

IMMUNOHISTOCHEMICAL ASSESSMENT OF HORMONE RECEPTOR STATUS OF BREAST CARCINOMA: INTEROBSERVER VARIATION OF THE QUICK SCORE

LAKMINI MUDDUWA, THUSHARIE LIYANAGE

ABSTRACT

BACKGROUND: Immunohistochemical (IHC) assessment of estrogen receptor (ER) and progesterone receptor (PR) status has become a routine practice to predict the likely outcome of Tamoxifen therapy. **AIMS:** To assess the interobserver variation in scoring hormone receptor status of breast carcinoma, using the Quick Score. **MATERIALS AND METHODS:** IHC-stained slides of breast carcinomas reported by the two authors during a 28-month period were included in the study. Both authors independently reassessed all the tumors. Both were blinded to each other's assessment. The Quick score with a 0-8 point scale was used to score the hormone receptor status. Weighted Kappa was calculated to assess the interobserver variation. **RESULTS:** A total of 210 breast carcinomas were included in this study. There was a substantial to almost perfect agreement between the two observers in scoring the hormone receptor status (kappa values; ER = 0.856, PR = 0.711). Both ER and PR showed an almost perfect agreement in assessing the intensity of staining (kappa value; ER = 0.882, PR = 0.840), while the scoring of proportion of cells gave lower Kappa values (kappa value; ER = 0.778, PR = 0.592). Interobserver agreement was less in scoring hormone receptor status of breast carcinomas after mastectomies compared with excision biopsies, wide local excisions and metastatic deposits in lymph nodes. Suboptimal fixation resulting in background staining has contributed to the variation. **CONCLUSION:** A substantial to almost perfect interobserver agreement was seen in assigning an overall Quick score. Detection of complete negative and strong expression had a moderate to substantial agreement.

Key words: Breast carcinoma, hormone receptors, interobserver variation

DOI: 10.4103/0019-5359.49078

INTRODUCTION

The assessment of estrogen receptor (ER) status of breast carcinoma has become the

routine practice to predict the likely outcome of Tamoxifen therapy. The assessment of progesterone receptor (PR) status along with ER gives a stronger predictive power.^[1] Hormone receptor status of breast carcinoma was earlier assessed by biochemical methods. However, with recent advances in diagnostic techniques, more and more small biopsies are submitted for histopathological diagnosis.

Department of Pathology, Faculty of Medicine,
University of Ruhuna, Galle, Sri Lanka

Correspondence:

Dr. Lakmini Mudduwa, Department of Pathology, Faculty of
Medicine, University of Ruhuna, Galle, Sri Lanka.
E-mail: lakminimudduwa@yahoo.com

This has rendered biochemical assays less useful in the routine practice as the entire tumor may have to be sampled for accurate diagnosis, leaving no tissue for the biochemical assay.^[2] The tissue included in the homogenate for the biochemical assay may have benign breast tissue and fibroelastic tissue together with neoplastic tissue in different proportions, masking the true level of hormone expression in the tumor.^[2] Further, the immunohistochemical (IHC) assessment has replaced biochemical assay of hormone receptor status, which can even be done on tiny biopsies. IHC assessment permits visualization of the tumor cells expressing hormone receptors and gives the opportunity to semi-quantify against the cells that are negative for hormone receptors. IHC assessment is considerably less expensive compared to the biochemical assay and can be done on routinely processed tissue sections.^[3] With the widespread use of IHC assay, the need for effective quality assurance program has been emphasized. The United Kingdom National External Quality Assessment Scheme for Immunocytochemistry (UKNEQAS-ICC) is one such program, which mainly aims at the interlaboratory variation, thus the variation in the technical aspects.^[1,3]

There are many scoring systems in use, to semi-quantify the assessment of hormone receptor status of breast carcinoma on IHC slides. Some laboratories use H score, which is one of the most complex.^[2] The proportion of cells stained alone as a scoring system is also in use disregarding the intensity of staining. However, the Quick score, which considers both the proportion of cells and the intensity of staining, is used by many laboratories in

the United Kingdom as well as in mainland European laboratories.^[4] These laboratories have found that results obtained from Quick score correlates well with the biochemical assays and provides significant predictive and prognostic information.^[4] Benefits of adjuvant endocrine therapy is proportional to the quantity of receptors present.^[4] Therefore, having a scoring system that semi-quantifies the expression of hormone receptors is valuable to the oncologist in predicting the chances of response.

