Breast Cancer Histologic Grading using Digital Microscopy: Concordance and Outcome Association

Emad A Rakha[1*], Mohammed Aleskandarany[1*], Michael S Toss[1], Andrew R Green[1], Graham Ball[2], Ian O Ellis[1] and Leslie W Dalton[3]

[1] Division of Cancer and Stem Cells, School of Medicine, University of Nottingham, Nottingham City Hospital, Hucknall Road, Nottingham NG5 1PB, UK

[2] John van Geest Cancer Research Centre, School of Science and Technology, Nottingham Trent University, Nottingham, UK

[3] Department of Histopathology, South Austin Hospital, Austin, TX USA

*Both authors contributed equally to this study

Correspondence:
Professor Emad Rakha
Department of Histopathology,
Nottingham University Hospital NHS Trust,
City Hospital Campus, Hucknall Road, Nottingham,
NG5 1PB, UK
Tel: (44) 0115-9691169
Fax: (44) 0115- 9627768
Email: Emad.rakha@nuh.nhs.uk
emadrakha@yahoo.com

Keywords: Breast pathology, grade, agreement, virtual microscopy

Running Title: Virtual versus conventional grading of invasive BC

ABSTRACT

AIMS: Virtual microscopy utilising digital whole slide imaging (WSI) is increasingly used in breast pathology. Histologic grade is one of the strongest prognostic factors in breast cancer (BC). This study aims at investigating the agreement between BC grading using traditional light microscopy (LM) and digital whole slide imaging (WSI) with consideration of reproducibility and impact on outcome prediction.

METHODS: A large (n=1675) well-characterised cohort of BC originally graded by LM was re-graded using WSI. Two separate virtual-based grading sessions (V1 and V2) were performed with a three months washout period. Outcome was assessed using breast cancer specific and distant metastasis free survival.

RESULTS: The concordance between LM grading and WSI was strong (LM/SWI *Cramer's V*: V1=0.576, and V2=0.579). The agreement regarding grade components was as follows: Tubule formation=0.538, Pleomorphism=0.422 and Mitosis=0.514. Greatest discordance was observed between adjacent grades whereas high/low grade discordance was uncommon (1.5%). The intra-observer agreement for the two WSI sessions was substantial for grade (V1/V2 *Cramer's V*=0.676; *kappa*=0.648) and grade components (*Cramer's V* T=0.628, P=0.573 and M=0.580). Grading using both platforms showed strong association with outcome (All *p-value* <0.001). Although mitotic scores assessed using both platforms were strongly associated with outcome, WSI tends to underestimate mitotic counts.

CONCLUSIONS: Virtual microscopy is a reliable and reproducible method for assessing BC histologic grade. Regardless of the observer or assessment platform, histologic grade is a significant predictor of outcome. Continuing advances in imaging technology could potentially provide improved performance of WSI BC grading and in particular mitotic count assessment.

INTRODUCTION

Virtual microscopy (VM) using digital whole slide imaging (WSI) is a technology through which glass slides of pathologic specimens are digitally scanned at high-resolution for viewing on a computer screen. Applications of WSI in the clinical, educational, and research settings including image analysis applications are increasing and in some centres WSI has replaced conventional microscopy as a diagnostic tool used by pathologists [1-6]. However, one of main the concerns related to VM adoption in breast pathology, in addition to diagnosis, is the assessment of prognostic and predictive variables including histologic grade [6]. There is a perception that the quality of the images displayed by WSI may interfere with reliable histologic grading. In addition the interpretive ability of the reporting Pathologist assigning a “virtual grade” to each cancer remains largely unknown.

Therefore, to improve WSI performance, enhancement of the WSI platform and the training of histopathologists with the digital environment is recommended. However, testing performance and reproducibility of WSI in cases’ reporting is critically needed. This could be achieved via head-to-head comparison of WSI compared to traditional light microscopy (LM) to provide sufficient evidence prior to clinical adoption.

