# Optimizing breast cancer survival models based on conventional biomarkers and stromal parameters

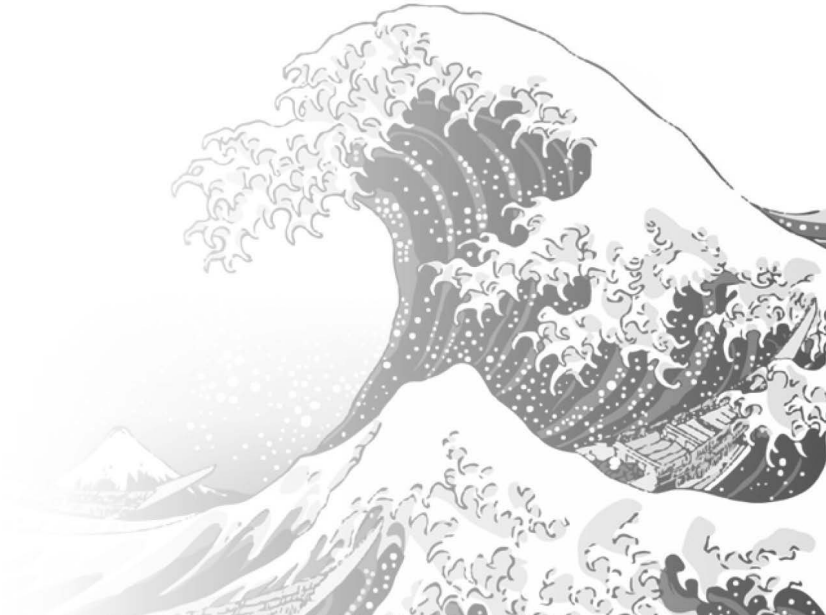
Dekker, T.J.A.; Dekker T.J.A.

# 8

# The impact of tumor heterogeneity, interobserver and interlaboratory variation on the robustness of Ki67 assays in breast cancer: Results from the RASTER study cohort

*Manuscript in preparation*

M.V. Nijenhuis, T.J.A. Dekker, C.A. Drukker, E. Groen, J. Sanders, S. Linn, V.T.H.B.M. Smit, E.J.Th. Rutgers, J. Wesseling



## Introduction

Cellular proliferation is an important feature of malignant tumors. The proliferation marker Ki67 is a nuclear protein that is present in all stages of the cell cycle except G0 [1]. The Ki67 labelling index (LI, i.e. percentage of nuclear antigen-positive cells) has been described as both a prognostic and predictive marker for breast cancer. Gene expression studies have demonstrated at least four molecular subtypes of breast cancer, i.e. Luminal A, Luminal B, HER2-positive, and basal-like [2]. These subtypes can be classified using immunohistochemical markers [3]. According to the 2013 and 2015 St Gallen guidelines, the decision on systemic treatment should be based on these so-called surrogate intrinsic subtypes determined by ER, PgR, HER2, and Ki67 assessment [4]. The ascribed importance of the Ki67 LI to distinguish lowly from highly proliferative breast cancer (Luminal A and B subtypes), and thus guiding decisions regarding chemotherapy, renders the reproducibility of Ki67 assays as highly clinically relevant [5].

If neoadjuvant systemic treatment or radiotherapy is considered, a core needle biopsy (CNB) is used to obtain a classifying diagnosis, including the assessment of prognostic and predictive biomarkers [4]. There is some debate on whether determining these biomarkers on CNB is superior or at least equal to the assessment in surgical specimens. The advantage of CNB might be superior fixation owing to a smaller tissue size. However, tumor heterogeneity might lead to both false-positive and false-negative test results. With minor adjustments to the testing algorithm, reliable ER and HER2 assays can be obtained when using CNBs [6]. Less data exists on how this might affect Ki67 assays.

Tissue microarrays (TMA) experiments allow large scale profiling of tissue samples, including multiple protein measurements. This technique can be used to study biological heterogeneity, similar to small biopsy samples. Our aim was to use this technique to assess the agreement of Ki67 LI between CNB and surgical specimens in node-negative breast cancers.

