67 proliferation in core biopsies vs. surgical samples - a model for neo-adjuvant breast cancer studies. BMC Cancer 2011; 11: 341.
43. Hashim YM, Trinkaus KM, Linehan DC, et al. Regional lymphadenectomy is indicated in the surgical treatment of pancreatic neuroendocrine tumors (PNETs). Ann Surg 2014; 259: 197-203.
44. Partelli S, Gaujoux S, Boninsegna L, et al. Pattern and clinical predictors of lymph node involvement in nonfunctioning pancreatic neuroendocrine tumors (NF-PanNETs). JAMA Surg 2013; 148: 932-939.
45. Jiang Y, Jin JB, Zhan Q, et al. Impact and Clinical Predictors of Lymph Node Metastases in Nonfunctional Pancreatic Neuroendocrine Tumors. Chin Med J (Engl) 2015; 128: 3335-3344.
46. Rindi G, Falconi M, Klersy C, et al. TNM staging of neoplasms of the endocrine pancreas: results from a large international cohort study. J Natl Cancer Inst 2012; 104: 764-777.
47. Scarpa A, Mantovani W, Capelli P, et al. Pancreatic endocrine tumors: improved TNM staging and histopathological grading permit a clinically efficient prognostic stratification of patients. Mod Pathol 2010; 23: 824-833.
48. Bettini R, Boninsegna L, Mantovani W, et al. Prognostic factors at diagnosis and value of WHO classification in a mono-institutional series of 180 non-functioning pancreatic endocrine tumours. Ann Oncol 2008; 19: 903-908.