Grading of breast cancer using the Nottingham combined histologic grade is one of the strongest prognostic factors in early stage disease [7-9]. Grade comprises one of the main components of several management decision tools [10-13] and it has recently been included in the American Joint Committee on cancer (*AJCC*) TNM staging system as a stage modifier [14,15]. However concordance of breast cancer grading among pathologists using glass slides shows moderate agreement with kappa values of 0.48 [16] to 0.53 [17]; the high concordance rates observed in grade 3 (kappa 0.60) and grade 1 (kappa 0.51) tumours whereas the lowest is observed with grade 2 (kappa 0.33) tumours [16]. The impact of introduction of WSI in routine practice on the concordance of grade and its performance as a prognostic factor remains to be defined.

Therefore, this study aims at comparing the histologic grading of breast cancers as assigned by an expert pathologist using WSI with the grade assessed in routine practice using LM. In addition to assessment of concordance, impact of different grading platforms on patient outcome was evaluated using the large well characterised Nottingham breast cancer cohort.

PATIENTS AND METHODS

This study has been performed on a large series (n=1675) of early stage invasive primary operable breast cancer patients presented to Nottingham City Hospital from 1999-2006. This is a well-characterised cohort of breast cancer with long term clinical follow-up (median 135 months) and detailed clinico-pathological profiles. Data included primary tumour histologic grade and grade components, tumour size and histotype, lymph node stage, nodal status, lymphovascular invasion, Nottingham Prognostic Index, molecular subtypes and outcome data was collected. The latter include breast cancer-specific survival (BCSS), defined as time (in months) from the date the primary surgical treatment to the time of death from breast cancer, and distant metastasis free survival (DMFS) was defined as the time (in months) from the surgery until the first event of distant metastasis. Patient and tumour demographics are summarised in Table 1.

This tumour cohort was originally graded using the Nottingham grading system during routine pathology reporting utilising all available tumour glass slides (average 4 slides per case) and light microscopy (LM) [8]. For the purpose of this study, data for the final grade as well as the individual grade components (tubule formation, nuclear pleomorphism and mitotic count scores) was retrieved from the patients' records. 1-3 tumour blocks per case were retrieved and freshly prepared H&E slides were reviewed. A representative slide per case was selected by a specialised breast pathologist (EA Rakha) without further glass slide grading. Glass slides were scanned into high-resolution (0.19 µm/pixel) digital images at 20x magnification using 3D Histech Panoramic 250 Flash II scanner (3DHISTECH Ltd., Budapest, Hungary). The whole slide digital images (WSI) were generated, stored and viewed using the 3D Histech Pannoramic Viewer (3DHISTECH Ltd., Budapest, Hungary; http://www.3dhistech.com/downloads) on a high resolution screen. The digital slide was graded using College of American Pathologists' criteria [18] which are essentially the same as the original Nottingham criteria [7]. Digital images were initially examined at low magnification where tubule formation was assessed. Also, low to intermediate magnification was performed for the identification of potential "hotspots" for mitotic counting. Essentially, for mitotic counting, the distance measure tool of the software was used. This was important for determining the number of mitotic figures in a given area.

To allow for intra-observer agreement of BC grading using WSI, the whole cohort was graded again by the same observer (Dr L. Dalton who is an experienced breast pathologists with special interest in breast cancer grading) using the same criteria twice after a 3-month washout time with no special training during that time. In both WSI grading sessions (V1 and V2), grade components were assigned blinded to the LM grade as well as other clinicopathological parameters.

This study was approved by Nottingham Research Ethics Committee 2 under the title of "Development of a molecular genetic classification of breast cancer".

Statistical analysis

Statistical analysis was performed using functions obtained from the open-source R statistical platform [19]. Since WSI and LM have a procedural difference, Cramer's V statistic was adopted to help judge strength of concordance [20]. The coefficient ranges from 0 (no association) to 1 (perfect association). The *kappa* statistic is technically a measure of concordance between two observers who are examining the same parameter following the same approach [21]. For these analyses, R library vcd had the function required for calculating the *Cramer's V* statistic (= function assocstats), while the function for kappa statistic was obtained from R library inter-rater reliability. Survival analysis was performed using SPSS

23 (SPSS 23 for Windows, Chicago, IL, USA) using log rank test and Kaplan Meier plots. Survival analysis (BCSS and DMFS) was performed on WSI grade as well as the WSI component scores. Likewise, survival analysis was performed for the glass-slide LM grade to include separate analysis of the component scores. Multivariate analysis was performed using cox proportional hazards analysis with inclusion of parameters significantly associated with outcome in univariate analysis. Statistical significance in survival stratification was calculated by the log-rank method and univariate cox regression analysis. A p-value of less than 0.05 (two tailed) was considered significant.