Secondly, the degree of interlaboratory variation regarding Ki67 assessments remains largely unknown. Two previous studies found that while intralaboratory reproducibility is good, only moderate interlaboratory agreement is achieved [7,8]. Because TMAs were used in both of these studies, outcomes might even underestimate the 'real world' interlaboratory variation that would occur when whole sections are used. More data is therefore needed on the reproducibility of Ki67 in well-defined clinical cohorts while using modern-day IHC methods on whole slides. For this purpose, we

also investigated the reproducibility of Ki67 assays between two different reference laboratories (namely the Netherlands Cancer Institute – Antoni van Leeuwenhoek and the Leiden University Medical Center) and between multiple observers from these two centers.

Lastly, the Ki67 LI has been advocated as an alternative to performing gene-signature assays [9,10] for prognostication in certain clinical scenarios. Although this data is mostly based on the correlation of this parameter with the 21-gene recurrence score, less data is available on the correlation between this parameter and the 70-gene signature.

## Methods

### *Patient population*

For this retrospective study, clinical and pathological data were used from the first consecutive 105 patients of the MicroarRAy PrognoSTics in Breast CancER (RASTER) study (ISRCTN71917916). Details on the patient series of the RASTER study and the 70-gene recurrence score have been reported previously [11,12]. Patients were diagnosed with node-negative breast cancer from 2004 to 2006 at 16 Dutch hospitals. For each tumor in this study, one formalin-fixed paraffin-embedded (FFPE) tumor block containing a representative part of the tumor was collected at The Netherlands Cancer Institute. Institutional approval for the RASTER study was obtained from the Institutional Review Board of The Netherlands Cancer Institute [11].



### *Ki67 samples*

From each FFPE tumor block, two tissue microarrays (TMAs) were constructed by transferring random tissue cylinders of 0.6mm from within the tumor area (six cylinders per tumor, three cylinders per TMA) using a tissue arrayer (Beecher Instruments, Sun Prairie, WI, USA). Sections of the paraffin blocks from all tumors and the TMAs were cut at 4µm thickness and mounted. At the NKI-AVL, slides were baked at 56°C, and then stained with the Immunologic Autostainer 480 (Labvision, Fremont, California, USA) using MIB-1 (dilution 1:100). At the LUMC, immunohistochemistry was performed with the Dako Link 48 autostainer (Dako, Glostrup, Denmark) using the MIB-1 antibody (dilution 1:200). Appropriate controls were used throughout. All TMA cores and whole-slides were reviewed for adequacy. Cores or whole-slides with high background staining, absence of tumor or loss of tissue cores were deemed unsuitable, and were excluded from further analysis.

### *Ki67 evaluation*

All surgical specimen samples were independently assessed by 6 blinded observers, including 4 dedicated breast pathologists (NKI-AVL: J.W., E.G., J.S.; LUMC: V.S.). The TMAs were independently assessed by 2 blinded observers (E.G. and J.S.). The ratio of Ki67-positive cells was estimated with 5% accuracy; therefore, only values ending with 5 or 0 were recorded. Each observer was asked to use his or her daily evaluation approach to quantify the proportion of Ki67-positive cells and to perform evaluation of all cases. A cut-off value of 15% for positive marker status was used. For assessing the Ki67 LI concordance between CNBs and resection specimens, we regarded the TMAs as virtual biopsies to investigate the possible influence of tumor heterogeneity on Ki67 results. For these TMA cores, the Ki67 indices for the 3 individual cores within each triplet were averaged to determine the mean Ki67 LI CNB score (hereinafter referred to as CNB mean) for that particular virtual biopsy. Secondly, the maximum CNB Ki67 LI (referred to as CNB max) consisted of the highest Ki67 LI score of the three cores. Both these scores were compared with the resection specimen and each other.

### *Statistical design*

The level of agreement was expressed by means of Cohen's $\kappa$. The $\kappa$ values were interpreted as reflecting slight (0-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80) and almost prefect (> 0.80) agreement between observations according to Landis and Koch [13]. Analyses were carried out using the IBM SPSS Statistics package version 20.0.0 (SPSS Inc, Chicago, IL). Two-sided $p$ values of $< 0.05$ were considered statistically significant. The correlation between the Ki67 LI and the 70-gene signature and the tumor grade were assessed by calculating chi-squared tests.

