

Effect of experience of pathologist on the inter-observer reproducibility of the Nottingham grade

Thusharie Liyanage, Lakmini Mudduwa

Department of Pathology, Faculty of Medicine, Karapitiya, Galle, Sri Lanka

Corresponding author: Dr. Thusharie Liyanage (thusharieliyanage@yahoo.com)

ABSTRACT

Introduction: Breast carcinoma is the most common solid malignant tumour. It is the commonest cancer among Sri Lankan females. Tumour grading is an important prognostic factor, determining the future management of the patient. Grading of the tumour should be consistent among the pathologists without significant inter-observer variation. This study examined whether the experience of pathologist has any effect on the inter-observer agreement of the histological grading system of invasive breast carcinoma.

Method: Two pathologists with different levels of experience in histopathology were involved in the study. They independently evaluated histological grade using the Nottingham modification of the Bloom and Richardson histological grading system on 47 cases of carcinoma of breast.

Results: Generalized kappa value for overall Nottingham grading was 0.61. Generalized kappa values were 0.5 and 0.51 for tubule formation and mitotic count respectively. The generalized kappa value for nuclear pleomorphism was 0.63.

Conclusion: The difference in the level of experience does not affect the final Nottingham grade of breast carcinoma.

Key words: *Nottingham grading, Inter-observer variation*

Introduction

Breast carcinoma is the most common solid malignant tumour and the leading cause of cancer death in women world over (1,2). In 2001 almost 240,000 women were diagnosed with breast cancer and over 40,000 died of the disease (3). It is the commonest cancer among Sri Lankan females as well (4).

The prognosis of breast carcinoma is determined by the pathological examination of the primary carcinoma and the axillary lymph nodes. The most important prognostic factor of breast carcinoma is the tumour stage at the time of diagnosis (3). For staging, the tumour size, lymph node metastases and distant metastases are assessed. Additional prognostic information is given by the histological subtype, histological grade, the presence or absence of lymphovascular invasion, oestrogen and progesterone receptor status, Her-2/neu status, p53

status, Ki 67, MIB 1 labeling and angiogenesis (2,5,6).

Tumour grade is one of the important prognostic factors which indicates the extent of tumour differentiation (1,3,7). The most widely used system over the years for microscopic grading of breast carcinoma is the Nottingham modification of the Bloom and Richardson system (1). It relies on architectural and cytological features, where three components of tumour morphology are assessed (1,5,7). Tubule formation, nuclear pleomorphism and the frequency of mitosis are assessed and given a score ranging from 1-3. The three scores are added together to give the overall histological grade.

In Sri Lanka the Nottingham grading system has been in use in the reporting of breast carcinomas for the past decade. As the tumour grade is used as an important parameter in making management decisions, it is important that the grading is

consistent among pathologists. An ideal grading system should have good reproducibility, which should not be affected by the experience of the user.

The subjective nature of a grading system can be minimized by introducing more precise definitions for the assignment of points within each category of morphological features assessed (8). When grading systems with specified guidelines are used, the reproducibility was found to be acceptable (9,10). This is one of the reasons the grading of breast carcinoma, with the introduction of Modified Bloom and Richardson system has received worldwide acceptance. Today this is the most widely used histological grading system for breast cancer (7,8,11). Over the past decades, studies have confirmed the usefulness and reproducibility of the Nottingham grading system (10,12,13,14) and it has been adopted for use in the breast cancer pathological reporting minimum dataset of the United Kingdom (15).

Bloom and Richardson initially introduced a grading system for breast carcinoma, which was later modified by Elston and Ellis and recognized as the Nottingham modification. They introduced clearly defined criteria, particularly by giving numerical limits to the measurement of tubule formation and mitotic counts (12). The modified system differs from the original Bloom and Richardson system mostly in the assessment of mitotic counts. Bloom and Richardson analyzed the relative numbers of both hyperchromatic nuclei and mitotic figures (8,12). In the modified grading system hyperchromatic nuclei and pyknotic nuclei were excluded from the mitotic count and only clearly identifiable mitotic figures are included (12). By excluding cells of prophase, possible confusion with apoptotic cells and intra-tumoural lymphocytes is avoided (8).

