

A simple digital image analysis system for automated Ki67 assessment in primary breast cancer

Anastasia Alataki,^{1,2}  Lila Zabaglo,^{1,2} Holly Tovey,³ Andrew Dodson¹ & Mitch Dowsett^{1,2}

¹Ralph Lauren Centre for Breast Cancer Research, Royal Marsden Hospital and The Institute of Cancer Research, ²The Breast Cancer Now Toby Robins Research Centre, The Institute of Cancer Research, and ³Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, UK

Date of submission 13 November 2020

Accepted for publication 14 February 2021

Published online Article Accepted 15 February 2021

Alataki A, Zabaglo L, Tovey H, Dodson A & Dowsett M

(2021) *Histopathology* 79, 200–209. <https://doi.org/10.1111/his.14355>

A simple digital image analysis system for automated Ki67 assessment in primary breast cancer

Aims: Ki67 is a well-established immunohistochemical marker associated with cell proliferation that has prognostic and predictive value in breast cancer. Quantitative evaluation of Ki67 is traditionally performed by assessing stained tissue slides with light microscopy. Automated image analysis systems have become available and, if validated, could provide greater standardisation and improved precision of Ki67 scoring. Here, we aimed to evaluate the use of the Cognition Master Professional Suite (CogM) image analysis software, which is a simple system for scoring Ki67 in primary breast cancer samples.

Methods and results: Sections from 94 core-cut biopsies, 20 excision specimens and 29 pairs of core-cut biopsies and excision specimens were stained for Ki67 with MIB1 antibody and the Dako EnVision FLEX Detection System. Stained slides were scanned to

convert them to digital data. Computer-based Ki67 scoring was performed with CogM. Manual Ki67 scoring assessment was conducted on previously stained sections from the same biopsies with a clinically validated system that had been calibrated against the risk of recurrence. A high correlation between manual and digital scores was observed [$r_{\text{Cores}} = 0.92$, 95% confidence interval (CI) 0.87–0.94, $P < 0.0001$; $r_{\text{Excisions}} = 0.95$, 95% CI 0.86–0.98, $P < 0.0001$] and there was no significant bias between them ($P = 0.45$). There was also a high correlation of Ki67 scores between paired core-cut biopsies and excision specimens when CogM was used ($r = 0.78$, 95% CI 0.59–0.89, $P < 0.0001$).

Conclusions: CogM image analysis allows for standardised automated Ki67 scoring that accurately replicates previously clinically validated and calibrated manual scores.

Keywords: automated image analysis, biopsy, breast cancer, immunohistochemistry, Ki67 scoring

Introduction

Uncontrolled sustained proliferation is one of the fundamental traits of human cancer cells.¹ The proliferation biomarker Ki67 is expressed in dividing cells but is absent in resting cells, and it has been used extensively for comparing

proliferation rates between tumour samples in breast cancer.² Ki67 can potentially be used for predicting long-term outcome,^{3–6} possibly for predicting responsiveness to certain therapies, including chemotherapy and endocrine therapy,^{2,7–10} for estimating residual risk,¹¹ and for evaluating treatment efficacy on the basis of its

Address for correspondence: M Dowsett, The Ralph Lauren Centre for Breast Cancer Research, The Royal Marsden Foundation Trust, 203 Fulham Road, Chelsea, London, SW3 6JJ, UK. e-mail: mitchell.dowsett@jcr.ac.uk

© 2021 The Authors. *Histopathology* published by John Wiley & Sons Ltd.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

dynamic changes during neoadjuvant endocrine therapy.^{2,12}

Ki67 immunohistochemistry is the most widely used assay for the measurement of cell proliferation in breast cancer.^{13,14} Evaluation of this marker is traditionally performed by visually assessing stained tissue slides with light microscopy. Proper quantification of protein expression relies on the analyst's interpretation of the results, methodology, and quality of the tissue.¹⁵ Moreover, visual scoring is often subject to considerable intraobserver and interobserver variability, particularly in multicentre studies, contributing to a lack of consensus for Ki67 values and cut-offs.^{16,17} Unlike other immunohistochemical markers, such as oestrogen receptor, Ki67 has not yet been subject to widespread standardisation, and requires more accurate and well-defined scoring. This has limited its application in both diagnostic and research settings.^{2,18}

During the last decade, progress in technology allowed the development of software systems for automated scoring of immunohistochemical biomarker expression, which could potentially enhance precision and reliability while enabling greater workloads to be handled more quickly. In this context, digital image analysis may enable the standardisation of Ki67 quantification, for which a number of approaches have been reported.^{19–22} The Cognition Master Professional Suite (CogM) is a collection of straightforward image analysis software tools for the presentation, evaluation and analysis of digital histological slides. One of the CogM modules, Ki67 Quantifier, has been previously validated in a neoadjuvant breast cancer clinical trial as a computer-based approach for Ki67 scoring based on a cell detection method.^{22–24}

The image analysis must be applicable to different tissue sample types, such as core-cut biopsies and excision specimens, and this may be particularly relevant when different sample types are taken for longitudinal comparative purposes.^{25,26} The possibility that image analysis is sensitive to artefacts that may be caused by differences in quality of fixation between different sample types needs to be recognised and accommodated.

