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

András Vörös[a]   Erika Csörgő[a]   Tibor Nyári[b]   Gábor Cserni[a, c]

Departments of [a]Pathology and [b]Medical Physics and Informatics, University of Szeged, Szeged, and
[c]Department of Pathology, Bács-Kiskun County Teaching Hospital, Kecskemét, Hungary

Abstract

Objective: The Ki-67 proliferation index has received an important role in treatment tailoring and molecular classification of estrogen receptor-positive breast carcinomas. The aim was to analyze the reproducibility of assessing proliferation on the basis of Ki-67 immunohistochemistry. *Methods:* Thirty core biopsy samples of breast cancer patients were analyzed after immunostaining with B56, SP6 and MIB-1 monoclonal Ki-67 antibodies. All samples were evaluated twice and independently by 3 pathologists, with each observer performing his daily routine practice. The ratio of Ki-67-positive cells was estimated with 5% accuracy. Correlation was calculated for the results of each investigator for all pairs of antibodies and for the results of each antibody for all pairs of investigators. Ki-67 scores were divided into categories of either 4 quarters or into 3 groups reflecting the St. Gallen consensus recommendations with 15 and 30% as cutoff values. The reproducibility of classifying the tumors into these categories was assessed with κ statistics. *Results:* Altogether, 540 evaluations were made. Good to excellent correlation (Spearman's and Pearson's coefficient range 0.74–0.92 and 0.73–0.93, respectively) was noted for the pairwise comparison of antibodies by observer and of observers by antibody. The inter- and intraobserver reproducibility of the Ki-67 score classification into equal quarters (1–25, 26–50, 51–75 and 76–100%) or into 3 categories with cutoffs at 15 and 30% was fair to poor in the middle categories, but moderate to substantial in the low and high ranges. Interobserver differences in practice potentially impacted on less consistent classification. *Conclusion:* Our results indicate that the three different Ki-67 antibodies tested do not substantially influence the reproducibility of the estimated proliferation rates. Although reproducibility is better in the clinically more relevant distinction of high versus low proliferation, without standardization, the current practice of Ki-67 assessment in many laboratories does not allow proper and consistent therapeutic decision-making.

Copyright © 2012 S. Karger AG, Basel

Introduction

Proliferation is an important feature of malignant tumors including breast cancer [1–4]. Various methods have been established for its quantification, including mi-

András Vörös
Department of Pathology, University of Szeged
Állomás u. 2
HU–6720 Szeged (Hungary)
E-Mail andrasvoros@libero.it

totic index counting, measurement of S-phase fraction by flow cytometry, counting of argyrophilic nucleolar organizer regions, positron emission tomography and immunohistochemistry of proliferation-associated antigens. Beresford et al. [4] have highlighted the limitations of many of these new methods and have suggested Ki-67 immunohistochemistry to be the standard proliferation assay due to its simplicity and wide availability. Ki-67 is a nuclear protein present in all stages of the cell cycle except G0 [5]. The anti-Ki-67 antibody labeling index (LI, percentage of nuclear antigen-positive cells) has been proposed as a prognostic and predictive marker for breast cancer, especially for estrogen receptor (ER)-positive, HER2-negative early breast carcinoma [1, 6–8]. Different studies have demonstrated that a high Ki-67 LI indicates an increased risk of recurrence [7, 9, 10], metastasis [7–9, 11–15] and faster progression of the disease. Some publications emphasize that adding chemotherapy to endocrine therapy is a recommended treatment in this population of patients [15]. Different cutoff values of Ki-67 LI have been proposed for the indication of chemotherapy for these patients [9, 12, 16–18]. Owing to the relevance of this factor in recommending the addition of chemotherapy, the reproducibility of Ki-67 LI evaluation is highly important.

With Ki-67 immunohistochemistry, proliferating tumor cells show nuclear positivity, and pathologists provide the percentage of cells with nuclear immunostaining (Ki-67 LI). This relatively subjective evaluation of Ki-67 LI can lead to a high degree of interobserver variability, e.g., there are pathologists who estimate the percentage of nuclear staining, whereas others count the number of positive cells in different fields of the tumor area [1, 19]. Besides such methodological differences, there are other factors including tissue fixation, heterogeneity of staining in different tumor areas (e.g., periphery vs. central part), heterogeneity of nuclear labeling intensity (according to the variation of the amount of protein during the cell cycle) and human or technical performance that also generate interobserver variability. Although a Ki-67 LI cutoff value was defined at the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer for recommending adjuvant chemotherapy in endocrine-responsive breast carcinomas [20–22], there are no standardized methods to eliminate the different influential factors mentioned above. In this study, we analyzed the reproducibility of the Ki-67 expression levels by types of antibody and investigator method in core biopsy samples of breast cancer patients.

