Good staining quality ensuring the reproducibility of Ki67 assessment

Yeh-Han Wang,[1,2,3] Chiung-Ru Lai,[4,5] Huang-Chun Lien,[6,7] Chih-Yi Hsu [3,4,5]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jclinpath-2019-206205).

[1]Department of Anatomic Pathology, Taipei Institute of Pathology, Taipei, Taiwan
[2]Institute of Public Health, National Yang-Ming University, Taipei, Taiwan
[3]College of Nursing, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan
[4]Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, Taipei, Taiwan
[5]School of Medicine, National Yang-Ming University, Taipei, Taiwan
[6]Graduate Institute of Pathology, National Taiwan University, Taipei, Taiwan
[7]Department of Pathology, National Taiwan University Hospital, National Taiwan University, Taipei, Taiwan

Correspondence to
Dr Chih-Yi Hsu, Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, Taipei 112, Taiwan; cyhsu@vghtpe.gov.tw

Received 5 September 2019
Revised 29 October 2019
Accepted 16 November 2019
Published Online First 3 December 2019

To cite: Wang Y-H, Lai C-R, Lien H-C, *et al*. *J Clin Pathol* 2020;**73**:413–417.

ABSTRACT

Aims Although Ki67 labelling index (LI) is a prognostic and predictive marker in breast cancer, its accuracy and reproducibility must be validated before its clinical application. We aimed to evaluate the agreement of Ki67 LI in clinical practice in Taiwan.

Methods We conducted a Ki67 immunohistochemistry (IHC) proficiency test. The participants performed the Ki67 IHC test and measured the Ki67 LI of 10 cases of breast cancer tissue on a microarray slide. The staining quality was centrally reviewed based on the Ki67 staining of the tonsil surface epithelium.

Results Ki67 staining and counting methods are diverse in Taiwan. The reproducibility of Ki67 LI was poor to good (intraclass correlation coefficient: 0.581, 95% CI 0.354 to 0.802). The reproducibility and agreement in the high staining quality group were significantly higher than those in the low staining quality group. The majority of the Ki67 LIs derived from the low staining quality group were underestimated. Different counting methods did not reveal significant differences when determining Ki67 LI with microarray sections.

Conclusions We suggest using the surface epithelium of the tonsil as external control and achieving optimal staining results that consist of a high positive parabasal layer, a low positive intermediate layer and a negative superficial layer. Good Ki67 staining quality can minimise the staining variations among different laboratories, and it is essential for the reproducibility of Ki67 LI.

INTRODUCTION

As the Ki67 labelling index (LI) has been considered a prognostic or predictive factor in breast cancer, its accuracy and reproducibility remain debatable in clinical practice due to lack of standardisation. Ki67 is not included in the biomarkers suggested by the American Society of Clinical Oncology and National Comprehensive Cancer Network guidelines to guide decisions regarding therapy for women with breast cancer.[1 2] Additionally, the cut-off point for a low Ki67 index changed from 15% (2009),[3] 14% (2011)[4] or 20% (2013)[5] to 20%–29% (2015)[6] in the St. Gallen International Expert Consensus. Consequently, the consensus panel raised an issue regarding the reproducibility of Ki67 LI and its application in clinical decision-making in 2017, rather than providing specific cut-offs.[7] Although the International Ki67 in Breast Cancer Working Group has provided certain recommendations regarding the use of Ki67 immunohistochemistry (IHC),[8–10] larger scale analytical and clinical validations are warranted before Ki67 LI can be employed in clinical applications.

The variations in Ki67 LI may result from preanalytical (specimen handling) and analytical (staining and counting) factors.[8] The analytic factors can be evaluated using multisite, interlaboratory comparison exercises, so-called proficiency tests (PT). For quality assurance and improvement, the Committee of Breast Pathology of Taiwan Society of Pathology (TSP) conducted the first Ki67 IHC PT to evaluate the accuracy and reproducibility of Ki67 immunostaining in 2018.

MATERIALS AND METHODS

The study was conducted in accordance with local ethical regulations, and no Institutional Review Board approval was required. A tissue microarray (TMA) with a core diameter of 2 mm was constructed by acquiring 10 cases of breast cancer from the surgical pathology archive of the Taipei Veteran General Hospital (TVGH), including 9 invasive carcinomas of no special type and 1 invasive lobular carcinoma. All the breast specimens processed in TVGH abided by American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) recommendations in regard to specimen handling for optimisation of preanalytic factors. In addition, a small piece of tonsil tissue that served as the external control was included.