## Results

### *Patient characteristics*

A total number of 105 invasive breast cancer patients were included. The median age was 50 years of age (range of 31-61 years old). Most patients (82%) received breast conserving surgery as primary local treatment. Three-fourths of the tumors were smaller than 20 mm (pT1). Samples included 86 invasive ductal carcinomas (82%), 13 invasive lobular carcinomas (12%), and 6 histologic subtypes of invasive carcinoma (mixed, mucinous, tubular, or with medullary features) (6%). Most tumors were ER-positive (84%), PgR-positive (68%), and HER2-negative (75%). Half of the carcinomas were

considered low-risk (49%) based on the 70-gene signature result. Patient and tumor characteristics are presented in table 1.

### *Agreement between core needle biopsies and surgical specimens*

The CNB scores of the TMAs were considered as independent cases, and were matched with the scores of the single corresponding surgical specimen. Two independent observers scored the CNB mean and CNB max which were compared with the Ki67 LI assessed in the matched surgical specimens. The overall discordance rate between CNB max and surgical specimen was 18.7% (95% CI 15-23), with a $\kappa$ value of 0.61. For CNB mean and surgical specimen the discordance rate was 19.2% (95% CI 15-23), with a $\kappa$ value of 0.59. The mean tumor size in the concordant group was 16.93 mm versus 17.83 mm in the discordant group ($P$=0.35).

The evaluations of observer A showed a discordance rate for CNB max of 16.8% (95% CI 11-22) and for CNB mean of 20.1% (95% CI 14-26). The $\kappa$ values were 0.64 and 0.56, respectively. The evaluations of observer B showed a discordance rate for CNB max of 20.6% (95% CI 15-27) and for CNB mean of 18.3% (95% CI 13-24). The $\kappa$ values were 0.58 and 0.62, respectively. There was no significant difference between the values of CNB max and CNB mean for observer A ($P$=0.21) and observer B ($P$=0.48) (table 2).



### *Interobserver agreement*

For determining the interobserver agreement of Ki67 LI assays when these are performed on surgical specimens, 6 independent observers evaluated 105 surgical specimens for Ki67 LI. Figure 1 shows the distribution of Ki67 LI based on the 15%-cutoff. The agreement between observer 1 and the other observers was considered substantial (0.73; 0.84; 0.64; 0.69; 0.67).

### *Agreement between laboratories for surgical specimens*

Table 4 shows the interlaboratory agreement of surgical specimens. The discordance rate was 12.3% (95% CI 6-19), with a $\kappa$ value of 0.74. The mean Ki67 score of the 13 discordant cases in laboratory 1 was 16.92 (range 15-20). In laboratory 2 the mean Ki67 score of the 13 discordant cases was 9.62 (range 5-10). Table 5 shows the 13 discordant cases and their characteristics. All the 13 surgical specimens were assessed as Ki67 LI $\geq$ 15 after staining in laboratory 1 and Ki67 LI $\leq$ 15 after staining in laboratory 2.

*Agreement between Ki67, tumor grade, and 70-gene signature*

Tables 6 to 8 show the agreement between the Ki67 labelling index, tumor grade and the 70-gene signature. As shown in table 6, the agreement for Ki67 Li and tumor grade was not significant ($P$=0.25). Table 7 shows no significant agreement for Ki67 Li and tumor grade ($P$=0.77). The agreement for tumor grade and 70-gene signature, as shown in table 8 was highly significant ($P < 0.01$).

## Discussion

According to current treatment guidelines, Ki67 LI combined with ER, PgR, and HER2 can be used to assign surrogate molecular subtypes, and may influence the decision whether a patient is advised to undergo adjuvant chemotherapy [14,15]. Unlike ER and PgR, there is no established cut-off value for classifying Ki67 as high or low. A Ki67 LI cut-off value of ≥ 14 was proposed for recommending adjuvant chemotherapy in endocrine-responsive breast carcinomas [4]. Cserni et al. stated that a cut-off should probably be an inclusive or non-inclusive number ending with 5 or 0 (like 5% or 10%), or more preferably ending with 0 (like 10% or 20%) [16]. Considering the need for reliable and reproducible breast cancer diagnostics, Ki67 assays must be subjected to stringent quality control. Several factors can hinder testing reproducibility including tissue fixation, choice of tissue (CNB vs surgical specimen), IHC protocols, and staining evaluation among others [17,18]. In this study, the influence of tumor heterogeneity, interobserver variation, and interlaboratory variation for Ki67 scoring were evaluated.