With the modified system, moderate to substantial agreement was found for histological grade in a study done in University of Virginia Health Science Center, USA in 1995. In this study, the median pairwise agreement was 71% (13). Studies using Bloom and Richardson system with its less precisely defined criteria have yielded agreements of 50.6%, 54%, 66%, 72% and 78% (13). However, there are no studies describing the effect of experience of the user on the inter-observer agreement.

Our study was undertaken to assess the effect of

user's experience on the inter-observer agreement in grading invasive breast carcinoma using Bloom and Richardson system with Nottingham modification and to assess the effect of the user's experience on the inter-observer agreement in scoring the 3 components of the grade.

Methods

All breast carcinomas diagnosed and reported by the second author, at the Department of Pathology, Faculty of Medicine, Galle, from January 2001 to September 2006 were included in the study. Both excision biopsies and mastectomy specimens were included.

Poorly fixed specimens were excluded as morphological features were not preserved adequately in these cases. All slides which were faded were restained with Haematoxylin and Eosin. All slides from each case were re-evaluated and graded by the authors independently. Both used the Nottingham modification of Bloom and Richardson grading system. They were blind to each other's grading and the original grading done at the time of reporting.

The level of experience in grading breast cancers was different among the two authors. One had independently graded breast carcinomas for over five years as a consultant histopathologist while the other had two and a half years experience of grading breast cancers as a postgraduate trainee in MD histopathology. The results were analyzed using the kappa coefficient of agreement between the two observers (16). Ethical approval was obtained from the Ethical Review Committee of the Faculty of Medicine, University of Ruhuna, Galle, prior to commencing the study.

Results

A total of 58 cases of breast carcinomas were retrieved from the files and four of them were found to be poorly fixed. The preservation of morphological features in these four specimens was not adequate for grading and they were excluded from the study. Two cases of mastectomy with no residual tumour and five cases of in-situ carcinoma were also excluded. After excluding unsuitable cases there were 47 cases of invasive carcinoma of breast for analysis of inter-observer variation. The vast

majority of patients had invasive duct carcinoma, not otherwise specified. There were a few special types of carcinoma which included one papillary carcinoma, two mucinous carcinomas and four lobular carcinomas.

The inter-observer variation of the overall histological grade and the three components of the Nottingham grade were assessed separately. Kappa value for the overall grading was 0.61 indicating substantial agreement.

Kappa values were 0.5 and 0.51 for tubule formation and mitotic count respectively revealing a moderate agreement. The agreement between the 2 observers was best in scoring pleomorphism which was 0.63 indicating a substantial agreement.

Discussion

Breast cancer treatment has become more complex over time. As treatment options increase the accuracy of prognostic information increases in importance since many therapeutic decisions are based on such data (8,11). In comparison with other prognostic parameters, the evaluation of histological grade is cheap and can be easily performed in all cases of breast cancer (9). The major disadvantage in histopathological grading in breast carcinoma is the subjective nature of the procedure and the problems in consistency and reproducibility (8,9). This leads to inter-observer variation as well as intra observer variation in results.

To obtain best results in grading and higher levels of inter-observer agreement, it is very important that careful attention is paid to specimen preparation. Fixation should be optimal (8,14). Autolysis artifacts should be kept to a minimum by slicing the specimen in the fresh state immediately after resection. The practice of immersion of the whole breast unsliced, in formalin should be discouraged as it leads to poor preservation of tissue, as well as completion of mitoses, which may give an erroneously wrong value during grading. The histological sections should be 4-5 μ thick. If the sections are too thick, the nuclear details will be obscured and will influence the nuclear grading as well as the mitotic count. Staining with Haematoxylin and Eosin is adequate and no special stains are required for grading of breast carcinomas.

The tumor grade is achieved as a composite score. Even though the pathologists may have disagreement in individual components sometimes they can come up with the same grade for the carcinoma, as it is the composite score which decides which grade is attributed to the individual tumour (7). Our study also supports this finding. It was also revealed in a study performed in Texas USA, where there was greater than 87% agreement by the pathologists as to the final combined Bloom and Richardson grade (10). It has been found that when pathologists disagree, they tend not to disagree by much (17).