Although the interlaboratory variation in IHC assessment of hormone receptor status has been extensively studied, addressing the technical variation, there are only a few studies available on the interobserver variation in scoring.^[2] Therefore, the aim of the present study was to assess the interobserver variation in assigning a score for the hormone receptor status of breast carcinoma, using the Quick Score, which has been recommended in some national guidelines.^[2]

MATERIALS AND METHODS

ER and PR status of breast carcinomas reported by the two authors in our laboratory between June 2006 and October 2008 were included in the study. Both authors independently re-assessed the ER and PR status of each case. They were blinded to the previous score and the each others' reassessment. The Quick score was used to give a score to each tumor, as it was the routine scoring system used in our laboratory. At the time of reporting of ER and PR status of each case, one representative block from several blocks with tumor (mastectomies, wide

local excisions and excision biopsies) or the single block available for Tru-cut biopsies had been selected for the assessment of ER and PR status. Therefore, each case included in the study had one slide each for ER and PR. The staining was evaluated on the invasive component only. Best-preserved and best-stained areas of the sections were assessed. A minimum of 100 cells were counted for each case. Only nuclear staining was considered as positive staining. A score for the proportion of stained cells (0 = no nuclear staining, 1 = <1% nuclear staining, 2 = 1%-10% nuclear staining, 3 = 11%-33% nuclear staining, 4 = 34%-66% nuclear staining and 5 = 67%-100% nuclear staining) and the intensity of staining (0 = no staining, 1 = weak staining, 2 = moderate staining and 3 = strong staining) were assigned to each tumor. The score for the proportion of cells stained and the score for the intensity of staining were added to obtain the total score, which range from 0 to 8. However, a score of 1 was not given as the overall Quick score is the sum of two scores. The authors adhered to the guidelines given in the working protocol published by Leak *et al*/ while assigning scores.^[4]

The routine laboratory protocol of staining had been adhered to stain all the blocks. From the selected blocks, 4- μ -thick sections were taken on poly-L-lysine-coated slides. IHC staining was performed using the Streptavidin Biotin method. Dako (Glostrup, Denmark) polyclonal rabbit antihuman progesterone receptor (A0098) and monoclonal mouse antihuman estrogen receptor α , clone 1D5 (M7047) and Universal LSAB2 kit/HRP Rabbit/mouse with Streptavidin/HRP (K0675) were used for IHC staining, which was performed manually by an experienced technical officer.

The interobserver variation was assessed using Kappa statistics. Weighted Kappa values were calculated for the total scores and scores for the percentage of cells stained and the intensity of staining for both ER and PR. As many clinicians empirically use the 10% of stained cells as the cut-off, Kappa value was calculated to assess the interobserver variability in identifying a total score of 4 or more (a Quick score of 4 tallies with 10% weakly stained cells). For this step, weighting was not done, as the Kappa values were calculated for two raters against two options. Kappa values were calculated for three categories of specimens received (1 – mastectomies; 2 - wide local excisions, excision biopsies and lymph node deposits; and 3 – Tru-cut biopsies) to evaluate the effect of type of surgery, as a surrogate to assess the effect of fixation on interobserver consistency. The cases with a discrepancy of ≥ 3 in the overall score, between the two observers were reassessed once again by the two observers at a double-headed microscope to find out the possible reasons for disagreement. This study was conducted following the approval obtained from the Ethical Review Committee of the institution.

RESULTS

The total number of breast carcinomas included in the study was 210. ER slides were available for all 210 cases but for PR only 204 were available. There were 146 (69.5%) mastectomy specimens, 40 (19.0%) excision biopsies, 13 (6.19%) Tru cut biopsies, 4 (1.9%) wide local excisions and 7 (3.3%) lymph node deposits. Nine laboratories including our own had referred wax blocks of breast carcinomas for IHC assessment of the hormone receptor status.

The overall Quick Score for both ER and PR had a substantial to almost perfect agreement [Table 1]. Of the two components of the Quick score, intensity of staining had a better interobserver agreement. A substantial agreement was seen in identifying high levels of hormone receptor expression, while identification of completely negative tumors had a substantial (ER) to moderate (PR) agreement. There was a substantial (ER) to moderate (PR) agreement at a cut off of ≥ 4 [Table 2].

Weighted Kappa values of the interobserver agreement for two of three groups of specimens are given in Table 3. The weighted Kappa was not calculated for the group that included Tru cut biopsies, as the total number of Tru cut biopsies was only 13. Therefore, the observers have not utilized the full range of the score for this group of specimens. There is a significantly better agreement in assigning

scores for the group 2 specimens, which included excision biopsies, wide local excisions and lymph node deposits compared with group 1 (mastectomies).

A discrepancy of 3 or more in the total score was observed in 47 cases of PR and 29 cases of ER. A discrepancy of 5 or 6 was seen in 31 of the above mentioned cases. This variation was mostly due to the background staining in the specimen tissue. Minor enhancement observed in the nuclear outline on a slide with background staining has contributed to the disagreement in most cases.