49. Richards-Taylor S, Ewings SM, Jaynes E, et al. The assessment of Ki-67 as a prognostic marker in neuroendocrine tumours: a systematic review and meta-analysis. J Clin Pathol 2016; 69: 612-618.
50. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. Genet Med 2009; 11: 3-14.
51. Guiu S, Michiels S, André F, et al. Molecular subclasses of breast cancer: how do we define them? The IMPAKT 2012 Working Group Statement. Ann Oncol 2012; 23: 2997-3006.
52. Duffy MJ, Sturgeon CM, Sölétormos G, et al. Validation of new cancer biomarkers: a position statement from the European group on tumor markers. Clin Chem 2015; 61: 809-820.
53. Polley MY, Leung SC, McShane LM, et al. An international Ki67 reproducibility study. J Natl Cancer Inst 2013; 105: 1897-1906.
54. Jennings L, Van Deerlin VM, Gulley ML. Recommended principles and practices for validating clinical molecular pathology tests. Arch Pathol Lab Med 2009; 133: 743-755.
55. Fitzgibbons PL, Bradley LA, Fatheree LA, et al. Principles of analytic validation of immunohistochemical assays: Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. Arch Pathol Lab Med 2014; 138: 1432-1443.
56. Varga Z, Diebold J, Dommann-Scherrer C, et al. How reliable is Ki-67 immunohistochemistry in grade 2 breast carcinomas? A QA study of the Swiss Working Group of Breast- and Gynecopathologists. PLoS One 2012; 7: e37379.
57. Arima N, Nishimura R, Osako T, et al. The importance of tissue handling of surgically removed breast cancer for an accurate assessment of the Ki-67 index. J Clin Pathol 2016; 69: 255-259.
58. Vyberg M, Nielsen S. Proficiency testing in immunohistochemistry--experiences from Nordic Immunohistochemical Quality Control (NordiQC). Virchows Arch 2016; 468: 19-29.
59. Romero Q, Bendahl PO, Fernö M, et al. A novel model for Ki67 assessment in breast cancer. Diagn Pathol 2014; 9: 118.
60. Sato M, Kojima M, Nagatsuma AK, et al. Optimal fixation for total preanalytic phase evaluation in pathology laboratories: a comprehensive study including immunohistochemistry, DNA, and mRNA assays. Pathol Int 2014; 64: 209-215.
61. Polley MY, Leung SC, Gao D, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol 2015; 28: 778-786.
62. Denkert C, Budczies J, von Minckwitz G, et al. Strategies for developing Ki67 as a useful biomarker in breast cancer. Breast 2015; 24 Suppl 2: S67-S72.
63. Vörös A, Csörgő E, Nyári T, et al. An intra- and interobserver reproducibility analysis of the Ki-67 proliferation marker assessment on core biopsies of breast cancer patients and its potential clinical implications. Pathobiology 2013; 80: 111-118.
64. Coates AS, Winer EP, Goldhirsch A, et al. Tailoring therapies – improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol 2015; 26: 1533-1546.
65. Garcia-Carbonero R, Sorbye H, Baudin E, et al. ENETS consensus guidelines for high-grade gastroenteropancreatic neuroendocrine tumors and neuroendocrine carcinomas. Neuroendocrinology 2016; 103: 186-194.
66. Pezzilli R, Partelli S, Cannizzaro R, et al. Ki-67 prognostic and therapeutic decision driven marker for pancreatic neuroendocrine neoplasms (PNENs): A systematic review. Adv Med Sci 2016; 61: 147-153.
67. Panzuto F, Boninsegna L, Fazio N, et al. Metastatic and locally advanced pancreatic endocrine carcinomas: analysis of factors associated with disease progression. J Clin Oncol 2011; 29: 2372-2377.
68. La Rosa S, Klersy C, Uccella S, et al. Improved histologic and clinicopathologic criteria for prognostic evaluation of pancreatic endocrine tumors. Hum Pathol 2009; 40: 30-40.
69. Pape UF, Jann H, Müller-Nordhorn J, et al. Prognostic relevance of a novel TNM classification system for upper gastroenteropancreatic neuroendocrine tumors. Cancer 2008; 113: 256-255.
70. Öberg K, Knigge U, Kwekkeboom D, et al. Neuroendocrine gastro-entero-pancreatic tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 2012; 23 Suppl 7: vii124-vii130.
71. O'Toole D, Couvelard A, Rebours V, et al. Molecular markers associated with response to chemotherapy in gastro-entero-pancreatic neuroendocrine tumors. Endocr Relat Cancer 2010; 17: 847-856.
72. Singla S, LeVea CM, Pokuri VK, et al. Ki67 score as a potential predictor in the selection of liver-directed therapies for metastatic neuroendocrine tumors: a single institutional experience. J Gastrointest Oncol 2016; 7: 441-448.
73. Gudlaugsson E, Skaland I, Janssen EA, et al. Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. Histopathology 2012; 61: 1134-1144.
74. Plancoulaine B, Laurinaviciene A, Herlin P, et al. A methodology for comprehensive breast cancer Ki67 labeling index with intra-tumor heterogeneity appraisal based on hexagonal tiling of digital image analysis data. Virchows Arch 2016; DOI: 10.1007/s00428-015-1865-x
75. Mikami Y, Ueno T, Yoshimura K, et al. Interobserver concordance of Ki67 labeling index in breast cancer: Japan Breast Cancer Research Group Ki67 ring study. Cancer Sci 2013; 104: 1539-1543.
76. Lopez XM, Debeir O, Maris C, et al. Clustering methods applied in the detection of Ki67 hot-spots in whole tumor slide images: an efficient way to characterize heterogeneous tissue-based biomarkers. Cytometry A 2012; 81: 765-775.
77. Khan Niazi MK, Yearsley MM, Zhou X, et al. Perceptual clustering for automatic hotspot detection from Ki-67-stained neuroendocrine tumour images. J Microsc 2014; 256: 213-225.
78. Christgen M, von Ahsen S, Christgen H, et al. The region-of-interest size impacts on Ki67 quantification by computer-assisted image analysis in breast cancer. Hum Pathol 2015; 46: 1341-1349.
79. Lu H, Papathomas TG, van Zessen D, et al. Automated Selection of Hotspots (ASH): enhanced automated segmentation and adaptive step finding for Ki67 hotspot detection in adrenal cortical cancer. Diagnostic Pathology 2014; 9: 216.
80. Świderska Z, Korzyńska A, Markiewicz T, et al. Comparison of the manual, semiautomatic, and automatic selection and leveling of hot spots in whole slide images for Ki-67 quantification in meningiomas. Anal Cell Pathol (Amst) 2015; 2015: 498746.
81. Couvelard A. Neuroendocrine tumours of the pancreas: recent developments in staging and grading. Diagn Histopathol 2011; 18: 1.
82. Coriat R, Walter T, Terris B, et al. Gastroenteropancreatic well-differentiated grade 3 neuroendocrine tumors: review and position statement. Oncologist 2016; 21: 1191-1199.
83. Carlinfante G, Baccarini P, Berretti D, et al. Ki-67 cytological index can distinguish well-differentiated from poorly differentiated pancreatic neuroendocrine tumors: a comparative cytohistological study of 53 cases. Virchows Arch 2014; 465: 49-55.
84. Focke CM, Decker T, van Diest PJ. Intratumoral heterogeneity of Ki67 expression in early breast cancers exceeds variability between individual tumours. Histopathology 2016; 69: 849-861.

85. Voss SM, Riley MP, Lokhandwala PM, et al. Mitotic count by phosphohistone H3 immunohistochemical staining predicts survival and improves interobserver reproducibility in well-differentiated neuroendocrine tumors of the pancreas. Am J Surg Pathol 2015; 39: 13-24.
86. Cserni G, Vörös A, Liepniece-Karele I, et al. Distribution pattern of the Ki67 labelling index in breast cancer and its implications for choosing cut-off values. Breast 2014; 23: 259-263.
87. Vörös A, Csörgő E, Kővári B, et al. Different methods of pretreatment Ki-67 labeling index evaluation in core biopsies of breast cancer patients treated with neoadjuvant chemotherapy and their relation to response to therapy. Pathol Oncol Res 2015; 21: 147-155.
88. Klimstra DS. Pathology reporting of neuroendocrine tumors: essential elements for accurate diagnosis, classification, and staging. Semin Oncol 2013; 40: 23-36.

Address for correspondence

Łukasz Liszka MD
Department of Pathomorphology in Katowice
Medical University of Silesia
Medyków 18
40-754 Katowice, Poland
tel./fax +48 32 25 25 080
e-mail: lliszka@mp.pl