The 2018 Ki67 IHC PT was freely provided as a selective testing item to the institutions that have already registered IHC items for breast cancer (eg, oestrogen receptor, progesterone receptor or human epidermal receptor 2). One unstained slide of the TMA sectioned at 4 µm in thickness was delivered to the participating laboratories. The presence of adequate cancer cells within the test slides was confirmed by the PT committee using microscopic examination of the first and the last section of the TMA with H&E staining.

The participants performed Ki67 IHC staining and Ki67 scoring independently, without any private communication. The results of Ki67 LI of each core were submitted online no later than early September 2018. The staining and counting information was collected, including staining methods and protocols, antibody used, counting area selection and measuring methods. The recommended answer form was a specific number of LI rather than a numerical range or three-tier grading. After completion of the online submission, the participants were requested to send the TMA slide back to the TSP office for a quality review.

Two pathologists of the Committee of Breast Pathology were responsible for the quality review of the Ki67 IHC PT slides that were sent back. The evaluation was based on the staining pattern

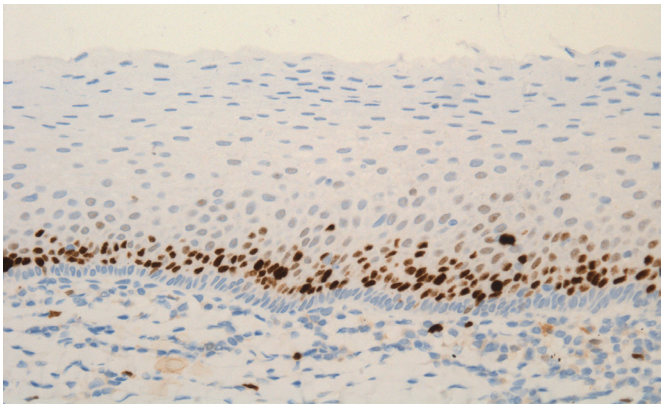

Figure 1 Ideal Ki67 staining pattern of the surface epithelium of the tonsil tissue.

of the surface epithelium of the tonsil tissue, which served as the external control in every test slide.[11] The scoring criteria were determined by the presence of the following: (1) high positive zone (the parabasal layer); (2) low positive zone (the intermediate layer); (3) negative zone (the superficial layer); (4) adequacy of positivity (a band-like pattern at parabasal layer and faint nuclear staining in >50% cells of the intermediate layer) and (5) ideal counter stain (figure 1). Each criterion was counted for one score in the quality review, and the highest score was 5. The final score in the staining quality was the average of that given by the two reviewers. The slides with a quality score of ≥4 were considered the high staining quality group, whereas those with a quality score of ≤2 were considered the low staining quality group. Subsequently, the slides with the highest quality score were manually counted, and the average results of Ki67 LI of these slides were provided as the reference data in the study. Manual counting was performed by calculating the Ki67 positive cells among 500–1000 tumour cells in the hotspot areas.

The responses of Ki67 LI with numerical data were collected for the analyses. When the LI result was given in a numerical range, the median of the range was used for the following analyses. The Fisher's exact or χ^2 test was used to compare the proportions of categorical variables. Continuous variables were analysed using the Mann–Whitney U test or the Kruskal–Wallis test. P values were derived from two-tailed tests, and $p<0.05$ was considered statistically significant.

To evaluate the interobserver reproducibility, the intraclass correlation coefficient (ICC) with a 95% CI was calculated using a two-way random-effects model. The ICC ranged between 0 and 1, with values closer to 1 representing better reproducibility. An ICC of <0.50 indicated poor, 0.50–0.75 indicated moderate, >0.75–0.90 indicated good and >0.90 indicated excellent reproducibility.[12] To satisfy the random-effects model assumptions of normality for ICC analyses, the Ki67 data were converted to a logarithmic scale by adding 0.1% and applying a log base 2 transformation. The differences of ICCs were tested by t-test on 1000 bootstrap replicates. Laboratories failed to complete all the LI measurement of the 10 cores in the TMA were excluded from the ICC analyses.

Ki67 LI values were categorised as low (<14%), intermediate (14%–29%) and high (>29%). Pairwise comparison of the LI categories between the participants and the reference result was analysed using kappa statistics. The kappa value ranged between 0 and 1, with values closer to 1 representing better agreement.

