



Original Article

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

Yen-Ying Chen<sup>a,b</sup>, Ling-Ming Tseng<sup>b,c</sup>, Ching-Fen Yang<sup>a,b</sup>, Pei-Ju Lien<sup>d</sup>, Chih-Yi Hsu<sup>a,b,\*</sup>

<sup>a</sup> Department of Pathology and Laboratory Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

<sup>b</sup> National Yang-Ming University School of Medicine, Taipei, Taiwan, ROC

<sup>c</sup> Division of General Surgery, Department of Surgery, and Comprehensive Breast Health Center, Taipei Veterans General Hospital, Taipei, Taiwan, ROC

<sup>d</sup> 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.

Copyright © 2016, the Chinese Medical Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**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](mailto:cyhsu@vghtpe.gov.tw) (C.-Y. Hsu).

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

1726-4901/Copyright © 2016, the Chinese Medical Association. Published by Elsevier Taiwan LLC. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

carcinomas of the breast to aid in treatment selection and to provide prognostic information.<sup>1–4</sup>

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.<sup>5</sup> 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. Multi-parameter 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.<sup>6–8</sup> Although treatments guided by IHC4 scores are more likely to be cost effective,<sup>9,10</sup> 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.<sup>3,4</sup> 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,<sup>1,2,11</sup> as was found between the study populations in the original IHC4 report.<sup>6</sup> 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.<sup>6</sup> 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,<sup>5,12–14</sup> which makes it difficult to follow the cut-off point. Although there are some recommendations from the International Ki67 in Breast Cancer Working Group,<sup>15</sup> 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.<sup>3,4</sup> One percent or more of tumor cells exhibiting nuclear staining was regarded as positive for ER and PR.<sup>3</sup> HER2 positivity was defined by complete intense membrane staining in  $> 10\%$  of tumor cells.<sup>4</sup> The percentages of Ki67 positive tumor cells derived from at least three high-power fields (400 $\times$ ) were averaged for the Ki67 labeling index using manual counting or image analysis (ImmunoRatio).<sup>16,17</sup>

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.<sup>18</sup> Harrell C index can range from 0.5 (no predictive discrimination) to 1.0 (perfect separation of patients with different outcomes).<sup>18</sup> The risk categories estimated by IHC4,<sup>6</sup> Magee equations,<sup>8</sup> or St. Gallen Consensus<sup>5,14</sup> 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

Table 1  
The risk groups classified by IHC4 score and Magee equation using original cut-off values.

|                                       | n (%)      | 5y-DRFS (%) | Univariate <sup>a</sup> |               | Multivariate <sup>b</sup> |               |
|---------------------------------------|------------|-------------|-------------------------|---------------|---------------------------|---------------|
|                                       |            |             | C                       | HR (p)        | C                         | HR (p)        |
| <b>IHC4 score</b>                     |            |             |                         |               |                           |               |
| <b>Ki67 multiplier 4<sup>c</sup></b>  |            |             |                         |               |                           |               |
| <−29.9                                | 307 (47.8) | 97.6        | 0.6982                  | 0.42 (0.064)  | 0.7238                    | 0.47 (0.114)  |
| −29.9–29.9                            | 245 (38.2) | 93.6        |                         | 1 (Reference) |                           | 1 (Reference) |
| > 29.9                                | 90 (14.0)  | 83.1        |                         | 3.22 (0.002)  |                           | 2.84 (0.009)  |
| <b>Ki67 multiplier 10<sup>d</sup></b> |            |             |                         |               |                           |               |
| <−29.9                                | 129 (20.1) | 99.2        | 0.7148                  | 0.21 (0.130)  | 0.7456                    | 0.23 (0.160)  |
| −29.9–29.9                            | 351 (54.7) | 95.5        |                         | 1 (Reference) |                           | 1 (Reference) |
| >29.9                                 | 162 (25.2) | 86.9        |                         | 3.60 (<0.001) |                           | 3.16 (0.002)  |
| <b>Magee equation</b>                 |            |             |                         |               |                           |               |
| <b>Magee 1<sup>e</sup></b>            |            |             |                         |               |                           |               |
| <18                                   | 286 (44.6) | 98.5        | 0.7166                  | 0.19 (0.002)  | 0.7464                    | 0.23 (0.008)  |
| 18–30                                 | 315 (49.1) | 91.9        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥31                                   | 41 (6.4)   | 78.9        |                         | 3.43 (0.003)  |                           | 3.47 (0.003)  |
| <b>Magee 2<sup>f</sup></b>            |            |             |                         |               |                           |               |
| <18                                   | 307 (47.8) | 97.2        | 0.6417                  | 0.35 (0.011)  | 0.6902                    | 0.43 (0.045)  |
| 18–30                                 | 301 (46.9) | 91.6        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥31                                   | 34 (5.3)   | 87.5        |                         | 1.90 (0.239)  |                           | 1.69 (0.351)  |
| <b>Magee 3<sup>g</sup></b>            |            |             |                         |               |                           |               |
| <18                                   | 335 (52.2) | 98.4        | 0.7347                  | 0.18 (0.001)  | 0.7530                    | 0.21 (0.002)  |
| 18–30                                 | 270 (42.1) | 90.9        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥31                                   | 37 (5.8)   | 76.4        |                         | 3.41 (0.003)  |                           | 3.29 (0.006)  |

C = Harrell C; 5y-DRFS = 5-year distant recurrence-free survival rate; ER = estrogen receptor; HER = human epidermal growth factor receptor; HR (p) = hazard ratio (significance); IHC = immunohistochemistry; PR = progesterone receptor.

