Available online at www.sciencedirect.com

ScienceDirect

Journal of the Chinese Medical Association 79 (2016) 649−655

www.jcma-online.com

Original Article

Adjust cut-off values of immunohistochemistry models to predict risk of distant recurrence in invasive breast carcinoma patients

Yen-Ying Chen [a,b], Ling-Ming Tseng [b,c], Ching-Fen Yang [a,b], Pei-Ju Lien [d], Chih-Yi Hsu [a,b,*]

[a] *Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, ROC*
[b] *National Yang-Ming University School of Medicine, Taipei, Taiwan, ROC*
[c] *Division of General Surgery, Department of Surgery, and Comprehensive Breast Health Center, Taipei Veterans General Hospital, Taipei, Taiwan, ROC*
[d] *Department of Nursing, Taipei Veterans General Hospital, Taipei, Taiwan, ROC*

Received February 16, 2016; accepted May 13, 2016

Abstract

Background: Multigene assays are recommended for hormone receptor-positive invasive breast carcinoma to determine the risk of recurrence, but they are highly expensive. We investigated the prognostic values of immunohistochemistry (IHC)-based prognostic models as an alternative to multigene assays.

Methods: The risk categories estimated by the IHC-based prognostic models were correlated to those estimated by the multigene assays in 71 cases and the follow-up results in 642 consecutive cases of HER2− luminal-type early breast cancer. Cut-off values of IHC-based models were adjusted based on survival outcome to reveal maximum Harrell C index or based on the maximum positive likelihood ratio correlated to multigene assay.

Results: All investigated IHC-based models could predict the risk of distant recurrence, but their cut-off values required adjustment. Using distant recurrence-free survival (DRFS) to refine the cut-off values could improve the prognostic values. Adjusting the cut-off values using the results of multigene assays, the positive predictive values of an estimate of low risk or low recurrence score ($\leq$ 21) were higher than 90%. On average, 23% of cases got different results of risk assessment after adjustment. Although cut-off values adjusted by multigene assay were not identical to those refined by survival, the adjusted values (17.1 and 23.8) and the refined values (17.5 and 24.5) of the best model (Magee Eq. 1) were close. Among all the evaluated models, Magee equation 2 was the only one without Ki67, and its prognostic values were the lowest. Using 20% as cut-off for Ki67 as suggested by St. Gallen consensus, we could confidently define luminal A cancer.

Conclusion: It is necessary to adjust the cut-off values of IHC-based prognostic models to fit the purpose. If the estimated risk is clearly high or low, it may be reasonable to omit multigene assays when cost is a consideration.

Keywords: breast neoplasms; gene expression profiling; immunohistochemistry; prognosis

1. Introduction

The histopathology of invasive breast cancer in women greatly impacts its management. In addition to traditional pathological parameters, such as histological type, grade, and stage, estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) status normally determined by immunohistochemistry (IHC) also play an important role. Current guidelines recommend that ER, PR and HER2 testing should be performed in all invasive

Conflicts of interest: The authors declare that they have no conflicts of interest related to the subject matter or materials discussed in this article.

* Corresponding author. Dr. Chih-Yi Hsu, Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, 201, Section 2, Shih-Pai Road, Taipei 112, Taiwan, ROC.
E-mail address: cyhsu@vghtpe.gov.tw (C.-Y. Hsu).

http://dx.doi.org/10.1016/j.jcma.2016.06.004

carcinomas of the breast to aid in treatment selection and to provide prognostic information.[1–4]

ER, PR, and HER2 testing defines the clinically useful subtypes of breast cancer, such as luminal, HER2, and triple-negative. There is still some uncertainty about the optimal treatment for patients with luminal-type tumors.[5] The St. Gallen International Expert Consensus suggests endocrine therapy for luminal A-like tumors defined by high receptor, low proliferation and low tumor burden ($\leq$ 3 positive nodes and tumor size $\leq$ 5 cm), and suggests adding cytotoxic chemotherapy for luminal B-like tumors with any of the markers indicative of lesser endocrine responsiveness. Multiparameter molecular (multigene) test if available is considered to have the highest efficacy. A low-risk result can support the omission of cytotoxic chemotherapy despite luminal B-like phenotype. However, multigene assays are highly expensive and not covered by the National Health Insurance of Taiwan.