RESULTS

In this study, a large 1675 retrospective cohort of early invasive primary operable BC were graded on high-resolution digital images acquired through WSI of representative slides. For this WSI grading, as for the original LM, the three-tier histologic grade of BC was used [7,8,18]. Tables 2 and 3 show cross comparison of WSI grade with LM grade as well as the cross comparison of the three components of the Nottingham grade. Table 4 shows the cross comparison of the sum of grade components (3-9 scale) for LM grade scores and WSI grade scores.

The agreement between WSI grading and glass slide/LM grading was moderate for the both WSI grading sessions when kappa statistic was used (V1/LM *kappa*=0.51, and V2/LM *kappa*=0.50). However, when *Cramer's V* statistic was used the *Cramer's V* for WSI with LM was 0.58 in both sessions respectively which is considered as a substantial concordance. If grade is reduced to a binary level of high (i.e. grade 3) versus not high (i.e. grade 1 &2), the *Cramer's V* was 0.66. The unweighted kappa statistic for WSI grade with LM grade was 0.51. The kappa statistics for component scores were: mitoses = 0.47; tubules = 0.49; and pleomorphism = 0.3.

Importantly, exact grade agreement between WSI and LM grading was reached in 68% of cancers. There was almost an even match in the number of cancers with low-intermediate discordance (255 cancers;

15.1%) as compared to intermediate-high discordance (265 cancers; 15.7%). There was 32.3% discordance between WSI and LM grade, which is largely between adjacent levels of grade: low versus intermediate, or intermediate versus high grade. The binary high versus low/intermediate discordance of grade was 17%. In this study only twenty-six (1.5%) grade assignments were attributable to high versus low-grade discrepancy and the proportion reduction of high in LM to low in WSI was highly significant (p< 0.00001).

The intra-observer agreement for the two virtual sessions (V1/V2) was higher than the values of agreement between WSI and LM but remained in the moderate concordance category (table 5).

Survival analysis

Survival analysis was performed on both the grade assigned by WSI and the original LM, Table 5. WSI grading in both grading sessions showed statistically significant differences for BCSS and DMFS as did the LM grading ($p = 1x10^{-13}$) (table 6). Individual WSI grade components showed statistically significant differences for BCSS and DMFS. WSI tubule formation showed a stronger association with BCSS than that of LM (Hazards ratio (HR)=2.8, 95%CI=1.9-4.0, and 1.9, 95%CI=1.5-2.4, for WSI and LM respectively). Similar results were observed for DMFS (HR=2.6, 95% CI=1.9-3.6, and 1.7 95%CI= 1.4-2.1). Figures1 and 2 show survival curves of the final WSI and LM based histologic grade as well as grade components and BCSS.

To assess the prognostic independency of BC grade assigned using LM and WSI V1 and V2 multivariate analyses were performed including other established prognostic variables in the models. LM grading as well as WSI V1 and V2 were significantly associated with BCSS (p value for the 3 grading methods were < 0.001) and DMFS ((p value < 0.001), independent of other variables, Table 7.

DISCUSSION

Currently, there is an increasing interest in using WSI for diagnostic and research purposes. However, it is crucial to ensure that diagnostic performance utilising virtual slides is at least equivalent to that of using conventional light microscopy. To validate the diagnostic concordance of WSI and LM, USCAP recommends 12 rigorously developed guidelines with the potential of providing pathology laboratories with a practical guide to validate WSI systems for diagnostic work [22]. These include, yet not limited to the number required for double reporting (at least 60 cases per application) and the washout period (at least 2 weeks). In the current study, more than 1600 breast cancers were regraded using WSI by expert pathologists and the results were compared to the original routine practice generated grade. The quantifiable three-tier system of Nottingham grade compiling the degree of tubule formation, nuclear pleomorphism and mitotic frequency scores, is an ideal parameter for comparing WSI with LM. To assess the intra-observer concordance and the impact of WSI training on, the whole cohort was grading again with a long washout interval of 3 months. The end-point for this study was the concordance statistics as well as the patients' clinical outcome. To the best of our knowledge, this is the largest study performing head-to-head comparison of breast cancer grading using WSI and LM including patients' survival as a study endpoint.