Materials and Methods

Patients and Samples

Core biopsy samples of patients with operable T2 ≥3 cm or T3–4 and/or N1–2 and M0 breast cancer candidates for neoadjuvant docetaxel-epirubicin with/without capecitabine chemotherapy were retrospectively analyzed [23]. Samples had been taken between January 2003 and December 2011 at the Department of Radiology, University of Szeged, or at the Bács-Kiskun County Teaching Hospital. The tumor samples were fixed in buffered formalin and embedded in paraffin. Samples have been routinely stained with hematoxylin and eosin and routinely immunostained for ER, progesterone receptor, HER-2 and topoisomerase IIα. For the purpose of this study, they were immunostained for Ki-67 with the following 3 antibodies: SP6 (monoclonal rabbit antibody; Hisztopatologia Kft., Pécs, Hungary), B56 (monoclonal mouse antibody; Hisztopatologia Kft.) and MIB-1 (monoclonal mouse antibody; Dako, Glostrup, Denmark) (fig. 1). Wet antigen retrieval consisted of pretreatment of all samples in a microwave oven in a citrate buffer with pH 6 for 30, 30 and 50 min in the case of MIB-1, B56 and SP6, respectively. All antibodies were diluted at 1:100. Expression of Ki-67 was determined using Dako EnVision FLEX/HRP, DAB+ Chromogen (Dako). In the present study, all samples were assessed independently by 3 pathologists at high-power magnification (×400). The ratio of Ki-67-positive cells was estimated with 5% accuracy; therefore, only values ending with 5 or 0 were recorded. Each observer was asked to use his daily evaluation approach to quantify the proportion of Ki-67-positive cells and to perform the evaluation of all cases stained with one antibody first, followed by all cases stained with the second and third antibody, in order to avoid bias arising from remembering the LI of a given sample. The assessments were done twice with an interval of at least 2 months.

The results of each investigator were analyzed for all pairs of antibodies, and the results of the different antibodies were analyzed for all pairs of investigators. To assess how the parameters correlate with each other, Spearman's rank correlation was used in these pairwise analyses. Similarly, assuming an ideal linear relationship between Ki-67 LI values, Pearson's coefficients were calculated for the same pairs. To further investigate the influence of the observers and the antibody on the Ki-67 LI value, two-way ANOVA was performed. The computations were done with the statistical software package SPSS 15.0 for Windows (SPSS Inc., Chicago, Ill., USA).