Table 1 Survey of staining and counting methods

Factors	N	%
Autostainer		
No	8	15.4
Yes	44	84.6
Antibody clones		
GM010	18	34.6
MIB-1	17	32.7
30-Sep	7	13.5
SP6	5	9.6
Others	5	9.6
Evaluation tools		
Visual estimate	47	90.4
Image analysis*	2	3.8
Manual counting	3	5.8
Cell number evaluated		
<500	27	51.9
≥500	25	48.1
Area selection		
Hotspot	30	57.7
Average	18	34.6
No consensus	4	7.7

*Using ImmunoRatio.[21]

RESULTS

Survey of staining and counting methods

A total of 52 institutions registered for the 2018 Ki67 LI PT, including 50 hospital-based pathology laboratories and 2 commercial laboratories. Autostainers were widely used. GM010 and MIB-1 were the commonly used clones of the Ki67 antibodies. In total, 47 (90.4%) laboratories evaluated LI with visual estimate and 30 (57.7%) laboratories counted the LI of the hotspot areas. The details of the brief survey are listed in the online supplementary table and summarised in table 1.

Review of staining quality

Of the 52 laboratories, 46 (88.5%) had sent their test slides back to the TSP office, and the average staining quality score was 3.1 (95% CI 2.7 to 3.4). The high staining quality group (score ≥4) consisted of 15 (32.6%) out of the 46 laboratories, while 11 (23.9%) laboratories were classified under the low staining quality group (score ≤2). The quality scores were not significantly related to the use of autostainers and antibody clones (table 2). A total of four (8.7%) slides were given the highest quality score after review, and the mean LI of central counting of each core was regarded as the reference result of each core. According to the reference results, three cores (#3, #5 and #8) were categorised as low LI and three cores (#1, #4 and #9) were categorised as high LI. The remaining four cores (#2, #6, #7 and #10) were intermediate in LI.

Distribution of Ki67 Li among participants

The distribution of the responded Ki67 LI derived from participants according to staining and counting methods are listed in table 3. The Ki67 LIs between the staining quality groups were significantly different, and Ki67 LI values of the high staining quality group were greater than those of the intermediate and low staining quality groups. Additionally, Ki67 LIs rendered using clone 30–9 were greater than those rendered using clone SP6 and clone GM010 ($p=0.004$ and $p=0.015$). Different evaluation tools and cell number evaluated did not reveal significant

Table 2 Staining quality scores by staining method

Factors	N	Median (range)	Mean±SD	P value
Total	46	3.0 (1.0–5.0)	3.1±1.1	
Autostainer				0.521
No	7	3.0 (1.0–4.0)	2.8±1.1	
Yes	39	3.0 (1.0–5.0)	3.1±1.1	
Antibody clones				0.161
GM010	16	3.3 (1.0–4.5)	2.8±1.2	
MIB-1	16	3.0 (1.0–5.0)	3.1±1.2	
30-Sep	5	4.0 (3.0–5.0)	4.2±0.8	
SP6	5	3.0 (2.0–3.0)	2.7±0.4	
Others	4	2.5 (2.0–4.0)	2.8±0.9	

differences in Ki67 LI. As a 2 mm core might not provide enough choice for area selection, the difference between counting by hotspot and average methods was not significant. The reference results and the distribution of participants' results are plotted in figure 2. Most of the responded Ki67 LI values from the low staining quality group were underestimated, as the median value was smaller than the reference result observed in every core.

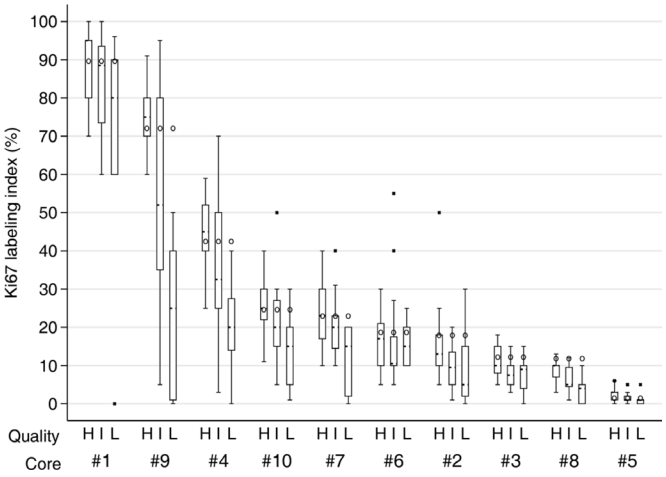

Figure 2 Distribution of the Ki67 LI grouped by staining quality. The boxes indicate the IQR, with adjacent values as whiskers, outlying values as x; median is indicated by the horizontal dotted line. Circles indicate the reference results of Ki67 LI (H, high; I, intermediate; L, low). LI, labelling index.