<sup>a</sup> Univariate analyses.

<sup>b</sup> Multivariate analyses with adjustment of chemotherapy and hormonal therapy.

<sup>c</sup> IHC4 score =  $94.7 \times [-0.1 \times \text{ER H-score}/30 - 0.079 \times \text{PR \%}/10 + 0.586 \times \text{HER2} + 0.24 \times \ln(1 + 4 \times \text{Ki67})]$ .

<sup>d</sup> IHC4 score =  $94.7 \times [-0.1 \times \text{ER H-score}/30 - 0.079 \times \text{PR \%}/10 + 0.586 \times \text{HER2} + 0.24 \times \ln(1 + 10 \times \text{Ki67})]$ .

<sup>e</sup> Magee 1 =  $15.31385 + 1.4055 \times \text{Nottingham score} - 0.01924 \times \text{ER H-score} - 0.02925 \times \text{PR H-score} + \text{HER2}$  (0 for negative, 0.77681 for equivocal, 11.58134 for positive) +  $0.78677 \times \text{tumor size} + 0.13269 \times \text{Ki67}$ .

<sup>f</sup> Magee 2 =  $18.8042 + 2.34123 \times \text{Nottingham score} - 0.03749 \times \text{ER H-score} - 0.03065 \times \text{PR H-score} + \text{HER2}$  (0 for negative, 1.82921 for equivocal, 11.51378 for positive) +  $0.04267 \times \text{tumor size}$ .

<sup>g</sup> Magee 3 =  $24.30812 - 0.02177 \times \text{ER H-score} - 0.02884 \times \text{PR H-score} + \text{HER2}$  (0 for negative, 1.46495 for equivocal, 12.75525 for positive) +  $0.18649 \times \text{Ki67}$ .

positive likelihood ratio (LR+) was calculated as sensitivity divided by (1–specificity), while the greater LR+ value indicated an increased probability that the target was present. Cut-off values of Ki67, IHC4 scores, and Magee equations were adjusted based on survival outcome to reveal maximum Harrell C index or based on the maximum positive likelihood ratio correlated to multigene assay. As cytotoxics may be added in patients with 21-gene recurrence scores (RS) > 25,<sup>14</sup> one case of intermediate risk with RS = 26 was regarded as high risk. Eight cases of intermediate risk with RS ranging from 18 to 21 were regarded as low risk in the correlation analyses. The p-values were derived from two-tailed tests, and p < 0.05 was considered significant.

### 3. Results

#### 3.1. Correlation of IHC4 scores and Magee equations with DRFS

The distributions of risk categories of the 642 cases classified by IHC4 score and Magee equations using their original cut-off values are listed in Table 1. Although the DRFS of the

high-risk group either defined by IHC4 scores or Magee Eqs. 1 and 3 was shorter than those of intermediate- and low-risk groups, the proportion of high-risk groups revealed great differences which ranged from 5.8% to 25.2%. Also, the survival differences between intermediate and low-risk groups were not significant in IHC4 scores.

The values calculated by IHC4 scores and Magee equations all showed significant and continuous association with recurrence (Table S3). Their prognostic values represented by Harrell C were not significantly different, except that the prognostic value of Magee Eq. 2 was inferior to those of Magee Eqs. 1 and 3 (p = 0.001 and p = 0.015, respectively). The prognostic value with adjustment of chemotherapy and hormonal therapy of Magee Eq. 2 was also inferior to that of Magee Eq. 1 (p = 0.013).

#### 3.2. The cut-off values of IHC4 scores and Magee equations refined by DRFS

The cut-off values of IHC4 scores and Magee equations could be optimized by testing different cut-off values to give the maximum Harrell C value. Using refined cut-off values,

risk category was changed in an average of 21% of the cases. All the Harrell C values were increased and the hazard ratios between intermediate- and low-risk groups classified by IHC4 scores became significant (Table 2). The survival curves stratified by risk groups showed a similar trend (Figs. S1–S2). The survival difference between intermediate- and low-risk groups was not significant in Magee Eq. 2. Its prognostic value was the lowest and was significantly inferior to those by Magee Eq. 1 and 3. The proportions of risk categories classified by different IHC4 scores and Magee equations were closer. The proportions of high risk ranged from 22.4% to 27.4%, while those of low risk ranged from 35.4% to 43.2%.