Ki-67 scores were divided into four quarters (0–25, 26–50, 51–75 and 76–100% Ki-67 LI) to allow categorical data analyses. Four equally sized categories were arbitrarily chosen to limit their number and to allow a better analysis of the consistency of rating into a given category. As the 15% limit has been proposed as a cutoff for the low proliferation category and the 30% limit as a cutoff for the high proliferative category, grouping according to these two marginal values (into 3 categories of unequal size) was also evaluated. Interobserver, interreagent and intraobserver agreements were assessed with κ statistics according to Fleiss [24]. Interobserver reproducibility was also evaluated by determining pairwise weighted κ values [24]. These weighted κ values do not only take into account the classification into another category (non-agreement) but also give weight to the 'distance' between the ordinal categories that have been used for classifying the Ki-67 values (e.g., having two ratings in neighboring categories is better than having them in categories separated by another category).
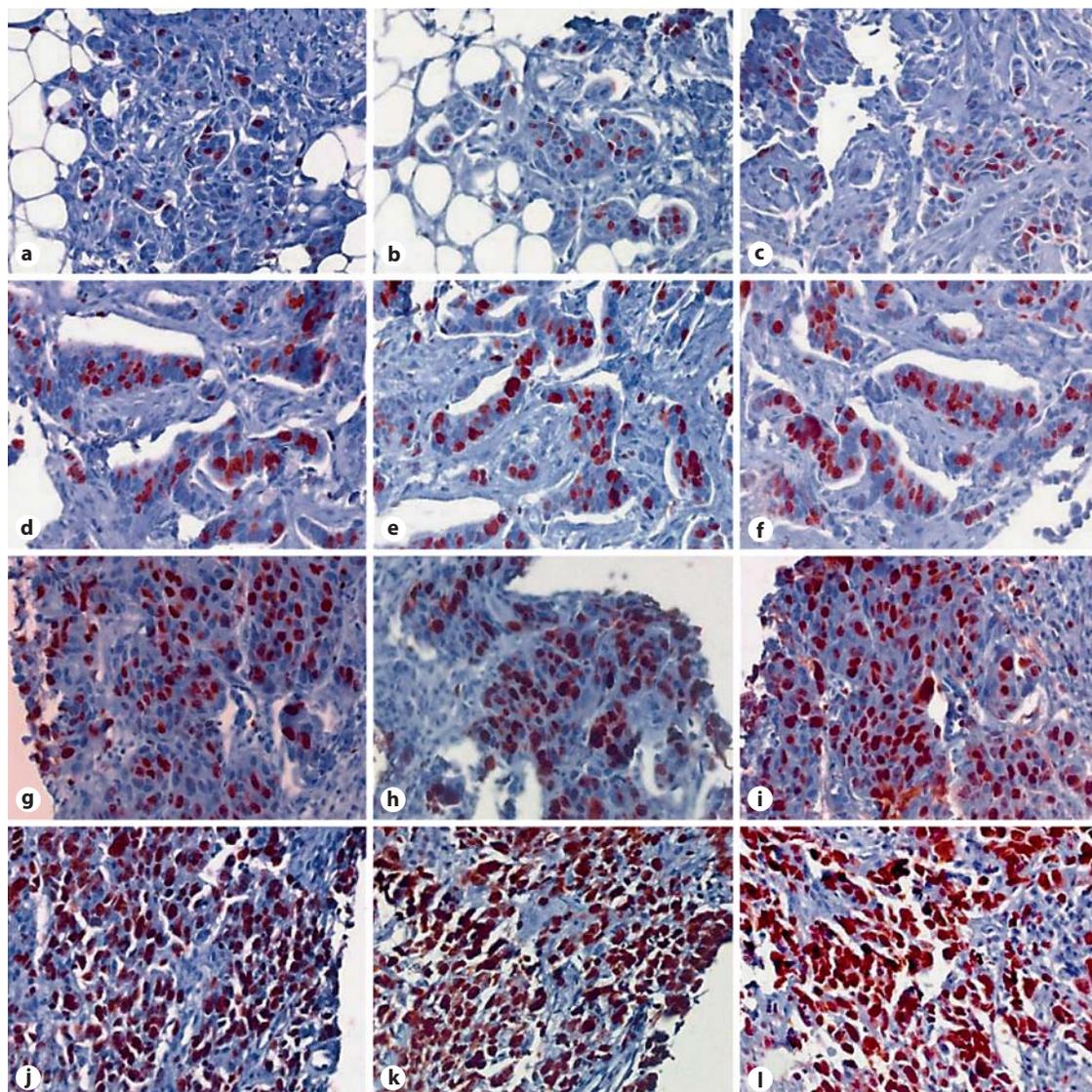
Fig. 1. Examples of tumor samples belonging to each quarter category (1–25%, **a–c**; 25–50%, **d–f**; 51–75%, **g–i**; 76–100%, **j–l**) of Ki-67-positive cells on the basis of majority opinion, immunostained with the 3 different Ki-67 antibodies (SP6, **a**, **d**, **g**, **j**; B56, **b**, **e**, **h**, **k**; MIB1, **c**, **f**, **i**, **l**). ×20.

The κ values were interpreted as reflecting slight (0–0.2), fair (0.21–0.4), moderate (0.41–0.6), substantial (0.61–0.8) and almost perfect (>0.8) agreement between observations according to Landis and Koch [25].

Results

Thirty core biopsy samples of breast cancer patients were analyzed. The mean age ± SE of the population was 46 ± 2 years (range 26–70). Samples included 28 invasive ductal carcinomas and 2 invasive lobular carcinomas. Altogether, 270 evaluations were made by 3 independent pathologists with expertise in breast pathology (G.C., A.V. and E.C.), and these were repeated a second time to better assess intraobserver variability. Mean ± SE Ki-67 LI values of the first and the second evaluations by the 3 pathologists were 45 ± 2, 52 ± 2 and 56 ± 2% for the SP6, B56 and MIB-1 antibodies, respectively.

Spearman's rank correlation was used for comparison of the different observers and antibodies. Each anal-

Table 1. κ values per observer and antibody according to classification into four quarters categories and into 3 categories determined by 2 cutoff values

a Quarter categories

	0–25%	26–50%	51–75%	76–100%	Overall
All[1]	0.47	0.24	0.14	0.52	0.34
Observer 1	0.63	0.32	0.15	0.51	0.39
Observer 2	0.45	0.13	0.03	0.54	0.3
Observer 3	0.58	0.38	0.28	0.55	0.44
MIB-1	0.51	0.24	0.15	0.54	0.36
B56	0.48	0.27	0.11	0.52	0.35
SP6	0.44	0.2	0.15	0.57	0.32

b Categories determined by cutoff values

	0–15%	16–30%	>30%	Overall
All[1]	0.42	0.19	0.52	0.39
Observer 1	0.61	0.38	0.61	0.52
Observer 2	0.42	0.13	0.51	0.37
Observer 3	0.4	0.24	0.54	0.41
MIB-1	0.35	0.26	0.56	0.41
B56	0.38	0.1	0.53	0.37
SP6	0.45	0.29	0.56	0.44

[1] All refers to the 18 assessments of the same cases with 3 antibodies, by 3 observers, twice. The remaining values are based on 6 assessments each: for the observers, 3 antibodies, twice; for the antibodies, 3 observers, twice.