For economic reasons, the use of prognostic models composed of four immunohistochemical markers (ER, PR, HER2, and Ki67) and pathological findings, such as IHC4 scores and Magee equations, work similarly to the multigene assay to provide information for prognostic and clinical judgments.[6–8] Although treatments guided by IHC4 scores are more likely to be cost effective,[9,10] IHC markers require standardization before widespread use. ER, PR, and HER2 are the leading breast cancer markers, and have readily available guideline recommendations for IHC testing.[3,4] Ki67 is not included in the American Society of Clinical Oncology and National Comprehensive Cancer Network guidelines because it shows greater variation in measurement and needs larger-scale analytical and clinical validation,[1,2,11] as was found between the study populations in the original IHC4 report.[6] Ki67 levels were on average about two and a half times higher due to manual readings and the use of the MIB1 antibody; therefore the multiplier was changed to four for Ki67 derived from MIB1 instead of 10 for that derived from the SP6 antibody and image analysis to make about 20 points of reduction in the IHC4 score.[6] Additionally, the cut-off point for a low Ki67 index changed from 15% (2009), 14% (2011), or 20% (2013) to 20–29% (2015) in the St. Gallen International Expert Consensus,[5,12–14] which makes it difficult to follow the cut-off point. Although there are some recommendations from the International Ki67 in Breast Cancer Working Group,[15] controversies continue to exist regarding counting only hot spots or all slide areas. Validation of local IHC results is needed before they can be applied to clinical decision making.

This study aimed to correlate the risk estimation derived from the IHC to those from multigene-expression assays for external references and correlate with the follow-up result for clinical validation. The cut-off values for IHC result to define luminal A tumors were tested.

2. Methods

The study protocol was approved by the Institutional Review Board of Taipei Veterans General Hospital, Taipei, Taiwan, R.O.C. Clinicopathological information of 642 consecutive patients with HER2− luminal-type (ER+ or PR+) early breast cancer who underwent surgery at Taipei Veterans General Hospital from 2010 to 2012 were retrieved from the medical records for survival analyses and clinical validation (Table S1). The median follow-up time was 52.7 months and distant recurrences were observed in 34 (5.3%) of cases. The second study cohort included 71 women with newly diagnosed HER2− luminal-type (ER+ or PR+) invasive carcinoma who had available multigene assay results (21-gene: 30 cases; 70-gene: 41 cases), collected from October 2009 to December 2015 (Table S2). The follow-up time of these 71 cases was relatively short (median, 31 months; range, 2–76 months). There was neither local nor distant recurrence. Among the cohort of 71 cases, 29 cases with results of 21-gene assay and longer follow-up time (median, 57 months) were included in the first dataset for clinical validation.

The original histopathological slides, including immunohistochemical stains for ER (clone 6F11; Leica Biosystems, Newcastle, UK, 1:100), PR (clone 16; Leica Biosystems, 1:150), HER2 (A0485; Dako, Glostrup, Denmark, 1:900), and Ki67 (clone MIB-1; Dako, 1:75), were evaluated by authors YYC and CYH without knowledge of the 21-gene or 70-gene assay results. The evaluations of ER, PR, and HER2 followed previously reported instructions.[3,4] One percent or more of tumor cells exhibiting nuclear staining was regarded as positive for ER and PR.[3] HER2 positivity was defined by complete intense membrane staining in $>$ 10% of tumor cells.[4] The percentages of Ki67 positive tumor cells derived from at least three high-power fields (400×) were averaged for the Ki67 labeling index using manual counting or image analysis (ImmunoRatio).[16,17]

Fisher's exact test was used to compare the distributions of categorical variables. Differences between continuous variables were compared using the Kruskal–Wallis test. Distant recurrence-free survival (DRFS) was measured from the date of surgery to the date of distant recurrence. Contralateral disease, other second primary cancers, and death before distant recurrence were considered censoring events. Locoregional recurrences were not considered events or censoring events. Survival curves were plotted using the Kaplan–Meier method, and their differences were calculated by log-rank test. Cox regression model was used to evaluate the hazard of recurrence. The prognostic values were compared using the Harrell C index, which is a rank parameter that measures the ordinal predictive power of a survival model by determining the probability of concordance between the predicted and the observed survival.[18] Harrell C index can range from 0.5 (no predictive discrimination) to 1.0 (perfect separation of patients with different outcomes).[18] The risk categories estimated by IHC4,[6] Magee equations,[8] or St. Gallen Consensus[5,14] were correlated to the multigene assay results. The details of IHC4 scores and Magee equations are listed in the footnotes of Table 1. The agreement of risk classifications was measured using kappa statistics, which were calculated as (observed agreement−agreement by chance) divided by (1−agreement by chance). The kappa statistics can range from −1 to +1, while the greater values reflect stronger agreement. The