Here, a study was undertaken to evaluate the use of CogM and a modified staining protocol for Ki67 scoring in core-cut biopsies and excision specimens, and compare the results with those obtained with the same samples that were previously analysed by visual scoring. Importantly, the manual scoring and staining had been conducted with methodology on which the International Ki67 in Breast Cancer Working

Group Party (KiBCWG) had based its recommendations, and the scores derived had been calibrated against outcome in the PeriOperative Endocrine Therapy for Individualised Care (POETIC) adjuvant endocrine trial of nearly 4500 patients.²⁷ Given the demonstration in that trial that Ki67 analyses performed after 2 weeks of treatment with an aromatase inhibitor (AI) provided substantial additional prognostic information, scoring of such samples was included. Additionally, Ki67 scoring with CogM was compared between core-cut biopsies and excision specimens taken at the time of surgery from the same patients.

Materials and methods

TISSUE SAMPLES

The specimens used in the present study were derived from patients who took part in the POETIC trial (Trial Number CRUK/07/015)²⁸ and the POETIC Pilot study.²⁹ Ethical approval was provided by the London–South East Research Ethics Committee (REC reference 08/H1102/37), and all patients provided signed consent for the use of their tumour tissue for research purposes.

Archival formalin-fixed paraffin-embedded tissue blocks from four cohorts of postmenopausal patients with primary hormone-receptor positive invasive breast cancer were utilised in this study. All tissue blocks had previously been stained for Ki67 with the monoclonal antibody clone MIB1 (Agilent Dako, Stockport, UK; M7240 at a dilution of 1:50) and the EnVision REAL Detection System (now no longer available) in an automated staining system (Dako Autostainer; Dako, Glostrup, Denmark). Cohort 1 comprised 94 core-cut biopsies taken after 2 weeks of treatment with an AI. Cohort 2 comprised 20 excision specimens from patients who had either received or not received an AI for 2 weeks. Cohort 3 comprised 11 pairs of core-cut biopsies and excision specimens of patients who had either received or not received an AI for 2 weeks. Cohort 4 comprised 18 pairs of core-cut biopsies and excision specimens of patients to whom no presurgical therapy had been administered. In this fourth cohort, core-cut biopsies were taken immediately after resection (sample A) and following X-ray of the excised tumour for margin clearance (sample B), and, similarly to the main surgical specimen, they were placed in formalin (sample C).²⁹ Clinical and pathological information on the above patients was obtained from the hospital records and POETIC trial records (Table 1).

Table 1. Clinical and pathological features of breast cancer patients in the study

Clinicopathological characteristics	Number	%
Histological subtype		
Ductal	132	92.3
Lobular	8	5.6
Papillary	2	1.4
Mixed ductal-lobular	1	0.7
Tumour grade		
1	12	8.4
2	73	51.0
3	56	39.2
Not known	2	1.4
Tumour size (mm)		
≤20	55	38.5
20–50	79	55.2
>50	9	6.3
HER2 status		
Negative	108	75.5
Positive	34	23.8
Unknown	1	0.7

All patients were oestrogen receptor-positive.

SAMPLE PREPARATION

Blocks from the four cohorts that had previously been analysed for Ki67 as described above were first sectioned and stained with haematoxylin and eosin to assess the quality of each sample, such as tumour content, tumour cellularity, and crush artefacts.³⁰ Sections were cut and stained as above, but with MIB1 antibody at a dilution of 1:50 and the EnVision FLEX Detection System (Agilent Dako).

VISUAL SCORING

To account for the biological heterogeneity of Ki67 expression that frequently occurs across a breast tumour section, five representative fields of invasive breast cancer were identified by the use of low-power magnification. Counting was performed by use of a high-powered objective ($\times 40$) of a bright-field microscope, and invasive tumour cells of the entire fields were scored in each sample. This was the methodology

on which the KiBCWG had based its recommendations, in which 100 cells from four representative fields of invasive tumour cells were scored.³¹ The number of Ki67-positive nuclei were counted irrespective of the intensity of staining. Ki67 positivity was calculated as the percentage of the total number of Ki67-positive invasive tumour cells in all assessed fields relative to the total number of invasive tumour cells. A random selection of ~10% of the sections scored by the main observer were visually reappraised by a second observer, and an agreement between them was reached.

IMAGE ACQUISITION AND COMPUTER-ASSISTED DIGITAL SCORING

Each stained section was digitised with a NanoZoomer-XR (Hamamatsu Photonics, Hamamatsu, Japan) at $\times 20$ magnification. The quality of the images generated was reviewed manually prior to performance of the automated image analysis. Slides were scanned again if images were out of focus. Digital image viewing was achieved by using NDP.VIEW2 software (Hamamatsu Photonics).

Ki67 scoring was conducted with an automated image analysis approach implemented with the CogM software and the Ki67 Quantifier module (VMscope, Berlin, Germany)^{23,24} (Figure 1). To reflect the proliferation status of each tumour, four representative fields of the whole section were manually selected and captured in images displayed with NDP.VIEW2 software. The entire section image was examined at different magnification levels by zooming in and out in order to determine how the Ki67 staining was distributed in the invasive tumour component. If biological heterogeneity of Ki67 staining was present, the proportion of invasive tumour areas with different levels/frequencies of Ki67 positivity (high, medium, low, or negligible) was estimated. On the basis of these estimates, the four fields were selected to capture the full range of Ki67 staining frequencies and reflect the proportion of different Ki67 staining frequencies in the examined tumour. NDP.VIEW2 software was used to export selected representative fields as JPEG image files, which were subsequently analysed with CogM. Areas of non-tumour cells, infiltrating lymphocytes and intraductal components of breast carcinoma were avoided. For a few samples in which CogM clearly recognised these cell types as tumour cells, the observer manually excluded these areas. Cell membrane expression of Ki67 was detected in a small proportion of samples, but was considered to be an artefact.³² When membrane staining was observed in several areas of the tissue, the whole sample was