Regarding the concordance between CNB and resection specimen, moderate to substantial concordance was reached. Reliance on CNB for determining Ki67 status leads however to a high rate of both false-negative and false-positive test results when considering the resection specimen as the gold standard. In other studies, the discordance between CNB and surgical specimen for Ki67 LI was found to range from 14-21%, which is in line with our results [19-22].

Secondly, these discordances also occurred during scoring by two independent observers when using two different scoring methodologies. CNB max or CNB mean showed no significant difference which indicates that choice of the highest score or the mean score of multiple CNBs has no influence on the accordance with the surgical specimen. Therefore, hotspots do not seem to play a decisive role in Ki67 scoring of CNBs.

It should be noted that the different scores between CNBs and resection specimens might also have to do with intra-observer discordances, as both observers scored the CNB and resection specimens separately and in a blinded manner. Differences in scoring modalities between observers can have a substantial effect on testing reproducibility [23]. Historically, manual counting of a certain number of nuclei on high-power fields has been used to determine the Ki67 labelling index. This method is labor-intensive and therefore not often used in routine practice. Alternatively, some pathologists choose to estimate the labelling index by scanning the slide, so-called 'eyeballing'. In experienced eyes, this approach may be sufficient to separate the obvious Ki67-high cases from the lower cases, but it is criticized for a lack of precision and reproducibility. Published literature has reported conflicting reports whether thorough counting as opposed to eyeballing improves scoring reproducibility [8,24]. Eyeballing is typically done at a smaller magnification than counting, making it easier to integrate slight regional variations, and score more consistent average values. Our study confirmed that interobserver variability could have a significant influence on testing reproducibility in daily pathology practice even when asked to score a selected tumor area (TMA cores in this study).

Thirdly, we examined the concordance between Ki67 assays assessed over two different testing centers, both performing IHC procedures in completely independent manner on whole slides. An interlaboratory comparison tests both the uniformity in IHC procedures as well as scoring evaluation. Remarkably, the discordance between the two centers was in a similar range as the interobserver variability on selected tissues. This suggests that introducing separate IHC procedures does not introduce a significant increase in testing discordance. This is further supported by the fact that the mean scores for all discordant cases were closely around the cut-off point. No discordant cases contained high percentage discordances regarding Ki67 labelling indices (as would be expected in the cases of unreliable IHC procedures). These data suggest that using modern-day IHC with autostainers can lead to reproducible IHC results over multiple centers. Unfortunately, results from our study support results by other authors that manual Ki67 scoring leads to relatively high interobserver discordance, even among dedicated breast pathologists. A solution might be to introduce automated Ki67 scoring in clinical practice. Such a solution has been shown to be feasible in a study performed as part of the GeparTrio trial and is already commonplace in some institutions [25]. Automated analysis might not be necessary in cases with a high Ki67 labelling index (> 35% for instance), but might be used in cases where the Ki67 score is closer to the cut-off.

Finally, we examined the correlation between Ki67 and the well-established tumor grade. No correlation was shown. Also, for the relatively new but widely used and accepted 70-gene signature [26], no correlation was shown with Ki67 Li. Therefore, we consider it highly questionable if Ki67 can be used as decisive marker for guiding systemic treatment strategies.

Our study has some limitations. First, this was a retrospective evaluation of data collected from a dataset; therefore, this study suffers from the bias associated with any retrospective study [27]. Second, the International Ki67 in Breast Cancer Working Group mentions anecdotal evidence that Ki67 scores are generally lower on TMAs [28]. This may be viewed as a methodological disadvantage. Lastly, some have suggested that CNB have superior fixation and might therefore also be superior for detecting biomarkers. This could not be tested in our current study as the biopsies were virtually created with the use of TMAs.