Table 2. κ values reflecting intraobserver reproducibility of the classification of Ki-67 LIs into 4 or 3 categories

a Quarter categories

	0–25%	26–50%	51–75%	76–100%	Overall
Observer 1	0.76	0.34	0.13	0.51	0.42
Observer 2	0.4	0.01	0	0.61	0.26
Observer 3	0.55	0.32	0.3	0.51	0.41

b Categories determined by cutoff values

	<16%	16–30%	>30%	Overall
Observer 1	0.64	0.5	0.69	0.61
Observer 2	0.37	0.07	0.57	0.36
Observer 3	0.45	0.32	0.64	0.5

ysis was performed on the first and the second evaluation separately and then combined. Interobserver comparisons for the dual (combined) evaluation suggested that the correlation between ratings was good to excellent (coefficients ranging between 0.74 and 0.91; p < 0.0001). Interantibody comparisons yielded coefficients ranging between 0.8 and 0.92 (p < 0.0001), suggesting excellent interantibody correlation. Pearson's coefficients ranged from 0.73 to 0.91 (p < 0.0001) for observer pairs and from 0.79 to 0.93 (p < 0.0001) for antibody pairs, suggesting similarly good to excellent correlation. When all 18 ratings (3 observers, 3 antibodies, twice) of the 30 cases were considered, allocation of the cases determined by the majority was as follows: 1st quarter (0–25% Ki-67 LI), 5 cases; 2nd quarter, 12 cases; 3rd quarter, 6 cases; 4th quarter, 7 cases. Only 2 tumors had all ratings falling into the same quarter: 1 belonging to the lowest and 1 to the highest proliferation group. For other categorization, the majority of classifications suggested that 3 cases had a Ki-67 LI ≤15%, 4 cases fell into the 16–30% LI category, and 23 cases had a high proliferation, with an LI >30%. All 10 cases with a 100% agreement in the classification belonged to the highly proliferative tumors.

On the basis of the κ values, the reproducibility of the Ki-67 LI values by quarter category distribution was generally moderate to substantial in the lowest and the highest ends, but was only slight to fair in the middle quarter categories (table 1a). This resulted in overall κ values reflecting only fair reproducibility both for given observers and for given antibodies. The same statement can be made after the evaluation of the dual assessment of all cases with all antibodies (18 evaluations per case). Pairwise weighted κ values showed somewhat better interobserver consistency (range 0.43–0.75), reflecting moderate to substantial reproducibility. When looking at the distributions determined by the 15 and 30% cutoff values, the reproducibility of the highly proliferative classification remained moderate/substantial, but the low and intermediate proliferation groups were less reproducible (table 1b). Pairwise weighted κ values ranged between 0.3 and 0.78 for these categories and also suggested fair to substantial reproducibility. The better reproducibility reflected by the weighted κ values indicates that inconsistent categorizations were often one category away from each other, i.e., low proliferation rates were unlikely to be ranked as high and vice versa.

Similar statements can be made concerning the intraobserver agreements determined for all 90 immunostained samples (table 2): tumors with a high and a low

Ki-67 LI can be more reproducibly identified than tumors falling into the intermediate (26–75 or 16–30%) range both on the basis of the quarter-based 4-tiered distribution or the two cutoffs defined by 3-tiered distribution. However, differences between observers can also be identified. Observer 2 demonstrated a somewhat worse performance than the other 2 observers: he had only 47% identical categorizations on a quarter basis (in contrast to 57 and 58% for the 2 other observers) and 63% identical categorizations on the basis of the 3-tiered classification (in contrast to 80 and 83% for the 2 other observers).

Two-way ANOVA has confirmed the fact that observer 2 rated the cases significantly lower than the other 2 observers ($p = 0.004$ for first observations and $p = 0.008$ for second observations) and also suggested that one antibody (SP6) resulted in significantly lower mean Ki-67 LI values (45.2% overall vs. 52.3% for B56 and 55.7% for MIB-1; $p = 0.017$ for first-round observations and $p = 0.01$ for second-round observations).