DISCUSSION

Variation in the assessment of hormone receptor status of breast carcinoma has two components: a technical variation and an interpretive variation.^[2] A study conducted within the European Union revealed that

Table 1: Weighted Kappa values for overall and for each components of the Quick score

	<i>Kappa value</i>	<i>Strength of agreement*</i>
ER		
Overall score	0.856	Almost perfect
Score for intensity of staining	0.882	Almost perfect
Score for proportion of stained cells	0.778	Substantial
PR		
Overall score	0.711	Substantial
Score for intensity of staining	0.840	Almost perfect
Score for proportion of stained cells	0.592	Moderate

*Guidelines for strength of agreement.^[9]

Table 2: Kappa values calculated for the different levels of the Quick score

	<i>Kappa value</i>	<i>Strength of agreement*</i>
ER		
Identifying completely negative expression (0)	0.627	Substantial
Cut-off ≥ 4	0.686	Substantial
Identifying high levels (7-8)	0.801	Substantial
PR		
Identifying completely negative expression (0)	0.585	Moderate
Cut-off ≥ 4	0.529	Moderate
Identifying high levels (7-8)	0.644	Substantial

*Guidelines for strength of agreement.^[9]

Table 3: Kappa values calculated for the groups of specimens

	<i>Kappa value</i>	<i>Strength of agreement*</i>
ER (overall)		
Group 1: Mastectomy	0.801	Substantial
Group 2: Excision biopsies etc.	0.925	Almost perfect
Group 3: Tru cut biopsies†		
PR (overall)		
Group 1: Mastectomy	0.608	Moderate
Group 2: Excision biopsies etc.	0.841	Almost perfect
Group 3: Tru cut biopsies†		

*Guidelines for strength of agreement.^[9] †Kappa was not calculated, as the group included a limited number of specimens

interobserver agreement in using Quick Score within laboratories is fair (kappa value = 0.39). The two components of the Quick score, intensity of staining (kappa value = 0.23) and proportion of stained cells (kappa value = 0.40), also had a fair agreement. The present study shows considerably better agreement compared with the European Union study.^[2] However, the two studies are not identical, as the European Union study^[2] included many observers, whereas we have assessed the consistency between two observers. We have used the weighted Kappa, which reflects the actual agreement better than unweighted Kappa. The assessment of the intensity of staining had a better agreement over the proportion of cells probably because of the lower number of levels (0-3) considered for the assessment of intensity as compared with proportion of cells that is assessed at 6 levels (0-5).

The present study also shows a substantial agreement in identifying completely negative ER tumors and the tumors with high levels of ER expression. Therefore, the completely negative and high levels of ER expression can be relied upon for therapeutic decisions. For PR, agreement was only moderate for the recognition of completely negative tumors. The most probable reason is that background staining, which was a problem seen more

often with PR, as we had used a polyclonal antibody for PR for routine IHC staining. At the most commonly used empirical cut-off, 10% of stained cell (4 in Quick score) also showed a substantial agreement with ER. Therefore, the Quick score appears as a reliable scoring system at the therapeutic decision-making level. As the number of categories in a scoring system increases, the variation in the results also increases. Often the middle group of a scoring system shows more variation than the two extremes. However, this has not much affected the interobserver agreement, as reflected in the Kappa for ER (0.686) at a score of 4. Semi-quantifying hormone expression similar to the Quick score is also important to predict the chances of response to Tamoxifen therapy.^[4]

The exact threshold of ER and PR staining, which should be used to distinguish between endocrine-responsive tumors is unknown.^[5] Even very low hormone receptor expressing tumors are also found to have some response to endocrine therapy. It has been found that some shrinkage of tumor is seen in elderly patients even though ER is low.^[4] Therefore, 10% positively stained cells have been empirically selected as a reasonable threshold for definite endocrine therapy.^[5,6] This tallies with 4/8 (weakly positive in 10% of cells) of the Quick score.

Reassessment of the Quick score, using a multi-headed microscope revealed that suboptimal fixation resulting in background staining has significantly contributed to the variation in interpretation. This is substantiated by the kappa values for the specimens grouped according to the expected level of fixation [Table 3]. It is well known that the best preservation is seen in Tru cut biopsies, and the next best in the excision biopsies and wide local excisions, over the mastectomies. Our laboratory received wax blocks of breast carcinomas for IHC hormone receptor assessment from 9 laboratories, as ours is the only IHC laboratory in the province. These wax blocks of specimens had been subjected to varying durations of fixation and processing schedules and delayed fixation. These factors individually or in combination may have had an effect on the reliability of IHC assay.^[7] To minimize the effects of such factors, the IHC staining methodology can be adjusted for the in-house specimens, but it is difficult to optimize the methodology to make it suitable for 9 different laboratories.