Reproducibility of Ki67 Li

The study excluded 10 participants who did not complete the Ki67 LI measurement of all the 10 cores from the ICC analyses. Of them, five, two and two were in the low, intermediate and high staining quality group, respectively, while one did not return the slide for quality review. Reproducibility of Ki67 LI among the PT participants was poor to good (ICC: 0.581, 95% CI 0.354 to 0.802). In the subgroup analyses, the reproducibility in the high staining quality group (ICC: 0.840, 95% CI 0.788 to 0.905) was significantly higher than that in the low staining quality group (ICC: 0.181, 95% CI 0.037 to 0.550, p<0.001). The ICC value of the high staining quality group was less than the ICC value (0.961, 95% CI 0.947 to 0.985, p<0.001) derived from the manual counting of the four slides of the highest quality score for the reference result.

Table 3 Comparison of the Ki67 LI grouped by staining and counting methods

Factors	Ki67 LI, median (IQR)	P value
Autostainer		0.412
No	15.0 (4.0–30.0)	
Yes	15.0 (7.0–33.0)	
Antibody clones		0.044*
GM010	15.0 (5.0–30.0)	
MIB-1	17.0 (7.0–35.0)	
30-Sep	20.0 (10.0–40.0)	
SP6	10.0 (5.0–30.0)	
Others	10.0 (10.0–30.0)	
Evaluation tools		0.635
Visual estimate	15.0 (6.0–31.0)	
Image analysis	17.5 (10.5–57.0)	
Manual counting	15.0 (5.0–29.0)	
Cell number evaluated		0.218
<500	15.0 (5.0–33.0)	
≥500	15.0 (7.0–30.0)	
Area selection		0.053
Hotspot	15.0 (8.0–32.0)	
Average	14.0 (5.0–30.0)	
Not specified	12.5 (5.0–40.0)	
Staining quality		<0.001*
Low	10.0 (2.0–20.0)	
Intermediate	15.0 (6.0–30.5)	
High	20.0 (10.0–47.5)	

*P<0.05
LI, labelling index.

Distribution and agreement of Ki67 Li categories

The distribution of Ki67 LI categories grouped by staining quality was significantly different (p=0.043, table 4). Low Ki67 LI (<14%) results were more frequently reported in the low staining quality group. Conversely, the proportion of high Ki67 LI (>29%) reported by the high staining quality group was 1.9 times that of the low staining quality group. The LI categories of the result of participants with the highest quality score were 100%, which agreed with the results of the central review of the same slide. The pairwise kappa values derived from the agreement of LI categories between the participants' result and reference result revealed a wide range (0.000–1.000). The kappa values of the high staining quality group (median: 0.697, IQR: 0.552–0.706) were significantly greater than those of the low staining quality group (median: 0.518, IQR: 0.333–0.697) (p=0.029). The kappa values of the intermediate staining quality group (median: 0.701, IQR: 0.420–0.706) were not significantly

Table 4 Frequency of the Ki67 LI categories stratified by staining quality

	Staining quality			
	Low	Intermediate	High	Total
# of slides	11	20	15	46
# of Ki67 LI results	96 (100%)	196 (100%)	148 (100%)	440 (100%)
Ki67 LI categories				
Low (<14%)	51 (53.1%)	93 (47.4%)	57 (38.5%)	201 (45.7%)
Intermediate (14%–29%)	26 (27.1%)	43 (21.9%)	35 (23.6%)	104 (23.6%)
High (>29%)	19 (19.8%)	60 (30.6%)	56 (37.8%)	135 (30.7%)

LI, labelling index.

different from those of the high and those of the low staining quality group (p=0.527 and p=0.137).

DISCUSSION

The PT surveyed both staining and counting factors and demonstrated an overview of clinical practice of Ki67 LI in breast pathology in Taiwan. Our results indicated that the staining quality plays a key role in the reproducibility and agreement of Ki67 LI in TMAs. Both the reproducibility and agreement of the high staining quality group were significantly higher than those of the low staining quality group. The majority of the Ki67 LIs derived from the low staining quality group were underestimated. With the experience from the PT, improvement of the staining quality of Ki67 would be a prior task to enhance the reliability of the Ki67 LI reporting in clinical practice.