WSI grading showed moderate concordance with LM grading comparable to concordance rate reported among different pathologists who graded breast cancer using conventional microscopy [16,17] . Exact grade agreement between WSI and LM grading was reached in 68% of cases. This magnitude of concordance is in-line with a prior reproducibility study [23]. Since WSI has a procedural difference, compared with LM, some emphasis was given to *Cramer's V* as measure of concordance. Multiple authorities considered a *Cramer's V* of a value greater than 0.5 to be the break point for acceptable concordance [24,25]. In the current study, WSI grade as compared to LM grade had *Cramer's V* of 0.58 at the ternary level and 0.66 at a binary level. These figures indicate high levels of reproducibility and demonstrate WSI reliability as a platform for grading breast cancer taking into account the inherent discordance levels in grade assignment between different observers using a single platform. Detailed analysis of discordance at the level of individual cases awaits further study, to include evaluation by recently introduced technologies. In this study, the true merit of WSI as compared to LM was further studied as regard to whether both offered a comparable level of patients' survival stratification utilising the large number of cases with long-term follow-up data. Both WSI and LM showed significant association with patients' outcome as well the individual grade components assessed by both platforms. Interestingly, tubule formation as assessed by WSI showed stronger association with outcome compared VLM assessment. Of note our study demonstrated that morphology is easy enough to be amendable to survival analysis while technically difficult molecular assays are not [26,27].

The intra-observer agreement for the two WSI sessions was moderate and showed similar association with outcome. These results support the fact that the level of concordance is to a large extent related to observer performance and the subjective nature of grade rather than the platform used. This together with the limitation of the current study which include 1) grade was assessed by different observers, 2) original grade was assessed using an average of four tumour tissue slides per case whereas WSI grade was assessed on a single slide, and 3) WSI scan magnification used was 20x rather than 40x which is considered ideal for assessment of mitotic counts. In fact, among the three Nottingham grade components, the most challenging component to evaluate by WSI was mitotic counts. There was difficulty in discerning mitotic figures from apoptotic cells. Although this was largely attributed to resolution, the inability of WSI to provide different planes may have contributed as an additional hurdle. Therefore, assessment of mitotic counts using 40x magnification may help resolving this issue. However, the large number of cases in this study and the repeated grade by the same observer using WSI has potentially overcome these limitations. The reasons for the tendency of lower mitotic scores in WSI compared to LM are likely related to the use of a single slide per case and the lower magnification used in WSI.

In breast cancer grading there will be, without doubt, some discordance between grade assignments by WSI as well as to LM grading. Also, comparisons among biomarkers tested for diagnostic and research purposes share this possibility of discordance [25,26,28]. However, at the level of an individual patient especially in the diagnostic setting, discordance is usually met with caution and concordance is sought for. Therefore, sustained effort is critically needed for improvement in concordance, or at least for an improved understanding in the meaning of discordance. In the current study, grading was validated as a ternary scheme and as binary scheme to assess for concordance of both grading platforms. Previous studies addressing binary biomarkers have compared their results with grade by collapsing grade into a binary scheme. For illustrative purposes, we did the same, and showed strong concordance of WSI with LM, no matter if low and intermediate grade were combined, or intermediate with high grade; concordance of grade was Cramer's V= 0.55 if low combined with intermediate.