The individual evaluation patterns were also evaluated. It turned out that all observers aimed to quantify the area with the highest number of stained cells; very weakly, faintly staining nuclei were generally discarded. All observers estimated the proportion of positive cells in relation to approximately 100 tumor cells of the chosen area, but there were also differences in the way of estimating this proportion. Observer 1 counted 100 cells and recounted the number of positive nuclei in the same 100 cells. Care was taken to keep in mind the tumor structures included and excluded in the 100-cell-containing area to allow the count of the positive nuclei in the same 100 cells. Observer 2 counted 10 tumor cells to get an impression of the area these cells occupied and included an area of tumor cells 10 times greater than this basic count, and finally, effectively counted the number of positive nuclei in the area estimated to contain 100 tumor cells. Observer 3 made an approaching guess in a larger area of the hot spot, then counted the proportion of stained nuclei in 20 or 30 cells and multiplied it by 4 or 3 to get a percentage estimate. For cases with a very high LI, a reverse count was made by evaluating the unstained nuclei. Therefore, the used methods were comparable in some respects, but differed in their details of cell counting. All were generally based on different approximations of the proportion of nuclear staining in relation to 100 cancer cells in hot spot areas and neither tried to better estimate this ratio by counting 1,000 cells as suggested by some authors [8, 26].

Discussion

Proliferation has been proposed as an important prognosticator of breast carcinomas, and the Ki-67 LI has been implied as a factor enabling a distinction between tumors with a high and a low proliferation. In this respect, Ki-67 has been used in various settings including the addition of adjuvant chemotherapy to hormonal therapy for hormone-sensitive breast carcinomas (St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer, 2009), the distinction of luminal A and luminal B subsets of ER-positive and HER-2 negative carcinomas [26] or the distinction between histological grade 2 carcinomas with an outcome similar to grade 1 or grade 3 cancers [27].

The present study suggests that the correlation between assessments by different observers or using different antibodies is good or excellent, but that interobserver reproducibility of the Ki-67 LI may only be fair in everyday pathology practice. This may cast some doubt on the general usefulness of this marker in its present state.

The data also suggest that the classification of tumors into low and high proliferation categories is better (moderate to substantial) than that of tumors with an intermediate proliferative activity. This proved true both for cutoffs of 25 and 75% (per equal quarters assessment) and for cutoffs at 15 and 30% as proposed by the St. Gallen Expert Consensus [20]. This may also mean that the range where the distinction is clinically important is characterized by better reproducibility.

We also noted a rather consistent intraobserver reproducibility variation. The κ values relating to observer 2 were consistently lower than the values of the other 2 observers, although the high and low proliferation categories had higher κ values even for this observer. This may reflect the differences in the technique of evaluating the Ki-67 LI. All observers made the LI estimation on the basis of approximately 100 cells in a hot spot area. Observer 1 counted real percentage values by counting 100 cells and recounting the positive (or at times the negative) nuclei in the same cells. Observer 2 approximated the area containing 100 cells by first delineating an area with one tenth of this population and then counted the positive tumor cell nuclei in this area. Finally, observer 3 counted the proportion of immunostained tumor cell nuclei in one fifth or one third of the area and then extrapolated this result to 100 cells. The two latter techniques spare time. Interestingly, observer 3 had κ values very comparable with those of observer 1. The most 'time-consuming' evaluation, that of observer 1, took 3–4 min per case

and 90–100 min for scoring the 30 cases in the series once.

MIB-1 is generally used in everyday practice to determine the Ki-67 LI on formalin-fixed and paraffin-embedded material. This antibody was generated using recombinant technology and reacts with the immunodominant area of the Ki-67 nuclear antigen [28]. B56 is directed against the same area and also represent an immunoglobulin G1 mouse monoclonal antibody. Although SP6 is a rabbit monoclonal antibody directed against the C terminus of the Ki-67 protein, it has recently been reported to correlate well with MIB-1 staining [29]. SP6 showed significantly lower mean Ki-67 LI values in both the first- and the second-round assessments than the other antibodies, but this did not influence reproducibility. In keeping with the above, the 3 antibodies used had very similar overall κ values reflecting fair to moderate reproducibility for the whole range of the cases. Again, reproducibility of classifying the cases into a low and a high proliferation category according to either cutoffs tested was better than classifying the cases into 'grey zone' intermediate categories.