Standardisation of Ki67 IHC staining is difficult due to the variation in the staining platforms, protocols and antibody clones between the laboratories.[13 14] Previous large-scale studies dealing with interlaboratory reproducibility of Ki67 staining revealed substantial variabilities.[15 16] The differences remained significant even laboratories using the same antibody and in the laboratories with prior participation in external quality assurance programme.[16] They suggested either standardisation of Ki67 LI determination or fully aware of lab-specific reference values.[15 16] In this study, neither the use of autostainers nor any Ki67 antibody clone could guarantee a high staining quality. As this study did not intend to establish a uniform staining protocol or to rank the clones of the antibody from the survey across laboratories, we instead recommend applying an identical standard on the quality assessment of the external control for optimisation of Ki67 staining quality. As we demonstrated in the quality review for PT, the surface epithelium of the tonsil tissue is a good external control tissue for Ki67 IHC. An optimal Ki67 IHC stain on the tonsillar epithelium should demonstrate strong positive staining over the parabasal layer with a band-like pattern and moderate or weak nuclear staining in >50% cells of the intermediate layer. The superficial layer should be negative without nonspecific staining in the background (figure 1). Irrespective of the platforms or antibody clones used, achieving optimal staining results could minimise variations result from staining methods.

The complexity of the counting method for Ki67 LI makes it difficult to implement; it includes evaluation tools, area selection and a count of the number of cells evaluated. Although the p value (0.053) of area selection was not significant in the present study, differences in Ki67 LI among counting methods did exist. The International Ki67 in Breast Cancer Working Group has conducted several studies on the standardisation problem and developed various standardisation methods.[9 10 17] The ICC of the present study (0.581, 95% CI 0.354 to 0.802) was similar to that obtained by counting locally stained slides in a previous international study using 1 mm core TMAs (0.59, 95% CI 0.37 to 0.68).[17] Our interlaboratory ICC (0.840, 95% CI 0.788 to 0.905) obtained by the high staining quality group was higher than the interlaboratory ICC obtained by counting central stained slides in the international study (0.71, 95% CI 0.47 to 0.78).[17] However, counting factors might not be well explored in the current study because we used a TMA with a core diameter of 2 mm, which facilitated area selection. Another study using a 1.4 mm core TMA also revealed a good interobserver reproducibility in measuring the same staining, but a high interlaboratory variability derived from staining.[15] Area selection of whole tissue sections in the clinical practice is more complex, and the assessment method may have direct influences on Ki67-LI and therapeutic decision making.[18–20] Although digital image analysis (ImmunoRatio)[21] was applied in only 3.8% of our participants and its benefits did not appear, the use of automated systems to measure Ki67 LI presents good interlaboratory reproducibility in some studies.[22–24] Furthermore, the use of automated systems is thought to reduce the time and labour required for manual counting.[22]

In addition to the usage of a TMA, there are other limitations to this study. One limitation is the study size based on 10 tissue samples. Increasing the number of samples will increase the workload for the participants. Therefore, we used 10 samples with an even distribution of low, medium and high Ki67 LI in an attempt to make a balance between the quality and quantity of the responses. Besides, the influences of tissue processing, fixation and other preanalytical factors are not tested in our study. Further, we are unable to comment on image analysis and automated systems because only two of our participants used image analysis as an assist tool.

In summary, the staining and counting methods of Ki67 are diverse in Taiwan. Our results revealed a poor-to-good reproducibility without standardisation of the staining and counting methods across laboratories. The staining quality was found to be essential for the reproducibility. We suggest using the surface epithelium of the tonsil as external control and achieving optimal staining results that consist of a high positive parabasal layer, a low positive intermediate layer and a negative superficial layer. Good Ki67 staining quality can minimise the staining variations among different laboratories.

Take home messages

- Ki67 staining and counting methods are diverse.
- The Ki67 labelling indexes (LIs) derived from the low staining quality group are usually underestimated.
- A good Ki67 staining quality is essential for the reproducibility of Ki67 LI.
- Using tonsil surface epithelium as control to achieve optimal staining that consists a high positive parabasal layer, a low positive intermediate layer and a negative superficial layer.

Handling editor Cheok Soon Lee.