The authors' experience is that the time taken to transport specimens from the operation theater to the laboratory varies from one hospital to another. Many specimens are sent unsliced and by the time it is sliced by the pathologist, the tumor autolysis has already started as the tissue penetration by formalin is slow. This affects the tumors in mastectomy specimens more than the excision biopsies or wide local excisions. The most superficial lymph nodes in an axillary clearance also are better fixed compared to the main tumor in the mastectomy specimen. Preparation of 10% buffered formalin is done unsupervised in some hospitals and may not comply with the standards. Some specimens,

e.g., small excision biopsies and Tru cut biopsies may be overfixed due to delays in the transport.

Suboptimal fixation causes background staining in specimen tissue, while the positive control shows appropriate specific staining. This could be due to specimen held for too long in formalin, which is a cross-linking fixative, causing "masking" of antigenic determinants due to aldehyde cross-linking and increased hydrophobicity of tissues.^[8] If the sectioned portion of tissue has not been well penetrated by the fixative, loss of antigenicity occurs.^[8] Both these factors may have contributed to the background staining in the cases where there was a significant disagreement. Polyclonal antibodies can have cross-reactivity, giving rise to non-specific staining, which further reduces the interobserver agreement.^[8]

CONCLUSION

The Quick score is a reliable method to score the hormone receptor status of breast carcinoma. There is a substantial to almost perfect agreement between observers in assigning an overall Quick score. Assessment of the proportion of stained cells with six levels reduces the consistency. Assessment of the intensity of staining with four levels shows a better concordance. Optimal fixation of tissues is essential to improve the interobserver agreement.

ACKNOWLEDGEMENTS

We thank Dr. Neil Thalagala, Consultant Community Physician, National Programme Manager - Child Health, Family Health Bureau, Sri Lanka, for assistance in the statistical analysis, Mrs. G. G. D. D. Gunawardhane for staining slides and Dr. N. J. Asanthi and Dr. L. R. L.

Liyanaarachchi of the Department of Pathology, Faculty of Medicine, University of Ruhuna, Sri Lanka, for assistance in the collection of data.

REFERENCES

1. Rhodes A, Jasani B, Balaton AJ, Barnes DM, Miller KD. Frequency of oestrogen and progesterone receptor positivity by immunohistochemical analysis in 7016 breast carcinomas: Correlation with patient age, assay sensitivity, threshold value, and mammographic screening. *J Clin Pathol* 2000;53:688-96.
2. Wells CA, Sloane JP, Coleman D, Munt C, Amendoeira I, Apostolikas N, *et al.* Consistency of staining and reporting of oestrogen receptor Immunohistochemistry within the European Union-an inter-laboratory study. *Virchows Arch* 2004;445:119-28.
3. Rhodes A, Jasani B, Barnes DM, Bobrow LG, Miller KD. Reliability of immunohistochemical demonstration of oestrogen receptors in routine practice: Interlaboratory variance in the sensitivity of detection and evaluation of scoring systems. *J Clin Pathol* 2000;53:125-30.
4. Leake R, Barnes D, Pinder S, Ellis I, Anderson L, Adomson R, *et al.* Immunohistochemical detection of steroid receptors in breast cancer: A working protocol. *J Clin Pathol* 2000;53:634-5.
5. Awada A, Castro Jr G. de. An integrated approach for tailored treatment in breast cancer. *Ann Oncol* 2005;16:ii203-8.
6. Pestalozzi BC, Luporsi-Gely E, Jost LM, Bergh J. ESMO Minimum clinical recommendations for diagnosis, adjuvant treatment and follow up of primary breast cancer. *Ann Oncol* 2005;16:i7-9.
7. Rhodes A, Jasani B, Balaton AJ, Miller KD. Immunohistochemical detection of oestrogen and progesterone receptors: Correlation of standards achieved on in house tumours with that achieved on external quality assessment material in over 150 laboratories from 26 countries. *J Clin Pathol* 2000;53:292-301.
8. Atwood KN. Trouble shooting. In Boenisch T, editor. *Immunohistochemical staining methods*. 3rd ed. California: Dako Cytomation; 2001. p. 47-62.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.

Source of Support: Nil, **Conflict of Interest:** None declared.

Staying in touch with the journal

1) Table of Contents (TOC) email alert

Receive an email alert containing the TOC when a new complete issue of the journal is made available online. To register for TOC alerts go to www.indianjmedsci.org/signup.asp.

2) RSS feeds

Really Simple Syndication (RSS) helps you to get alerts on new publication right on your desktop without going to the journal's website. You need a software (e.g. RSSReader, Feed Demon, FeedReader, My Yahoo!, NewsGator and NewzCrawler) to get advantage of this tool. RSS feeds can also be read through FireFox or Microsoft Outlook 2007. Once any of these small (and mostly free) software is installed, add www.indianjmedsci.org/rssfeed.asp as one of the feeds.