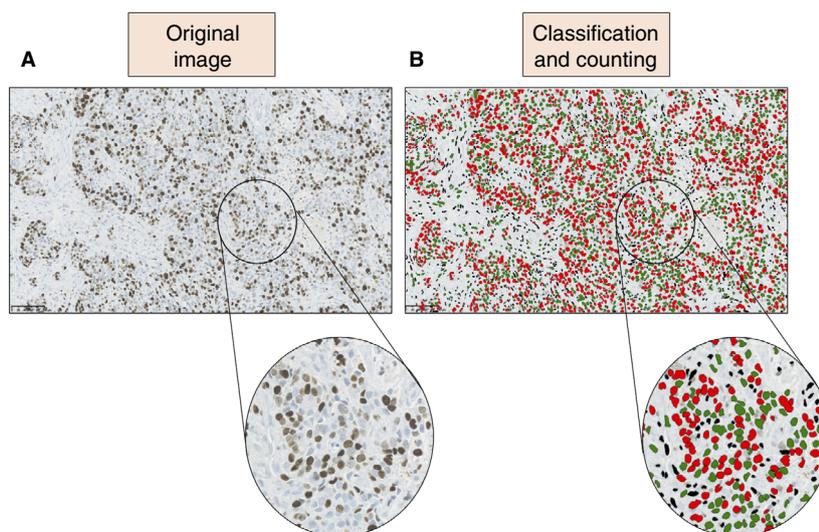


Figure 1. Example of the visualisation of Ki67 scoring results. A, Original image showing breast cancer tissue stained for Ki67 with MIB1 antibody (brown stain) and counterstained with Mayer's haematoxylin (blue stain). B, Analysis results obtained with the Cognition Master Professional Suite (CogM) Ki67 Quantifier software (green, Ki67-negative tumour cells; red, Ki67-positive tumour cells; black, non-tumour cells). The two areas circled in black are shown at higher magnification. Scale bar: 100 μm .

excluded from analysis. In cases in which only a few cells showed membrane staining, samples were still included in the analysis with exclusion of the relevant areas only. The time taken for image assessment in this study was approximately 5–10 min/section, depending on the size and type of tissue (core-cut biopsy or excision specimen). This included the time taken for the scanning of slides (1–3 min/slide), the selection and capture of fields (2–4 min/slide), and automated scoring (2–3 min/slide). Scoring was performed by a single observer, with confirmation of areas for scoring by a second observer when assignment of an area for analysis was uncertain. The percentage of Ki67 positivity was calculated in the same way as with the visual scoring system.

STATISTICAL ANALYSIS

Comparisons of Ki67 scores between manual and CogM assessment, as well as between core-cut biopsies and excision specimens, were conducted with Wilcoxon's signed-rank test with the raw Ki67 values. The statistical difference between groups was considered to be significant when the *P*-value was <0.05 . All statistical tests were two-sided by default. Calculations were performed with GRAPHPAD PRISM 7. Ki67 values were then log-transformed for normality, and correlations between manual and CogM assessment and different tumour tissue types were assessed by the use of Pearson's correlation coefficient. Bland–

Altman plots were produced from the log-transformed Ki67 values to investigate agreement of measurements between the two scoring methods and between the two tissue types. To assess the agreement between scoring systems after dichotomisation of Ki67 values, Cohen's kappa statistic was used.

Results

PROTOCOL DEVELOPMENT

As an initial pilot study showed lower Ki67 scores with CogM than with manual scoring (Figure S1), the pilot study was extended to assess a range of MIB1 antibody dilutions with the EnVision FLEX detection kit (Tables S1 and S2; Figures S2 and S3). A primary antibody dilution of 1:50, when used with the EnVision FLEX Detection System, was found to give higher visual scores that compensated for the lower image analysis scores obtained with CogM, such that scores obtained with the latter approximated to those obtained with previous manual scoring. The 1:50 dilution of the MIB1 antibody was therefore used in all comparisons below.

CORRELATION BETWEEN KI67 SCORING WITH THE MANUAL AND COGM METHODS

Ki67 scores obtained with the CogM method were compared with those determined previously with the

manual method (Figure 2). The mean number of tumour cells scored between the two methods indicated averages of 7691 cells/sample and 875 cells/sample scored with the CogM method and the manual method, respectively (Table S3). There was no significant difference between the manual and CogM scores for either the core-cut biopsies of Cohort 1 (Wilcoxon's signed rank test, $P = 0.4474$) (Figure 2A) or the excision specimens of Cohort 2 (Wilcoxon's signed rank test, $P = 0.9058$) (Figure 2D). The mean Ki67 values (\pm standard error of the mean) obtained with the manual and CogM methods for core-cut biopsies were 15.61 ± 1.49 and 15.15 ± 1.42 , respectively [mean difference, 0.45 ± 0.54 ; 95% confidence interval (CI) -0.60 to $+1.50$]. Similarly, the mean Ki67 values for excision specimens obtained with the manual and CogM methods were 17.76 ± 3.35 and 19.28 ± 3.93 ,

respectively (mean difference, -1.52 ± 1.99 ; 95% CI -5.42 to $+2.38$).