Acknowledgements The authors thank Hui-Mai Chen and the Committee of Breast Pathology of Taiwan Society of Pathology for technical assistance. They also thank the Department of Biostatistics Task Force of Taipei Veterans General Hospital for statistical consultation.

Contributors Conceptualisation and design: Y-HW, C-RL, H-CL and C-YH. Data acquisition, analysis and interpretation: Y-HW and C-YH. Manuscript preparation, editing and approval: Y-HW, C-RL, H-CL and C-YH.

Funding This study was funded by Taipei Veterans General Hospital (V108C-161) and Taipei Institute of Pathology (TIP-107–001).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iD

Chih-Yi Hsu http://orcid.org/0000-0002-1667-4291

REFERENCES

1 Krop I, Ismaila N, Andre F, *et al*. Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer: American Society of clinical oncology clinical practice guideline focused update. *JCO* 2017;35:2838–47.

2 National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: breast cancer (version 3.2019), 2019. Available: https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf [Accessed Oct 27, 2019].
3 Goldhirsch A, Ingle JN, Gelber RD, *et al*. Thresholds for therapies: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2009. *Ann Oncol* 2009;20:1319–29.
4 Goldhirsch A, Wood WC, Coates AS, *et al*. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2011. *Ann Oncol* 2011;22:1736–47.
5 Goldhirsch A, Winer EP, Coates AS, *et al*. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2013. *Ann Oncol* 2013;24:2206–23.
6 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;26:1533–46.
7 Curigliano G, Burstein HJ, Winer EP, *et al*. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen international expert consensus conference on the primary therapy of early breast cancer 2017. *Ann Oncol* 2017;28:1700–12.
8 Dowsett M, Nielsen TO, A'Hern R, *et al*. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in breast cancer Working group. *J Natl Cancer Inst* 2011;103:1656–64.
9 Polley M-YC, Leung SCY, Gao D, *et al*. An international study to increase concordance in Ki67 scoring. *Mod Pathol* 2015;28:778–86.
10 Leung SCY, Nielsen TO, Zabaglo L, *et al*. Analytical validation of a standardized scoring protocol for Ki67: phase 3 of an international multicenter collaboration. *npj Breast Cancer* 2016;2.
11 Hsu C-Y, Yang C-F, Liao L-R, *et al*. Tonsil surface epithelium is ideal for monitoring Ki-67 immunohistochemical staining. *Histopathology* 2013;63:810–6.
12 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
13 Ács B, Kulka J, Kovács KA, *et al*. Comparison of 5 Ki-67 antibodies regarding reproducibility and capacity to predict prognosis in breast cancer: does the antibody matter? *Hum Pathol* 2017;65:31–40.
14 Jang MH, Kim HJ, Chung YR, *et al*. A comparison of Ki-67 counting methods in luminal breast cancer: the average method vs. the hot spot method. *PLoS One* 2017;12:e0172031.
15 Mengel M, von Wasielewski R, Wiese B, *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* 2002;198:292–9.
16 Focke CM, Bürger H, van Diest PJ, *et al*. Interlaboratory variability of Ki67 staining in breast cancer. *Eur J Cancer* 2017;84:219–27.
17 Polley M-YC, Leung SCY, McShane LM, *et al*. An international Ki67 reproducibility study. *J Natl Cancer Inst* 2013;105:1897–906.
18 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;46:1341–9.
19 Romero Q, Bendahl P-O, Fernö M, *et al*. A novel model for Ki67 assessment in breast cancer. *Diagn Pathol* 2014;9:118.
20 Focke CM, van Diest PJ, Decker T. St Gallen 2015 subtyping of luminal breast cancers: impact of different Ki67-based proliferation assessment methods. *Breast Cancer Res Treat* 2016;159:257–63.
21 Tuominen VJ, Ruotoistenmäki S, Viitanen A, *et al*. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res* 2010;12.
22 Koopman T, Buikema HJ, Hollema H, *et al*. Digital image analysis of Ki67 proliferation index in breast cancer using virtual dual staining on whole tissue sections: clinical validation and inter-platform agreement. *Breast Cancer Res Treat* 2018;169:33–42.
23 Acs B, Pelekanou V, Bai Y, *et al*. Ki67 reproducibility using digital image analysis: an inter-platform and inter-operator study. *Lab Invest* 2019;99:107–17.
24 Rimm DL, Leung SCY, McShane LM, *et al*. An international multicenter study to evaluate reproducibility of automated scoring for assessment of Ki67 in breast cancer. *Mod Pathol* 2019;32:59–69.