As mentioned above two reasons are thought to be responsible for underperforming of WSI in the assessment of histologic grade which are the technology itself or WSI and the reader. This study demonstrates that grading using WSI is not only reproducible but also provides significant survival information comparable to glass slides. The concordance rate between glass slides grading and WSI was comparable to these reported using glass slides as the only tool and the intra-observer concordance using WSI was even higher than that reported by multiple readers using glass slides [29,30]. This study in addition to providing evidence for the reproducibility and reliability of WSI in grading breast cancer could prompt the question of what would be the minimal number of cases, randomly selected, which would be expected to show if a histopathologist would show ability to predict survival using WSI grade. If low enough then WSI may be a method to test competence at the level of survival prediction and not just concordance. The use of WSI technology also opens up opportunities for computer assisted classification of histologic grade with inherent improved standardisation and reproducibility of evaluation and potential for refinement of methodology.

Take home messages

- Regardless of the observer or assessment platform, histologic grade is a significant predictor of outcome.

- Virtual microscopy is a reliable and reproducible method for assessing breast cancer histologic grade.

- Higher magnification (x40) is recommended to produce adequate resolution for an accurate grading

- Continuing advances in imaging technology could potentially provide improved performance of whole slide imaging breast cancer grading and in particular mitotic count assessment.

REFERENCES

1. Kayser K. Introduction of virtual microscopy in routine surgical pathology--a hypothesis and personal view from Europe. *Diagnostic pathology.* Apr 30 2012;7:48.

2. Allen TC. Digital pathology and federalism. *Arch Pathol Lab Med.* Feb 2014;138(2):162-165.

3. Hedvat CV. Digital microscopy: past, present, and future. *Arch Pathol Lab Med.* Nov 2010;134(11):1666-1670.

4. Rocha R, Vassallo J, Soares F, Miller K, Gobbi H. Digital slides: present status of a tool for consultation, teaching, and quality control in pathology. *Pathology, research and practice.* 2009;205(11):735-741.

5. Brachtel E, Yagi Y. Digital imaging in pathology--current applications and challenges. *Journal of biophotonics.* Apr 2012;5(4):327-335.

6. Al-Janabi S, Huisman A, Van Diest PJ. Digital pathology: current status and future perspectives. *Histopathology.* Jul 2012;61(1):1-9.

7. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology.* Nov 1991;19(5):403-410.

8. Rakha EA, El-Sayed ME, Lee AH, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. *J Clin Oncol.* Jul 1 2008;26(19):3153-3158.

9. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res.* 2010;12(4):207.

10. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat.* 1992;22(3):207-219.

11. Wishart GC, Bajdik CD, Dicks E, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer.* Aug 21 2012;107(5):800-807.

12. Carlson RW, Brown E, Burstein HJ, et al. NCCN Task Force Report: Adjuvant Therapy for Breast Cancer. *J Natl Compr Canc Netw.* Mar 2006;4 Suppl 1:S1-26.

13. Curigliano G, Burstein HJ, E PW, 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.* Aug 1 2017;28(8):1700-1712.

14. Giuliano AE, Connolly JL, Edge SB, et al. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin.* Jul 8 2017;67(4):290-303.

15. American Joint Committee on Cancer (AJCC). AJCC Cancer Staging Manual. 8th ed. New York: Springer: 2017.

16. Rakha EA, Bennett RL, Coleman D, Pinder SE, Ellis IO, Pathology UKNCCfB. Review of the national external quality assessment (EQA) scheme for breast pathology in the UK. *J Clin Pathol.* Jan 2017;70(1):51-57.

17. Sloane JP, Amendoeira I, Apostolikas N, et al. Consistency achieved by 23 European pathologists from 12 countries in diagnosing breast disease and reporting prognostic features of carcinomas. *Virchows Archiv-an International Journal of Pathology.* 1999;434(1):3-10.

18. Lester SC, Bose S, Chen YY, et al. Protocol for the examination of specimens from patients with invasive carcinoma of the breast. *Arch Pathol Lab Med.* Oct 2009;133(10):1515-1538.

19. *R: A language and enviornment for statistical computing* [computer program]. R Foundation for Statistical Computing 2013.

20. McHugh ML. The chi-square test of independence. *Biochemia medica.* 2013;23(2):143-149.

21. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology.* Aug 2003;228(2):303-308.

22. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med.* Dec 2013;137(12):1710-1722.