Ki67 scores obtained with the manual method were highly correlated with scores obtained with the CogM method for both core-cut biopsies (Pearson's $r = 0.9145$; 95% CI 0.8739 – 0.9425 ; $P < 0.0001$) and excision specimens (Pearson's $r = 0.9450$; 95% CI 0.8635 – 0.9784 ; $P < 0.0001$) (Figure 2B,E). Figure 2C and Figure 2F show that there was no significant bias on Bland–Altman analysis [core-cut biopsies—mean difference for $\log(\text{Ki67})$ -0.0084 , 95% CI -0.0484 to 0.0316 ; excision specimens—mean difference for $\log(\text{Ki67})$ 0.0466 , 95% CI -0.0687 to 0.162].

A cut-off level for Ki67 of 10% after use of an AI has been used in several studies,^{3,28,33,34} whereas a lower cut-off of 8% was found to provide even greater information in the POETIC trial.²⁷ We therefore

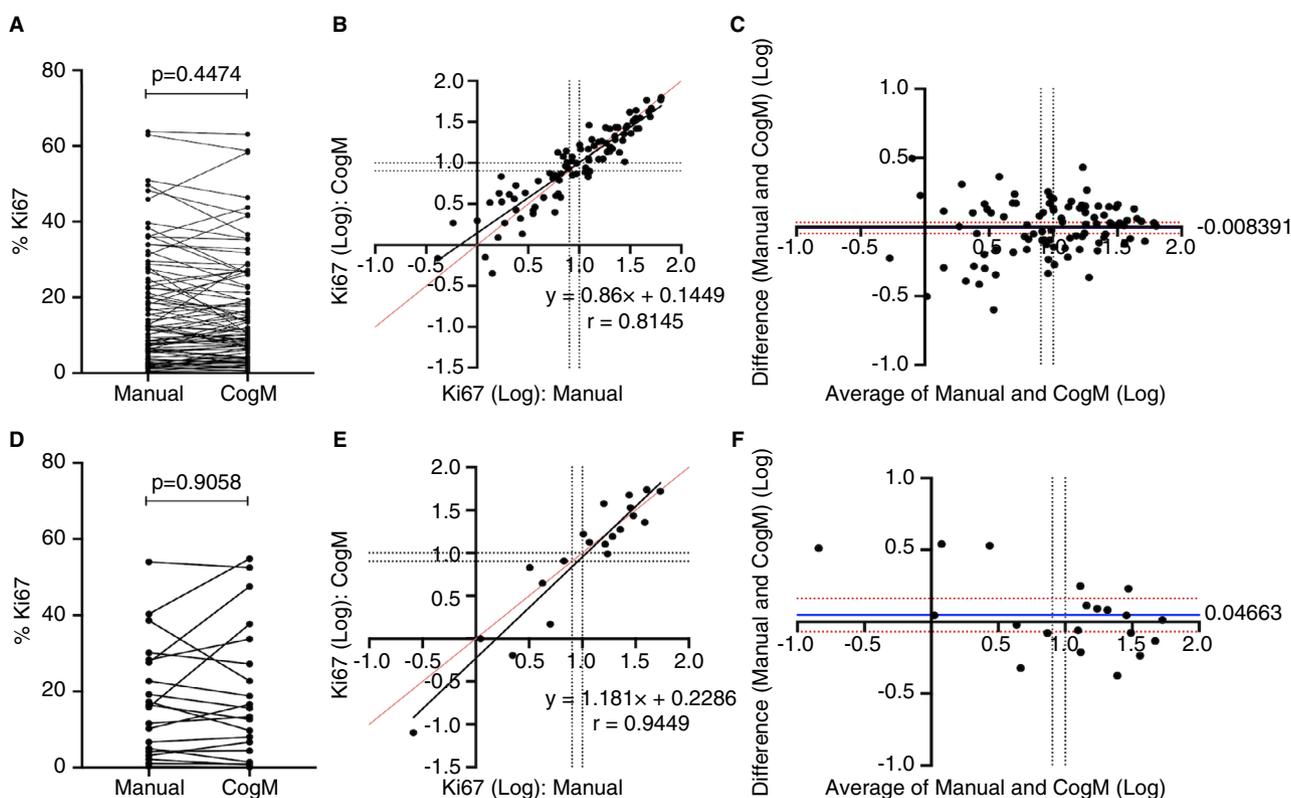


Figure 2. A,D, Comparisons of Ki67 scoring performed with the manual and Cognition Master Professional Suite (CogM) methods for individual core-cut biopsies (A) and excision specimens (D). P -values for the analysis were calculated with a two-sided Wilcoxon's matched-pairs signed-rank non-parametric test. B,E, Scatterplots showing the correlation between manual and CogM scoring for log-transformed Ki67 for core-cut biopsies (B) and excision specimens (E). Horizontal and vertical dotted lines indicate the 8% and 10% cut-off points. The correlation coefficient and slope were calculated with Pearson's correlation and linear regression parametric analyses, respectively. The pink line is the line of identity ($y = x$), and the black line represents the line of best fit. C,F, Bland–Altman plots of log-transformed data showing the difference in Ki67 scores between manual and CogM assessment against the average of the two for core-cut biopsies (C) and excision specimens (F). The vertical dotted lines indicate the 8% and 10% cut-off points, and the blue and red horizontal lines are drawn at the mean difference and the 95% confidence interval for the mean, respectively. r , correlation coefficient.