It is also important to highlight that our findings were retrieved on core biopsy samples. Core biopsies might not always contain 1,000 tumor cells for counting, as non-tumor tissue may be a common part of such specimens. Of the different factors potentially influencing the Ki-67 LI, tissue fixation, heterogeneity of different tumor areas and heterogeneity of nuclear labeling intensity are infrequent in such small samples, and this may improve reproducibility, although, on the other hand, the representative nature of a small sample may be questionable. For example, Romero et al. [30] showed significant differences between core biopsy and surgical sample proliferation values, which may confirm our assumption that core biopsy specimens allow better reproducibility on the basis of a smaller sample with lesser variability.

This is not the first study to look at the reproducibility of the Ki-67 LI. Several other studies casted doubts on the reproducibility of Ki-67 LI evaluation [1, 19, 30]. Although the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer (2009 and 2011) recommended an accurate methodical proceeding (e.g., counting 1,000 tumor cells), the everyday use and ability to follow this recommendation is questionable. As suggested in the present series, even intralaboratory differences in assessment may exist (pathologists may tend to use less time-consuming methods). Despite only fair overall reproducibility of the Ki-67 LI into categories of different proliferative activity, the results between the observers and the antibodies were at least moderately correlated, and the identification of tumors with the highest proliferation was better (moderate to substantial) than that of tumors with intermediate proliferation according to both approaches tested (although this is the result of a universal statistical occurrence, we should be more aware of this phenomenon in clinical decision making and acknowledge that intermediate or 'grey-zone' categories are generally less reproducible than low- and high-value categories, as exemplified by our data). Therefore, it seems that clinically important categorizations (high vs. non-high proliferation) can be more consistently made even with the methods described, although it is believed that substantial to excellent reproducibility would make the use of Ki-67 as a biomarker more confident. The histological grade of breast carcinomas is also a potential factor that may affect therapeutic decisions, but grading was not much better reproducible among expert breast pathologists, with overall κ values ranging from 0.43 to 0.59, and grade 2 tumors being less reproducibly classified as such than grade 1 or grade 3 tumors [31]. Furthermore, although Ki-67 LI has been reported as a potential tool to discriminate between grade 2 tumors of better and worse prognosis [27], the reproducibility of Ki-67 LI in this group of tumors was recently found to be inconsistently reproducible [32]. Even those things that seem easy, like identifying breast carcinomas with 90–100% staining for ERs, may seem suboptimally reproducible on the basis of κ values: the consistency among expert breast pathologists was not much better with a κ value of 0.52, although finding a good cutoff (10% at that time) allowed excellent reproducibility of categorizing the tumors into negative versus positive [33].

In conclusion, our results suggest that reproducibility of the Ki-67 LI is less than optimal even in core biopsies of breast cancer patients. However, the reproducibility of classifying tumors into a clinically more important highly proliferative category (like >15, >25 or >30%) is better than that of the overall classification and is not dependent on the antibody used (at least true for the 3 antibodies tested). At present, the consistency of Ki-67 LI determination in the routine work of some (probably numerous) laboratories (including ours) does not allow error-free therapeutic decision making on a yes or no basis. Our results also indicate that similar but slightly differing individual practices of Ki-67 LI evaluation by different observers may influence reproducibility, and therefore, reasonable standardization and the recommendation of a workable uniform method should be encouraged.

References

1 Luporsi E, André F, Spyratos F, Martin PM, Jacquemier J, Penault-Llorca F, Tubiana-Mathieu N, Sigal-Zafrani B, Arnould L, Gompel A, Egele C, Poulet B, Clough KB, Crouet H, Fourquet A, Lefranc JP, Mathelin C, Rouyer N, Serin D, Spielmann M, Haugh M, Chenard MP, Brain E, de Cremoux P, Bellocq JP: Ki-67: level of evidence and methodological considerations for its role in the clinical management of breast cancer: analytical and critical review. Breast Cancer Res Treat 2012; 132:895–915.

2 Oakman C, Bessi S, Zafarana E, Galardi F, Biganzoli L, Di Leo A: Recent advances in systemic therapy: new diagnostics and biological predictors of outcome in early breast cancer. Breast Cancer Res 2009;11:205.

3 Rahmanzadeh R, Hüttmann G, Gerdes J, Scholzen T: Chromophore-assisted light inactivation of pKi-67 leads to inhibition of ribosomal RNA synthesis. Cell Prolif 2007;40: 422–430.

4 Beresford MJ, Wilson GD, Makris A: Measuring proliferation in breast cancer: practicalities and applications. Breast Cancer Res 2006;8:216.

5 Gerdes J, Li L, Schlueter C, Duchrow M, Wohlenberg C, Gerlach C, Stahmer I, Kloth S, Brandt E, Flad HD: Immunobiochemical and molecular biologic characterization of the cell proliferation-associated nuclear antigen that is defined by monoclonal antibody Ki-67. Am J Pathol 1991;138:867–873.