Several studies done before have revealed that nuclear pleomorphism is the least reproducible of the three components (8,13). However if strict criteria are used, as we have done in this study, substantial agreement can be achieved for nuclear pleomorphism as well (7). The present study shows substantial agreement in assigning a score for nuclear pleomorphism with strict adherence to criteria given in the Nottingham modification.

Fixation of the specimen has an effect on identification of tubule formation and mitotic count. When the specimen is not fixed properly the tumour shows fixation artifacts like separation of tumour cell clusters from the stroma, single cell separation and condensation of the nuclei. It is very difficult to assign a correct score to tubule formation with fixation artifacts. When the nuclei are condensed it will adversely affect the mitotic count as normal mitoses will be hidden resulting in an underscoring of the mitotic count.

High degree of agreement between two individual observers is achieved by pathologists who are working in the same department very closely in routine practice, often presenting each other's work at meetings with continuous feedback (18). This also may explain the substantial agreement achieved in the present study for nuclear pleomorphism, as both authors work in the same department with continuous feedback of each other's work.

A study done in Japan revealed that repeated slide conferences were an effective means of standardizing the subjective histopathological criteria used to assess tumours. The level of agreement did tend to increase as the number of repeated slide conferences increased (19). It was shown that excellent agreement of histological grade

was more likely for extremely low grade and extremely high grade tumours. If a tumour is of intermediate nature, half of the observers will assign a score of 2 and half a score of 3. This will lead to inter-observer variation and such a case would be judged as having a concordance of only 50% or less in terms of percentage agreement or by kappa statistics (19).

The results of the present study revealed that the final Nottingham grade had substantial agreement with a kappa score of 0.61. The agreement was moderate in assigning a score to tubule formation and mitotic count. However, a substantial agreement was present in assigning a score to the nuclear pleomorphism.

In our study there was only moderate agreement on tubule formation with a kappa value of 0.5. This was mainly because the first author had erroneously calculated clefts within the tumour as tubules. Due to this reason many tumours which the second author had scored as 3 (Tubule formation < 10% tumour) the first author has assigned a score of 2 (Tubule formation 10-75% of tumour.) This can be attributed to the lesser experience that the first author had compared to the second author. Clear lumina must be visible and shrinkage artifacts due to clefts should not be mistakenly attributed to tubules (8).

Several studies have shown that mitotic count is the most important constituent of histological grade (2), but there are well known problems with the reproducibility of the mitotic count. In our study there was moderate agreement on mitotic count with a kappa value of 0.51. The variables responsible for differences among observers in mitotic counts are numerous (13). This include quality of staining, quality and the thickness of the section, non-random distribution of mitotic figures within the histological section, tumour cell size, properties of tumour such as necrosis, severe inflammation, fibrosis, and the amount of stroma, and the microscopic field area (9,13,20). Interobserver variation in mitotic counts may also occur from variation in the quality of specimens, as well as from the properties of the light microscope, and the intermediate nature of the tumour itself (21). Mitotic counts will be affected by delay in fixation, and therefore it is advisable to avoid delays in fixation whenever possible, and to keep the specimens in the refrigerator until fixation (2).

Other measures which will lead to increase in the agreement between pathologists are limiting the number of slides per case, identifying specific areas on the slides to be evaluated, issuing a training set of slides, and holding microscopic teaching sessions for all participants (13). However such measures would be more artificial and less likely to reflect the “everyday” practice of grading breast cancer.

Conclusion

The study indicates that there can be minimal variation in assigning a score to the individual three components of Nottingham grading between the two users due to difference in the level of experience. However, this variation does not affect the final grade of the breast carcinoma.

Acknowledgements

We wish to thank Dr. W.A.A. Wijayasiri, Dr. Nayana Pallethanna Department of Community Medicine for the advice given on the statistical analysis.

I also wish to thank all the staff members in the histopathology laboratory, who helped me in re-staining the faded slides.