Table 2. Comparison of Ki67 scoring between the manual and Cognition Master Professional Suite (CogM) methods in 94 core-cut biopsies

Core-cut biopsies				
A	CogM		Total	
	<8	≥8		
Manual	<8	32	7	39
	≥8	6	49	55
	Total	38	56	94
				Concordance ratio: 86.2%

B	CogM		Total	
	<10	≥10		
Manual	<10	38	6	44
	≥10	6	44	50
	Total	44	50	94
				Concordance ratio: 87.2%

Concordance against discordance between manual and CogM scoring are shown following dichotomisation of data at (A) 8% and (B) 10% cut-off points.

assessed the concordance between manual and CogM scores after data had been dichotomised at the 8% and 10% cut-off points. The concordance rates between scores for core-cut biopsies were 86.2% and 87.2%, respectively. The kappa values of 0.714 [standard error (SE) 0.073; 95% CI 0.570–0.858] and 0.744 (SE 0.069; 95% CI 0.608–0.879) indicated good agreement between the scores (Table 2). Similarly, the concordance rates between scores for excision specimens were 95% for both cut-off points, with kappa values of 0.886 (SE 0.110; 95% CI 0.671–1.000) and 0.894 (SE 0.103; 95% CI 0.692–1.000), respectively, indicating almost perfect agreement (Table 3). Those cases that were discordant could be seen to fall close to the cut-off points (Figure 2).

COMPARISON OF KI67 EXPRESSION BETWEEN SURGICAL CORE-CUT BIOPSIES AND EXCISION SPECIMENS DETERMINED WITH AUTOMATED IMAGE ANALYSIS SOFTWARE

Ki67 staining and CogM assessment of core-cut samples A and B from 15 patients of Cohort 4 were conducted, and comparison between them was assessed. There was no significant difference in Ki67 scores

Table 3. Contingency table with data derived from 20 excision specimens examining the relationship between the manual and Cognition Master Professional Suite (CogM) methods for Ki67 scoring

Excision biopsies				
A	CogM		Total	
	<8	≥8		
Manual	<8	6	1	7
	≥8	0	13	13
	Total	6	14	20
				Concordance ratio: 95%

B	CogM		Total	
	<10	≥10		
Manual	<10	7	0	7
	≥10	1	12	13
	Total	8	12	20
				Concordance ratio: 95%

Concordance rates between scores calculated with the manual and CogM methods are shown following dichotomisation of data at (A) 8% and (B) 10% cut-off points.

between samples A and B (Wilcoxon's signed rank test, $P = 0.6387$) (Figure S4A), and scores were highly correlated (Pearson's $r = 0.7429$; 95% CI 0.3723–0.9092; $P = 0.0017$) (Figure S4B). Hence, comparisons with sample C, the matched excision sample, were performed with the mean Ki67 scores of samples A and B.

Ki67 scores were higher in core-cut biopsies than in excision specimens of Cohorts 3 and 4 (Wilcoxon's signed rank test, $P = 0.0408$) (Figure 3A), but there was good correlation (Pearson's $r = 0.7841$; 95% CI 0.5860–0.8938; $P < 0.0001$) (Figure 3B), and the Bland–Altman analysis showed a bias of 0.1586 (95% CI 0.0667–0.2894) between them for log-transformed Ki67 scores (Figure 3C). Dichotomisation of data at the 8% and 10% cut-off points resulted in concordance rates of 79.3% and 72.4%, respectively (Table 4). The kappa values were 0.603 (SE 0.133; 95% CI 0.343–0.863) and 0.482 (SE 0.133; 95% CI 0.221–0.743), suggesting moderate agreement.

Discussion

Analytical variability within and particularly between centres has limited the widespread use of Ki67 in the