6 Clahsen PC, van de Velde CJ, Duval C, Pallud C, Mandard AM, Delobelle-Deroide A, van den Broek L, van de Vijver MJ: The utility of mitotic index, oestrogen receptor and Ki-67 measurements in the creation of novel prognostic indices for node-negative breast cancer. Eur J Surg Oncol 1999;25:356–363.

7 Goodson WH 3rd, Moore DH 2nd, Ljung BM, Chew K, Mayall B, Smith HS, Waldman FM: The prognostic value of proliferation indices: a study with in vivo bromodeoxyuridine and Ki-67. Breast Cancer Res Treat 2000;59:113–123.

8 Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA: Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol 2010;11:174–183.

9 Fasching PA, Heusinger K, Haeberle L, Niklos M, Hein A, Bayer CM, Rauh C, Schulz-Wendtland R, Bani MR, Schrauder M, Kahmann L, Lux MP, Strehl JD, Hartmann A, Dimmler A, Beckmann MW, Wachter DL: Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment. BMC Cancer 2011;11:486.

10 Jung SY, Han W, Lee JW, Ko E, Kim E, Yu JH, Moon HG, Park IA, Oh DY, Im SA, Kim TY, Hwang KT, Kim SW, Noh DY: Ki-67 expression gives additional prognostic information on St. Gallen 2007 and Adjuvant! Online risk categories in early breast cancer. Ann Surg Oncol 2009;16:1112–1121.

11 Mohammed ZM, McMillan DC, Elsberger B, Going JJ, Orange C, Mallon E, Doughty JC, Edwards J: Comparison of visual and automated assessment of Ki-67 proliferative activity and their impact on outcome in primary operable invasive ductal breast cancer. Br J Cancer 2012;106:383–388.

12 Kashiwagi S, Yashiro M, Takashima T, Aomatsu N, Ikeda K, Ogawa Y, Ishikawa T, Hirakawa K: Advantages of adjuvant chemotherapy for patients with triple-negative breast cancer at stage II: usefulness of prognostic markers E-cadherin and Ki67. Breast Cancer Res 2011;13:R122.

13 Murri AM, Hilmy M, Bell J, Wilson C, McNicol AM, Lannigan A, Doughty JC, McMillan DC: The relationship between the systemic inflammatory response, tumour proliferative activity, T-lymphocytic and macrophage infiltration, microvessel density and survival in patients with primary operable breast cancer. Br J Cancer 2008;99: 1013–1019.

14 Viale G, Regan MM, Mastropasqua MG, Maffini F, Maiorano E, Colleoni M, Price KN, Golouh R, Perin T, Brown RW, Kovács A, Pillay K, Ohlschlegel C, Gusterson BA, Castiglione-Gertsch M, Gelber RD, Goldhirsch A, Coates AS, International Breast Cancer Study Group: Predictive value of tumor Ki-67 expression in two randomized trials of adjuvant chemoendocrine therapy for node-negative breast cancer. J Natl Cancer Inst 2008;100:207–212.

15 Viale G, Giobbie-Hurder A, Regan MM, Coates AS, Mastropasqua MG, Dell'Orto P, Maiorano E, MacGrogan G, Braye SG, Ohlschlegel C, Neven P, Orosz Z, Olszewski WP, Knox F, Thürlimann B, Price KN, Castiglione-Gertsch M, Gelber RD, Gusterson BA, Goldhirsch A, Breast International Group Trial 1-98: Prognostic and predictive value of centrally reviewed Ki-67 labeling index in postmenopausal women with endocrine-responsive breast cancer: results from Breast International Group Trial 1-98 comparing adjuvant tamoxifen with letrozole. J Clin Oncol 2008;26:5569–5575.

16 Spyratos F, Ferrero-Poüs M, Trassard M, Hacène K, Phillips E, Tubiana-Hulin M, Le Doussal V: Correlation between MIB-1 and other proliferation markers: clinical implications of the MIB-1 cutoff value. Cancer 2002;94:2151–2159.

17 Jones RL, Salter J, A'Hern R, Nerurkar A, Parton M, Reis-Filho JS, Smith IE, Dowsett M: Relationship between oestrogen receptor status and proliferation in predicting response and long-term outcome to neoadjuvant chemotherapy for breast cancer. Breast Cancer Res Treat 2010;119:315–323.

18 Dowsett M, A'Hern R, Salter J, Zabaglo L, Smith IE: Who would have thought a single Ki67 measurement would predict long-term outcome? Breast Cancer Res DOI: 10.1186/bcr2434.