We conclude that our data indicate that both tumor heterogeneity and interobserver scoring variation of Ki67 LI may have a significant influence on treatment advice. Also, Ki67 shows poor correlation with tumor grade and 70-gene signature result. Therefore, we cannot recommend the routine use of this marker to guide decisions regarding (neo)adjuvant treatment.

## References

1. Van Dierendonck et al.: Cell-cycle-related staining patterns of anti-proliferating cell nuclear antigen monoclonal antibodies. Comparison with BrdUrd labeling and Ki-67 staining. Am. J. Pathol. 138, 1165–1172 (1991)
2. Sørlie T et al.: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc. Natl. Acad. Sci. U.S.A. 98, 10869–10874 (2001)
3. Cheang MCU et al.: Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J. Natl. Cancer Inst. 101, 736–750 (2009)
4. Senkus E. et al: Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of oncology: official journal of the European Society for Medical Oncology / ESMO 24 Suppl 6, vi7–23 (2013)
5. Vörös A 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 80, 111–118 (2013)
6. Dekker TJA et al.: Quantitative assessment of lymph vascular space invasion (LVSI) provides important prognostic information in node-negative breast cancer. Ann. Oncol. 24, 2994–2998 (2013)
7. Mengel M et al.: Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. J. Pathol. 198, 292–299 (2002)
8. Polley, MYC et al.: An international Ki67 reproducibility study. J. Natl. Cancer Inst. 105, 1897–1906 (2013)
9. Allison KH et al.: Routine pathologic parameters can predict Oncotype DX recurrence scores in subsets of ER positive patients: who does not always need testing? Breast Cancer Res Treat 131, 413–424 (2012)
10. Geradts J et al.: The oncotype DX recurrence score is correlated with a composite index including routinely reported pathobiologic features. Cancer Invest. 28, 969–977 (2010)
11. Bueno-de-Mesquita JM et al.: Use of 70-gene signature to predict prognosis of patients with node-negative breast cancer: a prospective community-based feasibility study (RASTER). Lancet Oncol. 8, 1079–1087 (2007)
12. Drukker, CA et al: A prospective evaluation of a breast cancer prognosis signature in the observational RASTER study. Int. J. Cancer 133, 929–936 (2013)
13. Landis JR et al.: The measurement of observer agreement for categorical data. Biometrics 33, 159–174 (1977)
14. Luporsi E et al.: 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 132, 895–915 (2012)

15. De Azambuja E et al.: Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br. J. Cancer 96, 1504–1513 (2007)
16. Cserni G et al.: Distribution pattern of the Ki67 labelling index in breast cancer and its implications for choosing cut-off values. Breast 23, 259–263 (2014)
17. Lindboe CF et al.: Comparison of Ki-67 equivalent antibodies. J. Clin. Pathol. 55, 467–471 (2002)
18. Arima N et al.: The importance of tissue handling of surgically removed breast cancer for an accurate assessment of the Ki-67 index. J. Clin. Pathol. 69, 255–259 (2016)
19. Kalkman S et al.: Brief fixation does not hamper the reliability of Ki67 analysis in breast cancer core-needle biopsies: a double-centre study. Histopathology 66, 380–387 (2015)
20. Romero Q et al.: Ki67 proliferation in core biopsies versus surgical samples – a model for neo-adjuvant breast cancer studies. BMC Cancer 11, 341 (2011)
21. Greer LT et al.: Does breast tumor heterogeneity necessitate further immunohistochemical staining on surgical specimens? J. Am. Coll. Surg. 216, 239–251 (2013)
22. Chen X et al.: Preoperative core needle biopsy is accurate in determining molecular subtypes in invasive breast cancer. BMC Cancer 13, 390 (2013)
23. Varga Z et al.: Standardization for Ki-67 assessment in moderately differentiated breast cancer. A retrospective analysis of the SAKK 28/12 study. PLoS ONE 10, e0123435 (2015)
24. Varga Z 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 7, e37379 (2012)
25. Klauschen F et al.: Standardized Ki67 Diagnostics Using Automated Scoring – Clinical Validation in the GeparTrio Breast Cancer Study. Clin. Cancer Res. 21, 3651–3657 (2015)
26. Cardoso F et al.: 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. N. Engl. J. Med. 375, 717–729 (2016)
27. Niikura N et al.: Assessment of the Ki67 labeling index: a Japanese validation ring study. Breast Cancer (2014). doi:10.1007/s12282-014-0536-0
28. Dowsett M et al.: Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. Journal of the National Cancer Institute 103, 1656–1664 (2011)