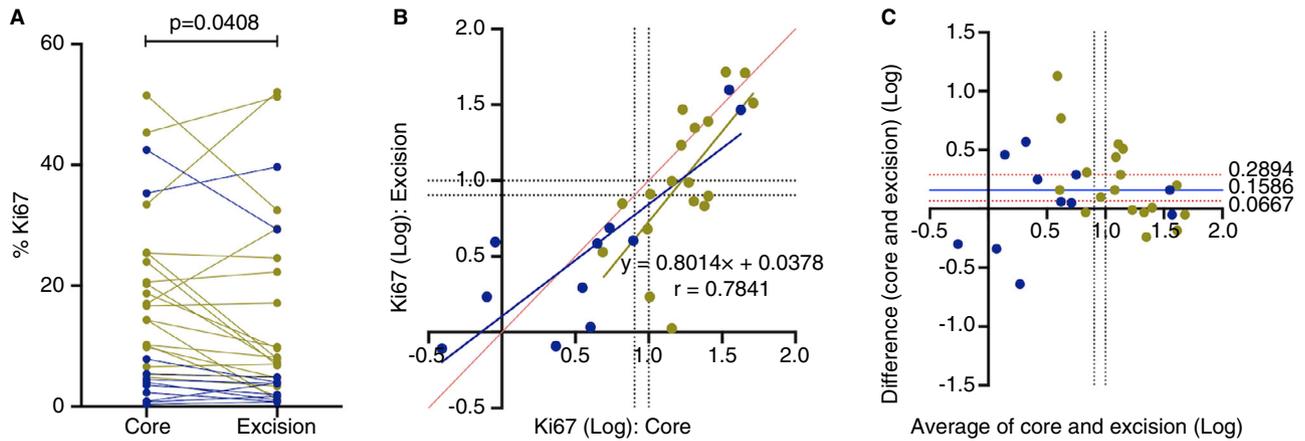


Figure 3. A, Comparison of Ki67 expression for individual core-cut biopsies and excision specimens measured following Cognition Master Professional Suite (CogM) analysis. The *P*-value for the analysis was calculated with a two-sided Wilcoxon’s matched-pairs signed-rank non-parametric test. B, Scatterplots indicating the correlation of log-transformed Ki67 scoring between core-cut biopsies and excision specimens when CogM was used. The horizontal and vertical dotted lines indicate the 8% and 10% cut-off points. The correlation coefficient and slope were calculated with Pearson’s correlation and linear regression parametric analyses, respectively. The straight lines represent the lines of best fit for each set of samples, and the pink line is the line of identity ($y = x$). C, Bland–Altman plots of log-transformed data showing the difference in Ki67 scores for core-cut biopsies and excision specimens when CogM was used. The vertical dotted lines indicate the 8% and 10% cut-off points, and the blue and red horizontal lines are drawn at the mean difference and the 95% confidence interval for the mean, respectively. Samples of Cohort 3 are labelled in dark blue, and those of Cohort 4 are labelled in green. *r*, correlation coefficient. [Colour figure can be viewed at wileyonlinelibrary.com]

Table 4. Contingency table with Ki67 values derived from 29 core-cut biopsies and excision specimens taken at the time of surgery

A	Core-cut biopsies		Total
	<8	≥8	
Excision specimens	<8	11	17
	≥8	0	12
	Total	11	29
Concordance ratio: 79.3%			

B	Core-cut biopsies		Total
	<10	≥10	
Excision specimens	<10	12	20
	≥10	0	9
	Total	12	29
Concordance ratio: 72.4%			

Concordance of Ki67 values between core-cut biopsies and excision specimens was assessed with CogM and dichotomisation of data at (A) 8% and (B) 10% cut-off points.

assessment of breast cancer specimens, despite its clearly strong prognostic significance and multiple other potential applications. The efforts of the

KiBCWG have led to improvements, particularly with a recommended manual scoring procedure, which was used as a reference method here. Nonetheless, it is clear that it is impossible to fully remove differences between analysts scoring visually, and, even with the same analyst, drift in scoring can happen over time. The availability of an automated, simple and stable system has multiple advantages.

Many image analysis systems exist, with most providing the opportunity to change settings for the large number of parameters that allow the score to be derived. In this regard, the ‘nailed-down’, non-adjustable software of CogM is an advantage for routine application between centres, because no opportunity exists for modification. Although the digital image scoring is quite quick, there are manual aspects, such as the selection of fields and exclusion of non-tumour cells in a few samples, that can slow the assessment down by a few minutes. The time taken for scoring of each section with CogM in this study was 5–10 min, and is likely to be acceptable in the working practice of most histopathologists.

Our data indicate that CogM, when used with a modified staining procedure, provides highly comparable data to those obtained from a set of samples from the POETIC trial, which provides a calibration of the Ki67 scores against clinical outcome. This observation is consistent with some previously published breast cancer studies, which showed that Ki67 scores obtained with automated scoring methods, including CogM, can be

highly concordant with visual assessment.^{19,20,22,35} Although there was some variability between the earlier manual scores and CogM scores, this may be because sections were not adjacent, as new sections were required to be cut deeper in the block for the modified staining and CogM analysis. It should also be noted that a large proportion of the samples assessed here were taken after AI treatment, which markedly reduces proliferation and Ki67 staining (by ~80% after 2 weeks). The precision of scoring is reduced with a lower numerator (numbers of positive cells), and the inclusion of samples taken after presurgical treatment is likely to have led to higher estimates of variability than would have been the case for samples from untreated patients. Nearly 10 times as many cells were scored by CogM, which provides greater precision in the data, and this is particularly advantageous when the proportion of positive cells is small.

The close relationship between the CogM data and the results from the POETIC trial indicate that the CogM method is highly suited to assess relationships with the risk of recurrence, and will show the same relationship with clinical outcome as recently reported for that trial.²⁷ It is of particular note that the trial demonstrated the prognostic information available from Ki67 analyses undertaken after 2 weeks of treatment with an AI, and that the information was additional to that available from samples taken before treatment. This approach to treating patients with an AI before surgery to derive this information on biological response to an AI and the consequent information about prognosis is becoming increasingly utilised, e.g. in managing oestrogen receptor-positive breast cancer patients in the COVID-19 pandemic,³⁶ and our data demonstrate the validity of the CogM method for that purpose.

Immunohistochemical Ki67 assessment with CogM was similar between matched surgical core-cut biopsies and excision specimens, and good correlation for Ki67 scores was obtained. The sample size was, however, relatively small, and there was a trend towards lower values in excision specimens. We have previously noted such a statistically significant difference in the much larger POETIC study,²⁷ which is probably explained by the longer time needed for formaldehyde to penetrate the larger volumes of excision tissue during fixation, with reduced proliferation in the non-fixed tissue and/or antigen loss.^{15,37,38} The impact of tissue sample type on measurements of biomarkers has been assessed previously,^{29,39} with expression of phosphorylated proteins, such as p-Akt and p-Erk1/2, being markedly reduced in excision specimens relative to core-cut biopsies.^{29,40}

To conclude, we have shown that CogM alongside a modified staining process is valid for evaluating the immunohistochemical expression of Ki67 in primary invasive breast cancer, with data that are similar to those obtained from previously assessed visual scoring, and that CogM can be applied with confidence to both core-cut biopsies and excision specimens, which is an important advantage for assessments in the presurgical setting.