19 Fasanella S, Leonardi E, Cantaloni C, Eccher C, Bazzanella I, Aldovini D, Bragantini E, Morelli L, Cuorvo LV, Ferro A, Gasperetti F, Berlanda G, Dalla Palma P, Barbareschi M: Proliferative activity in human breast cancer: Ki-67 automated evaluation and the influence of different Ki-67 equivalent antibodies. Diagn Pathol DOI: 10.1186/1746-1596-6-S1-S7.

20 Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thürlimann B, Senn HJ, Panel members: Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. Ann Oncol 2009;20:1319–1329.

21 Gnant M, Harbeck N, Thomssen C: St. Gallen 2011: Summary of the Consensus Discussion. Breast Care (Basel) 2011;6:136–141.

22 Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, Ellis M, Henry NL, Hugh JC, Lively T, McShane L, Paik S, Penault-Llorca F, Prudkin L, Regan M, Salter J, Sotiriou C, Smith IE, Viale G, Zujewski JA, Hayes DF, 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.

23 Nikolényi A, Sükösd F, Kaizer L, Csörgo E, Vörös A, Uhercsák G, Ormándi K, Lázár G, Thurzó L, Brodowicz T, Kahán Z: Tumor topoisomerase II alpha status and response to anthracycline-based neoadjuvant chemotherapy in breast cancer. Oncology 2011;80: 269–277.

24 Fleiss JL: The measurement of interrater agreement; in Fleiss JL (ed): Statistical Methods for Rates and Proportions, ed 2. New York, Wiley and Sons, 1980, pp 212–236.

25 Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977;33:159–174.

26 Cheang MCU, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, Perou CM, Ellis MJ, Nielsen TO: Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Inst Cancer 2009;101:736–750.

27 Aleskandarany MA, Rakha EA, Macmillan RD, Powe DG, Ellis IO, Green AR: MIB1/Ki-67 labelling index can classify grade 2 breast cancer into two clinically distinct subgroups. Breast Cancer Res Treat 2011;127:591–599.

28 Kubbutat MHG, Key G, Duchrow M, Schlüter C, Flad HD, Gerdes J: Epitope analysis of antibodies recognising the cell proliferation associated nuclear antigen previously defined by the antibody Ki-67 (Ki-67 protein). J Clin Pathol 1994;47:524–528.

29 Zabaglo L, Salter J, Anderson H, Quinn E, Hills M, Detre S, A'Hern R, Dowsett M: Comparative validation of the SP6 antibody to Ki67 in breast cancer. J Clin Pathol 2010; 63:800–804.

30 Romero Q, Bendahl PO, Klintman M, Loman N, Ingvar C, Rydén L, Rose C, Grabau D, Borgquist S: Ki67 proliferation in core biopsies versus surgical samples – a model for neo-adjuvant breast cancer studies. BMC Cancer 2011;11:341.

31 European Commission Working Group on Breast Screening Pathology, Sloane JP, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, Boecker W, Bussolati G, Coleman D, Connolly CE, Eusebi V, De Miguel C, Dervan P, Drijkoningen R, Elston CW, Faverly D, Gad A, Jacquemier J, Lacerda M, Martinez-Penuela J, Munt D, Peterse JL, Rank F, Tsakraklides V, Sylvan M, Zafrani B: Consistency achieved by 23 European pathologists from 12 countries in diagnosing breast disease and reporting prognostic features of carcinomas. Virchows Arch 1999;434:3–10.

32 Varga Zs, Diebold J, Dommann-Scherrer C, Frick H, Kaup D, Noske A, Obermann E, Ohlschlegel C, Padberg B, Rakozy C, Sancho Oliver S, Schobinger-Clement S, Schreiber-Facklam H, Singer G, Tapia C, Wagner U, Mastropasqua M, Viale G, Lehr H-A: 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.

33 Wells CA, Sloane JP, Coleman D, Munt C, Amendoeira I, Apostolikas N, Bellocq JP, Bianchi S, Boecker W, Bussolati G, Connolly CE, Dervan P, Drijkoningen M, Ellis IO, Elston CW, Eusebi V, Faverly D, Heikkila P, Holland R, Jacquemier J, Lacerda M, Martinez-Penuela J, De Miguel C, Peterse JL, Rank F, Reiner A, Saksela E, Sigal-Zafrani B, Sylvan M, Borisch B, Cserni G, Decker T, Kerner H, Kulka J, Regitnig P, Sapino A, Tanous AM, Thorstenson S, Zozaya E, European Working Group for Breast Screening Pathology: Consistency of staining and reporting of oestrogen receptor immunocytochemistry within the European Union – an interlaboratory study. Virchows Arch 2004;445: 119–128.