23. Shaw EC, Hanby AM, Wheeler K, et al. Observer agreement comparing the use of virtual slides with glass slides in the pathology review component of the POSH breast cancer cohort study. *J Clin Pathol.* May 2012;65(5):403-408.

24. Haibe-Kains B, Desmedt C, Loi S, et al. A three-gene model to robustly identify breast cancer molecular subtypes. *J Natl Cancer Inst.* Feb 22 2012;104(4):311-325.

25. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med.* Aug 10 2006;355(6):560-569.

26. Bartlett JM, Bayani J, Marshall A, et al. Comparing Breast Cancer Multiparameter Tests in the OPTIMA Prelim Trial: No Test Is More Equal Than the Others. *J Natl Cancer Inst.* Sep 2016;108(9).

27. 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(5):e37379.

28. Zhong F, Bi R, Yu B, Yang F, Yang W, Shui R. A Comparison of Visual Assessment and Automated Digital Image Analysis of Ki67 Labeling Index in Breast Cancer. *PLoS One.* 2016;11(2):e0150505.

29. Schuh F, Biazus JV, Resetkova E, et al. Histopathological grading of breast ductal carcinoma in situ: validation of a web-based survey through intra-observer reproducibility analysis. *Diagn Pathol.* Jul 10 2015;10:93.

30. Dalton LW, Gerds TA. The Advantage of Discordance: An Example Using the Highly Subjective Nuclear Grading of Breast Cancer. *Am J Surg Pathol.* Aug 2017;41(8):1105-1111.

Acknowledgment

We thank the Nottingham Health Science Biobank and Breast Cancer Now Tissue Bank for the provision of tissue samples.

Licence for Publication

Competing Interest

None declared.

Contributorship

All authors contributed to this study and all approved the final manuscript

Funding information

No funding was received for this study

Ethical approval information

This study was approved by Nottingham Research Ethics Committee 2 under the title ‘Development of a molecular genetic classification of breast cancer’, and in compliance with current ethical and legal guidelines of the United Kingdom

Data sharing statement

Data are available upon request and at the discretion of the authors.

Tables

Table 1: Characteristics of the breast cancer cohort

Parameters	Number of cases (%)
Age	
>50	1098 (65.6)
≤50	549 (32.8)
Unknown	28 (1.7)
Tumour size	
> 2.0cm	588 (35.1)
≤2.0cm	1058 (63.2)
Unknown	29 (1.7)
Lympho-vascular Invasion	
Negative	1197 (71.5)
Positive	450 (26.9)
Unknown	28 (1.7)
Lymph node status	
Negative	1132 (67.6)
Positive	515 (30.7)
Unknown	28 (1.7)
Lymph Node Stage	
1	1027 (62.4)
2	457 (27.3)
3	162 (9.7)
Unknown	29 (1.7)
Nottingham Prognostic Index	
Good	568 (33.9)
Moderate	820 (49)
Poor	256 (15.3)
Unknown	31 (1.9)
Histologic types	
Ductal NST	1258 (75.1)
Lobular	102 (6.1)
Tubular/Invasive Cribriform	60 (3.6)
Pure Mucinous	22 (1.3)
Invasive Micropapillary	13 (0.8)
Other types including Medullary-like	220 (13.1)
Distant metastasis	
Yes	357 (21.3)
No	1288 (76.9)
Unknown	30 (1.6)
Outcome Status at end of follow-up	
Alive	1190 (71)
Died from Breast cancer	297 (17.7)
Died from other causes	156 (9.3)
Unknown	32 (1.9)

Table 2: Cross comparison of Nottingham grade (2a) and grade component scores (2b-d) between virtual microscopy and traditional light microscopy

Table 2a

Grade **(Virtual Microscopy)**	**Grade (Light Microscopy)**			**Total percentage**
	Grade 1	**Grade 2**	**Grade 3**	
Grade 1	232	215	25	**28.2%**
Grade 2	39	420	213	**40.1%**
Grade 3	1	48	482	**31.7%**
Total Percentage	**16.2%**	**40.8%**	**43%**	**100%**
Percent exact agreement: 68%, Percent adjacent level: 30.5%, Percent high/low: 1.5%				