References

1. Breast. In: Rosai J. Ackerman's Surgical Pathology, 9th ed. New York. Mosby year book, 2004: 1763-1876.
2. Van Diest PJ, Van der Wall E, Baak JPA. Prognostic value of proliferation in invasive breast cancer. *Journal of Clinical Pathology* 2004; **57**: 675-681.
3. Susan CL. The breast. In: Vinay K, Abdul KA, Nelson F, Robbins pathologic basis of disease. 7th ed. Philadelphia: Saunders 2004: 1119-1154.
4. Cancer Incidence 1995 Sri Lanka, Cancer Registry, National Cancer Control Programme 2002.
5. Simpson JF, Gray R, Dressler LG *et al*, Prognostic value of Histologic grade and proliferative activity in axillary node positive breast cancer: Results from the Eastern Cooperative Oncology Group Companion Study, EST 4189. *Journal of Clinical Oncology* 2000; **18**: 2059-2069.
6. Schnitt SJ. Traditional and newer pathologic factors. *Journal of the National Cancer Institute Monograph* 2001; **30**: 22-26.

7. Sloane JP, Anderson TJ, Davies JD *et al.* Pathology reporting in breast cancer screening, National coordinating group for breast screening pathology. 2nd ed. Sheffield NHSBSP, 1995.
8. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. The value of histological grade in breast cancer: experience from a large study with long term follow up. *Histopathology* 1991; **19**: 403-410.
9. Boiesen P, Bendahl PO, Anagnostaki L *et al.* Histologic grading in breast cancer. Reproducibility between seven pathologic departments. *Acta Oncologica* 2000; **39**: 41-45.
10. Dalton LW, Page DL, Dupont WD. Histologic grading of breast carcinoma. A reproducibility study. *Cancer* 1994; **73**: 2765-2770.
11. Monticciolo DL. Histologic grading at Breast core needle biopsy: Comparison with results from the excised breast specimen. *The Breast Journal* 2005; **11**: 9-14.
12. Frkovic-Grazio S, Bracko M. Long term prognostic value of Nottingham histological grade and its components in early (pT1N0M0) breast carcinoma. *Journal of Clinical Pathology* 2002; **55**: 88-92.
13. Frierson HF, Wolber RA, Berean KW, Franquemont DW, Gaffey MJ. Interobserver reproducibility of the Nottingham modification of the Bloom and Richardson histologic grading scheme for infiltrating ductal carcinoma. *American Journal of Clinical Pathology* 1995; **103**: 195-198.
14. Robbins P, Pinder S, de Klerk N, Dawkins H, Harvey J. Histological grading of breast carcinomas: a study of interobserver agreement. *Hum.Pathol.* 1995; **26**: 873-879.
15. Royal College of Pathologists Working Group on Cancer Services. Standard and Minimum data sets for reporting common cancers. Minimum data set for breast cancer Histopathology reports. London: Royal College of Pathologists, 1998.
16. Practical statistics for medical research. Douglas GA. 1st ed. Chapman & Hall/CRC London, 1999: 403-409.
17. Dalton LW, Pinder SE, Elston CE, Ellis IO, Page DL. Histologic grading of breast cancer: linkage of patient outcome with level of pathologist agreement. *Mod Pathology* 2000; **13**: 730-735.
18. Douglas-Jones AG, Morgan JM, Appleton MAC *et al.* Consistency in the observation of features used to classify duct carcinoma in situ (DCIS) of the breast. *J Clinical Pathology* 2000; **53**: 596-602.
19. Tsuda H, Akiyama F, Kurosumi M *et al.* The efficacy and limitations of repeated slide conferences for improving Interobserver agreement when judging nuclear atypia of breast cancer. *Japanese Journal of Clinical Oncology*, **29**: 68-73.
20. Tsuda H, Akiyama F, Kurosumi M *et al.* Evaluation of the Inter observer agreement in the number of mitotic figures of breast carcinoma as simulation of quality monitoring in the Japanese National Surgical Adjuvant Study of Breast Cancer (NSAS-BC) protocol. *Japanese Journal of Cancer Res.* 2000; **91**: 451-457.
21. Tsuda H, Akiyama F, Kurosumi M *et al.* Establishment of histological criteria for high risk node negative breast carcinoma for a multi-institutional randomized clinical trial of adjuvant therapy. *Japanese Journal of Clinical Oncology* 1998; **28**: 486-491.