Conflicts of interest

A. Dodson is on the scientific advisory board for Visiopharm. M. Dowsett is on advisory boards for Radius, G1 therapeutics, AbbVie, Zentalis, and H3 Biomedicine, and receives lecture fees from Nanos-tring, Myriad, and Lilly. The salary of H. Toveys has been supported at one time or another by research grants paid to the Clinical Trials and Statistics Unit in the Institute of Cancer Research by: Pfizer, Janssen-Cilag Ltd, Merck, AstraZeneca, and Clovis. The remaining authors have no conflicts of interest to declare.

Author contributions

A. Alataki performed the research, analysed and interpreted the data, and wrote the paper. L. Zabaglo performed the initial stages of the research and supervised all stages of research. H. Tovey helped with sample logistics. A. Dodson performed initial stages of the research. M. Dowsett contributed to the conception and design of the research study, and revision and approval of the final version to be published.

Acknowledgements

We thank all participating patients in the POETIC trial (C1491/A8671/CRUK/07/015, C1491/A15955, and C406/A8962), from which samples were obtained for this study. This work was supported by the National Institute for Health Research Biomedical Research Centre at the Royal Marsden NHS Foundation Trust and the Institute of Cancer Research, London. We also thank Breast Cancer Now for funding this work as part of Programme Funding to the Breast Cancer Now Toby Robins Research Centre. The POETIC trial was supported by Cancer Research UK. A. Alataki was supported by a grant from the Le Cure Breast Cancer Research Fund.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011; **144**: 646–674.
- 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**: 1656–1664.
- Ellis MJ, Suman VJ, Hoog J et al. Ki67 proliferation index as a tool for chemotherapy decisions during and after neoadjuvant aromatase inhibitor treatment of breast cancer: results from the American College of Surgeons Oncology Group Z1031 Trial (Alliance). *J. Clin. Oncol.* 2017; **35**: 1061–1069.
- De Azambuja E, Cardoso F, De Castro G et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients. *Br. J. Cancer* 2007; **96**: 1504–1513.
- Inwald EC, Klinkhammer-Schalke M, Hofstädter F et al. Ki-67 is a prognostic parameter in breast cancer patients: results of a large population-based cohort of a cancer registry. *Breast Cancer Res. Treat.* 2013; **139**: 539–552.
- Petrelli F, Viale G, Cabiddu M et al. Prognostic value of different cut-off levels of Ki-67 in breast cancer: a systematic review and meta-analysis of 64,196 patients. *Breast Cancer Res. Treat.* 2015; **153**: 477–491.
- Criscitelli C, Disalvatore D, De Laurentiis M et al. High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in luminal B HER2 negative and node-positive breast cancer. *Breast* 2014; **23**: 69–75.
- Cohen AL, Factor RE, Mooney K et al. POWERPIINC (PreOperative Window of Endocrine TheRapy Provides Information to Increase Compliance) trial: changes in tumor proliferation index and quality of life with 7 days of preoperative tamoxifen. *Breast* 2017; **31**: 219–223.
- Denkert C, Budczies J, von Minckwitz G et al. Strategies for developing Ki67 as a useful biomarker in breast cancer. *Breast* 2015; **24**: S6–S72.
- Viale G, Regan MM, Maiorano E et al. Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1–98. *J. Clin. Oncol.* 2007; **25**: 3846–3852.
- Viale G, Regan MM, Dell'Orto P et al. Which patients benefit most from adjuvant aromatase inhibitors? Results using a composite measure of prognostic risk in the BIG 1–98 randomized trial. *Ann. Oncol.* 2011; **22**: 2201–2207.
- Dixon JM. Endocrine resistance in breast cancer. *New J. Sci.* 2014; **2014**: 1–27.
- 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**: 1–9.
- Yerushalmi R, Woods R, Ravdin PM et al. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol.* 2010; **11**: 174–183.
- Walker RA. Quantification of immunohistochemistry—issues concerning methods, utility and semiquantitative assessment I. *Histopathology* 2006; **49**: 406–410.
- 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**: 1–12.
- Polley M-Y, Leung SCY, Gao D et al. An international study to increase concordance in Ki67 scoring. *Mod. Pathol.* 2015; **28**: 778–786.
- 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.
- Cass JD, Varma S, Day AG et al. Automated quantitative analysis of p53, cyclin D1, Ki67 and pERK expression in breast carcinoma does not differ from expert pathologist scoring and correlates with clinico-pathological characteristics. *Cancers (Basel)* 2012; **4**: 725–742.
- Mohammed ZMA, McMillan DC, Elsberger B et al. 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.
- Konsti J, Lundin M, Joensuu H et al. Development and evaluation of a virtual microscopy application for automated assessment of Ki-67 expression in breast cancer. *BMC Clin. Pathol.* 2011; **11**: 3.
- Klauschen F, Wienert S, Schmitt WD et al. Standardized Ki67 diagnostics using automated scoring—clinical validation in the GeparTrio breast cancer study. *Clin. Cancer Res.* 2015; **21**: 3651–3657.
- Wienert S, Heim D, Kotani M et al. CognitionMaster: an object-based image analysis framework. *Diagn. Pathol.* 2013; **8**: 1–8.
- Wienert S, Heim D, Saeger K et al. Detection and segmentation of cell nuclei in virtual microscopy images: a minimum-model approach. *Sci. Rep.* 2012; **2**: 1–7.
- Smith IE, Walsh G, Skene A et al. A phase II placebo-controlled trial of neoadjuvant anastrozole alone or with gefitinib in early breast cancer. *J. Clin. Oncol.* 2007; **25**: 3816–3822.
- Gao Q, López-Knowles E, Cheang MCU et al. Impact of aromatase inhibitor treatment on global gene expression and its association with antiproliferative response in ER+ breast cancer in postmenopausal patients. *Breast Cancer Res.* 2019; **22**: 1–20.
- Smith I, Robertson J, Kilburn L et al. Long-term outcome and prognostic value of Ki67 after perioperative endocrine therapy in postmenopausal women with hormone-sensitive early breast cancer (POETIC): an open-label, multicentre, parallel-group, randomised, phase 3 trial. *Lancet Oncol.* 2020; **21**: 1443–1454.
- Dowsett M, Smith I, Robertson J et al. Endocrine therapy, new biologicals, and new study designs for presurgical studies in breast cancer. *J. Natl. Cancer Inst. Monogr.* 2011; **2011**: 120–123.
- Pinhel IF, MacNeill FA, Hills MJ et al. Extreme loss of immunoreactive p-Akt and p-Erk1/2 during routine fixation of primary breast cancer. *Breast Cancer Res.* 2010; **12**: 1–7.
- Zabaglo L, Salter J, Anderson H et al. Comparative validation of the SP6 antibody to Ki67 in breast cancer. *J. Clin. Pathol.* 2010; **63**: 800–804.
- Leung SCY, Nielsen TO, Zabaglo LA et al. Analytical validation of a standardised scoring protocol for Ki67