Table 2b

Tubule formation **(Virtual Microscopy)**	**Tubule formation (Light Microscopy)**			**Total percentage**
	Score 1	**Score 2**	**Score 3**	
Score 1	43	13	2	**3.5%**
Score 2	67	231	34	**19.8%**
Score 3	12	265	1008	**76.7%**
Total Percentage	**7.3%**	**30.4%**	**62.3%**	**100%**
Percent exact agreement: 76.6%, Percent adjacent level: 22.6%, Percent high/low: 0.8%				

Table 2c

Pleomorphism **(Virtual Microscopy)**	**Pleomorphism (Light Microscope)**			**Total Percentage**
	Score 1	**Score 2**	**Score 3**	
Score 1	18	210	69	**17.7%**
Score 2	6	250	336	**35.3%**
Score 3	0	48	738	**46.9%**
Total Percentage	**1.4%**	**30.3%**	**68.3%**	**100%**
Percent exact agreement: 60.1%, Percent adjacent level: 35.8%, Percent hi/low: 4.1%				

Table 2d

Mitotic counts **(Virtual Microscopy)**	**Mitotic counts (Light Microscope)**			**Total Percentage**
	Score 1	**Score 2**	**Score 3**	
Score 1	807	187	102	**65.4%**
Score 2	34	75	136	**14.7%**
Score 3	5	49	280	**19.9%**
Total Percentage	**50.5%**	**18.6%**	**30.9%**	**100%**
Percent exact agreement: 69.4%, Percent adjacent level: 24.3%, Percent hi/low: 6.3%				

*First Virtual scores (V1) were used here

Table 3: Concordance between light microscopy grade and its component scores with virtual microscopy grade and its component scores assessed using *Cramer's V* and *kappa* statistic.

Parameters	***Cramer's V* statistic**		***kappa* statistic**		
	Cramer's V	***Phi***	***Kappa***	**Confidence Interval**	**Z-value**
Grade	0.58	0.82	0.51	0.47 – 0.54	28.4
Mitosis scores	0.51	0.73	0.46	0.43 – 0.50	25.6
Tubules scores	0.53	0.75	0.48	0.44 – 0.52	23.1
Pleomorphism scores	0.41	0.58	0.27	0.24 – 0.31	14.1

Table 4: Cross comparison of the sum of grade components between virtual microscopy and light microscopy

Virtual Microscopy	**Light Microscopy**							**Total**
	3	**4**	**5**	**6**	**7**	**8**	**9**	
Score 3	7	26	5	3	1	0	0	**42**
Score 4	8	81	15	22	4	1	2	**133**
Score 5	7	88	86	74	19	14	9	**297**
Score 6	1	23	126	156	16	41	44	**407**
Score 7	0	9	22	88	17	72	57	**265**
Score 8	0	1	0	25	11	56	122	**215**
Score 9	0	0	3	1	5	43	264	**316**
Total*	**23**	**228**	**257**	**369**	**73**	**227**	**498**	**1675**

****p**<0.00001**

Table 5: Concordance of virtual microscopy grade and its component scores between first and second session of virtual scoring (intra-observer agreement of grade using virtual microscopy)

Parameters	***Cramer's V* statistic**		***kappa* statistic**		
	Cramer's V	***Phi***	***Kappa***	**Confidence Interval**	**Z-value**
Grade	0.68	0.96	0.65	0.60-0.68	37.3
Mitosis scores	0.58	0.82	0.60	0.56-0.63	32.7
Tubules scores	0.63	0.89	0.64	0.60-0.68	29.9
Pleomorphism scores	0.57	0.81	0.56	0.52-0.59	30.4

Table 6: Association between outcome, in terms of breast cancer specific survival (BCSS) and distant metastasis free survival (DMFS), and histologic grade; as assessed by virtual microscopy (VM) and light microscopy (LM)