Table 1. Baseline patient and tumor characteristics.

| Characteristic | No. | (%) |
|---|---|---|
| **No. of patients** | **105** | |
| **Median age at diagnosis, years** | **50 (31-61)** | |
| Lumpectomy | 86 | (82) |
| Mastectomy (+/- reconstruction) | 19 | (18) |
| pT1 (≤ 20) | 79 | (75) |
| pT2 (> 20-50) | 28 | (25) |
| pT3 (> 50) | 0 | |
| Ductal carcinoma | 86 | (82) |
| Lobular carcinoma | 13 | (12) |
| Other histological type | 6 | (6) |
| **Histological grade** | | |
| Grade 1 (good) | 21 | (22) |
| Grade 2 (intermediate) | 55 | (52) |
| Grade 3 (poor) | 28 | (27) |
| ER-negative | 17 | (16) |
| ER-positive | 88 | (84) |
| PR-negative | 34 | (32) |
| PR-positive | 71 | (68) |
| HER2-negative | 79 | (75) |
| HER2-positive | 10 | (10) |
| *Missing* | 16 | (15) |
| **70-gene signature result** | | |
| Low risk | 51 | (49) |
| High risk | 54 | (51) |

*Abbreviations*: ER: estrogen receptor; PR: progesterone receptor; HER2: human epidermal growth factor 2.

**Table 2.** Discordance between core needle biopsy and surgical specimen for Ki67 result.

| | | Surgical specimen | | | | Discordance rate | | |
|---|---|---|---|---|---|---|---|---|
| | | Ki67 LI | < 15% | ≥ 15% | Total (%) | (95% CI) | Kappa | *P*-value* |
| Observer A | CNB max | < 15% | 101 | 26 | 127 (69) | 16.8% (11-22) | 0.642 | 0.210 |
| | | ≥ 15% | 5 | 52 | 57 (31) | | | |
| | | Total (%) | 106 (58) | 78 (42) | 184 (100) | | | |
| | CNB mean | | | | | | | |
| | | < 15% | 106 | 37 | 143 (78) | 20.1% (14-26) | 0.561 | |
| | | ≥ 15% | 0 | 41 | 41 (22) | | | |
| | | Total (%) | 106 (58) | 78 (42) | 184 (100) | | | |
| | | | | | | | | 0.481 |
| Observer B | | < 15% | 82 | 15 | 97 (54) | 20.6% (15-27) | 0.584 | |
| | CNB max | ≥ 15% | 22 | 61 | 83 (46) | | | |
| | | Total (%) | 104 (58) | 76 (42) | 180 (100) | | | |
| | CNB mean | < 15% | 93 | 22 | 117 (65) | 18.3% (13-24) | 0.617 | |
| | | ≥ 15% | 11 | 54 | 65 (35) | | | |
| | | Total (%) | 104 (58) | 76 (42) | 180 (100) | | | |

*Abbreviations*: Ki67 LI: Ki67 labelling index; CNB: core needle biopsy; CNB max: maximum score out of 3 core biopsies;
CNB mean: mean score out of 3 core biopsies; CI: confidence interval
* *P*-values are based on the McNemar test (related samples)

**Table 3.** Discordance between observers for core needle biopsy Ki67 result.

| | | Observer B CNB max | | | | Discordance rate | |
|---|---|---|---|---|---|---|---|
| | | Ki67 LI | < 15% | ≥ 15% | Total (%) | (95% CI) | Kappa |
| Observer A | CNB max | < 15% | 97 | 28 | 125 (69) | 16.0% (11-21) | 0.669 |
| | | ≥ 15% | 1 | 55 | 56 (31) | | |
| | | Total (%) | 98 (54) | 83 (46) | 181 (100) | | |
| | | Observer B CNB mean | | | | | |
| | | Ki67 LI | < 15% | ≥ 15% | Total (%) | | |
| Observer A | CNB mean | < 15% | 116 | 24 | 140 (77) | 13.3% (8-18) | 0.686 |
| | | ≥ 15% | 0 | 41 | 41 (31) | | |
| | | Total (%) | 116 (64) | 65 (36) | 181 (100) | | |