- immunohistochemistry on breast cancer excision whole sections: an international multicentre collaboration. *Histopathology* 2019; 75: 225–235.
32. Del Sordo R, Sidoni A. MIB-1 cell membrane reactivity: a finding that should be interpreted carefully. *Appl. Immunohistochem. Mol. Morphol.* 2008; 16: 568.
 33. Robertson J, Dowsett M, Bliss J *et al.* Peri-operative aromatase inhibitor treatment in determining or predicting longterm outcome in early breast cancer—The POETIC Trial* (CRUK/07/015). *Cancer Res.* 2018; 78; Abstract GS1-03.
 34. Hofmann D, Nitz U, Gluz O *et al.* WSG ADAPT—adjuvant dynamic marker-adjusted personalized therapy trial optimizing risk assessment and therapy response prediction in early breast cancer: study protocol for a prospective, multi-center, controlled, non-blinded, randomized, investigator initiated phase II/III trial. *Trials* 2013; 14: 261.
 35. Thakur SS, Li H, Chan AMY *et al.* The use of automated Ki67 analysis to predict Oncotype DX risk-of-recurrence categories in early-stage breast cancer. *PLoS One* 2018; 13: 1–18.
 36. Dowsett M, Ellis MJ, Dixon JM *et al.* Evidence-based guidelines for managing patients with primary ER+ HER2– breast cancer deferred from surgery due to the COVID-19 pandemic. *npj Breast Cancer* 2020; 6: 21.
 37. Neumeister VM, Parisi F, England AM *et al.* A tissue quality index: an intrinsic control for measurement of effects of preanalytical variables on FFPE tissue. *Lab. Invest.* 2014; 94: 467–474.
 38. Goldstein NS, Ferkowicz M, Odish E *et al.* Minimum formalin fixation time for consistent estrogen receptor immunohistochemical staining of invasive breast carcinoma. *Am. J. Clin. Pathol.* 2003; 120: 86–92.
 39. Leal MF, Haynes BP, MacNeill FA *et al.* Comparison of protein expression between formalin-fixed core-cut biopsies and surgical excision specimens using a novel multiplex approach. *Breast Cancer Res. Treat.* 2019; 175: 317–326.
 40. Arnedos M, Nerurkar A, Osin P *et al.* Discordance between core needle biopsy (CNB) and excisional biopsy (EB) for estrogen receptor (ER), progesterone receptor (PgR) and HER2 status in early breast cancer (EBC). *Ann. Oncol.* 2009; 20: 1948–1952.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Box plots and scatterplot comparing Ki67 scoring results between the use of the manual method and use of the Cognition Master Professional Suite following staining with a 1:50 dilution of the MIB1 antibody along with the EnVision REAL Detection System.

Figure S2. Scatterplots and Bland–Altman plots comparing Ki67 scoring results derived from tissue microarray cores between the use of the manual method following staining with a 1:1200 dilution of the MIB1 antibody and use of the Cognition Master Professional Suite after staining with a range of MIB1 antibody dilutions. The EnVision FLEX Detection System was used in all cases.

Figure S3. Scatterplots and Bland–Altman plots comparing Ki67 scoring results derived from core-cut biopsies between the use of the manual method following staining with a 1:1200 dilution of the MIB1 antibody and use of the Cognition Master Professional Suite after staining with a range of MIB1 antibody dilutions. The EnVision FLEX Detection System was used in all cases.

Figure S4. Spaghetti plots and scatterplot comparing Ki67 scoring results between core-cut biopsies A and B following use of the Cognition Master Professional Suite and the EnVision FLEX Detection System.

Table S1. Ki67 scoring results following staining of tissue microarray controls with the MIB1 antibody and the EnVision FLEX Detection System.

Table S2. Ki67 scoring results after staining of core-cut biopsies with the MIB1 antibody.

Table S3. Numbers of tumour cells and Ki67-positive tumour cells scored with the Cognition Master Professional Suite and the manual method per sample.