Parameter	**BCSS**				**DMFI**			
	VM grading		**LM grading**		**VM grading**		**LM grading**	
	HR (95%CI)	***p*-value**	**HR (95%CI)**	***p*-value**	**HR (95%CI)**	***p*-value**	**HR (95%CI)**	***p*-value**
Grade*	**1.9 (1.6-2.3)**	**1x10^{-13}**	**2.4 (2.0-3.0)**	**1x10^{-13}**	**1.8 (1.5-2.1)**	**1x10^{-13}**	**2.1 (1.8-2.5)**	**1.x10^{-13}**
Tubules *	**2.8 (1.9-4)**	**5.9x10^{-9}**	**1.9 (1.5-2.4)**	**9.58x10^{-9}**	**2.6 (1.9-3.6)**	**4.9 x10^{-10}**	**1.7 (1.4-2.1)**	**4.6x10^{-8}**
Pleomorphism*	**1.8 (1.5-2.2)**	**6.2x10-^{11}**	**2.7 (2-3.7)**	**1x10^{-10}**	**1.6 (1.3-1.8)**	**1.6x10^{-9}**	**2.2 (1.7-2.9)**	**6.1x10^{-10}**
Mitosis*	**1.5 (1.3-1.7)**	**2.5x10^{-11}**	**1.7 (1.5-1.9)**	**1x10^{-13}**	**1.4 (1.3-1.6)**	**9.1x10^{-11}**	**1.6 (1.4-1.8)**	**1.0x10^{-13}**
Grade**	**1.9 (1.6-2.3)**	**1x10^{-13}**			**1.8 (1.5-2.1)**	**1.0x10^{-13}**		
Tubules**	**2.2 (1.6-2.9)**	**1.32x10^{-8}**			**2.1 (1.6-2.6)**	**3.6x10^{-9}**		
Pleomorphism**	**1.8 (1.5-2.2)**	**3.9x10^{-10}**			**1.6 (1.3-1.9)**	**7.42x10^{-9}**		
Mitosis**	**1.8 (1.5-2.2)**	**2.7x10^{-10}**			**1.4 (1.2-1.6)**	**2.2x10^{-9}**		

* VM grading first session (November 2016).
** VM grading second session (February 2017).

Table 7: Multivariate Cox proportional hazard analysis for predictors of breast cancer specific survival (BCSS) and distant metastasis free survival (DMFS) for histologic grade; as assessed by light microscopy and virtual microscopy sessions 1 and 2.

Variables	**BCSS**		**DMFS**	
	***p*-value**	**HR (95%CI)**	***p*-value**	**HR (95%CI)**
Light Microscopy				
Grade	< 0.001	1.9 (1.5- 2.3)	< 0.001	1.6 (1.4 – 2.0)
Size	0.004	1.4 (1.1- 1.8)	< 0.001	1.6 (1.3 – 2.0)
Nodal Stage	< 0.001	2.1 (1.5- 2.3)	< 0.001	2.0 (1.8 – 2.4)
ER status	0.002	0.7 (0.5- 0.9)	0.024	0.7 (0.6 – 0.9)
Virtual Microscopy 1				
Grade	< 0.001	1.7 (1.4- 2.0)	< 0.001	1.6 (1.3 – 1.8)
Size	0.002	1.5 (1.2- 1.9)	< 0.001	1.6 (1.3 – 2.0)
Nodal Stage	< 0.001	2.2 (2.0- 2.6)	< 0.001	2.1 (1.9 – 2.5)
ER status	0.010	0.7 (0.5- 0.9)	0.102	0.8 (0.6 – 1.0)
Virtual Microscopy 2				
Grade	< 0.001	1.7 (1.4- 2.0)	< 0.001	1.6 (1.3 – 1.8)
Size	0.005	1.4 (1.1- 1.8)	< 0.001	1.6 (1.3 – 2.0)
Nodal Stage	< 0.001	2.3 (1.9- 2.6)	< 0.001	2.1 (1.9 – 2.4)
ER status	0.010	0.7 (0.5- 0.9)	0.107	0.8 (0.6 – 1.1)

ER=Estrogen receptor

Figure legend

Figure 1: Association between histologic grade as assessed using digital slide imaging and traditional light microscope and breast cancer specific survival (BCSS)

Figure 2: Association between histologic grade components as assessed using digital slide image traditional light microscope and breast cancer specific survival (BCSS); tubule formation: a & b, pleomorphism: c & d and mitotic scores: e & f.