*Abbreviations*: Ki67 LI: Ki67 labelling index; CNB: core needle biopsy; CNB max: maximum score out of 3 core biopsies; CNB mean: mean score out of 3 core biopsies

**Table 4.** Concordance between laboratories for Ki67 result.

| | | Laboratory 1 | | | Discordance rate (95% CI) | Kappa |
|---|---|---|---|---|---|---|
| Laboratory 2 | Ki67 LI | < 15% | ≥ 15% | Total (%) | | |
| | < 15% | 60 | 0 | 60 (57) | 12.3% (6-19) | 0.738 |
| | ≥ 15% | 13 | 32 | 45 (43) | | |
| | Total (%) | 73 (70) | 32 (30) | 105 (100) | | |

*Abbreviations*: CNB: core needle biopsy; WS: whole slide; Ki67 LI: Ki67 labelling index

**Table 5.** Ki67 discordant cases in the interlaboratory comparison.

| Case | Ki67 LI NKI | Ki67 LI LUMC | 70-gene signature | Tumor size | ER | PR | HER2 | Histology |
|---|---|---|---|---|---|---|---|---|
| 1 | 20 | 10 | High | 10 | 1 | 1 | 1 | Ductal |
| 2 | 20 | 10 | High | 31 | 0 | 0 | 1 | Ductal |
| 3 | 15 | 10 | High | 15 | 0 | 0 | 0 | Ductal |
| 4 | 20 | 10 | Low | 15 | 1 | 0 | 0 | Ductal |
| 5 | 20 | 10 | Low | 50 | 1 | 1 | 9 | Ductal |
| 6 | 15 | 10 | High | 19 | 1 | 0 | 0 | Mucinous |
| 7 | 15 | 10 | Low | 13 | 1 | 1 | 0 | Lobular |
| 8 | 15 | 10 | High | 11 | 1 | 0 | 9 | Ductal |
| 9 | 15 | 10 | Low | 19 | 1 | 1 | 0 | Ductal |
| 10 | 15 | 5 | High | 15 | 1 | 1 | 0 | Ductal |
| 11 | 15 | 10 | High | 18 | 1 | 0 | 1 | Ductal |
| 12 | 20 | 10 | Low | 15 | 0 | 1 | 0 | Ductal |
| 13 | 15 | 10 | High | 12 | 1 | 1 | 0 | Ductolobular |

*Abbreviations:* Ki67 LI: Ki67 labelling index

**Table 6.** Agreement between tumor grade and Ki67 result.

| | | Ki67 LI | | | Pearson Chi-square |
|---|---|---|---|---|---|
| | | < 15% | ≥ 15% | Total (%) | |
| Tumor grade | 1 (Good) | 16 | 6 | 22 (21) | |
| | 2 (Intermediate) | 29 | 26 | 55 (52) | *P*=0.25 |
| | 3 (Poor) | 15 | 13 | 28 (27) | |
| | Total (%) | 60 (57) | 45 (43) | 105 (100) | |

*Abbreviations*: Ki-67 labelling index

**Table 7.** Agreement between 70-gene signature and Ki67 result.

| | | Ki67 LI | | | Pearson Chi-square |
|---|---|---|---|---|---|
| | | < 15% | ≥ 15% | Total (%) | |
| 70-gene signature | Low risk | 31 | 20 | 33 (31) | |
| | High risk | 29 | 25 | 54 (51) | *P*=0.77 |
| | Total (%) | 60 (57) | 45 (43) | 105 (100) | |

**Table 8.** Agreement between 70-gene signature and tumor grade.

| | | Tumor grade | | | Pearson Chi-square | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | Total (%) | |
| | | (Good) | (Intermediate) | (Poor) | | |
| 70-gene signature | Low risk | 17 | 31 | 3 | 33 (31) | |
| | High risk | 5 | 24 | 25 | 54 (51) | *P* < 0.01 |
| | | 22 (21) | 55 (52) | 28 (27) | 105 (100) | |

Figure 1. Distribution of Ki67 result across different observers and their agreement.