### 3.3. Correlation of IHC4 scores and Magee equations with multigene assays

PR, Ki67, IHC4 scores, and Magee equations were significantly related to the risk categories derived from multigene assays, while ER and the other clinicopathological features did not show a significance. The cut-off values of IHC4 scores and Magee equations could be adjusted by the maximum positive likelihood ratio in predicting the low- and high-risk categories derived from multigene assays (Table 3). Although the 21- and 70-gene assay were two different assays, the adjusted cut-off values did not change significantly when the 30 cases with results of the 21-gene assay were excluded. The cut-off values in predicting the high-risk category became slightly lower, while those in predicting the low-risk category were the same (Table S4). The positive predictive values of an estimate of low risk or low recurrent score ( $\leq 21$ ) were higher than 90%.

The multigene adjusted cut-off values were lower than the original ones in most of the models, except for IHC4 score using a Ki67 multiplier of 10. Applying the adjusted cut-off values to the cohort of 642 cases (Table 4), on average, 24% were reclassified into a different category. Magee Eq. 1 showed the maximum prognostic value (Harrell C) and classified the fewest individuals (31.3%) into the intermediate-risk group. Among models, the proportions of high-risk groups were relatively close (range, 20.9–29.0%) and the survival curves revealed a similar trend (Figs. S1–S2). Of note, the survival differences between low- and intermediate-risk groups were mostly insignificant in multivariate analyses with chemotherapy and hormonal therapy adjustment. Although the cut-off values adjusted by multigene assay (Table 4) were not identical to those refined by survival (Table 2), cut-off values of Magee Eq. 1 optimized by different methods were close (17.1 and 23.8 vs. 17.5 and 24.5, respectively). These optimized cut-off values were lower than the original values (18 and 31). Additionally, 27.4% and 17.5% of cases were upgraded to a higher risk category than the original one.

### 3.4. Cut-off value for Ki67 index to define luminal A tumors

We tested the criteria of St. Gallen consensus using different Ki67 values to define luminal A tumors, and found that using 20% as cut-off got the maximum positive likelihood ratio (Table S5). However, using median (25%) as a cut-off point showed the highest concordance (67.6%) with the

Table 2  
The risk groups classified by IHC4 score and Magee equation using optimized cut-off values for survival.

|                    | n (%)      | 5y-DRFS (%) | Univariate <sup>a</sup> |               | Multivariate <sup>b</sup> |               |
|--------------------|------------|-------------|-------------------------|---------------|---------------------------|---------------|
|                    |            |             | C                       | HR (p)        | C                         | HR (p)        |
| IHC4 score         |            |             |                         |               |                           |               |
| Ki67 multiplier 4  |            |             |                         |               |                           |               |
| ≤−40.1             | 227 (35.4) | 99.0        | 0.7481                  | 0.20 (0.032)  | 0.7714                    | 0.22 (0.047)  |
| −40.0–13.8         | 270 (42.1) | 94.5        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥13.9              | 145 (22.6) | 85.3        |                         | 3.38 (0.001)  |                           | 3.10 (0.003)  |
| Ki67 multiplier 10 |            |             |                         |               |                           |               |
| ≤−19.4             | 227 (35.4) | 99.0        | 0.7481                  | 0.20 (0.032)  | 0.7714                    | 0.22 (0.047)  |
| −19.3–34.5         | 270 (42.1) | 94.5        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥34.6              | 145 (22.6) | 85.3        |                         | 3.38 (0.001)  |                           | 3.10 (0.003)  |
| Magee equation     |            |             |                         |               |                           |               |
| Magee 1            |            |             |                         |               |                           |               |
| ≤17.5              | 277 (43.2) | 99.2        | 0.7812                  | 0.16 (0.016)  | 0.7841                    | 0.17 (0.024)  |
| 17.6–24.4          | 221 (34.4) | 94.5        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥24.5              | 144 (22.4) | 83.5        |                         | 3.71 (0.001)  |                           | 3.58 (0.001)  |
| Magee 2            |            |             |                         |               |                           |               |
| ≤13.9              | 229 (35.7) | 98.1        | 0.6916                  | 0.37 (0.091)  | 0.7250                    | 0.44 (0.157)  |
| 14.0–22.3          | 237 (36.9) | 94.3        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥22.4              | 176 (27.4) | 88.5        |                         | 2.46 (0.018)  |                           | 2.25 (0.034)  |
| Magee 3            |            |             |                         |               |                           |               |
| ≤17.0              | 277 (43.2) | 98.8        | 0.7481                  | 0.23 (0.026)  | 0.7733                    | 0.25 (0.036)  |
| 17.1–21.8          | 192 (29.9) | 94.3        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥21.9              | 173 (27.0) | 86.3        |                         | 2.93 (0.007)  |                           | 2.67 (0.016)  |

C = Harrell C; 5y-DRFS = 5-year distant recurrence-free survival rate; HR (p) = hazard ratio (significance); IHC = immunohistochemistry.

<sup>a</sup> Univariate analyses.

<sup>b</sup> Multivariate analyses with adjustment of chemotherapy and hormonal therapy.

Table 3  
Cut-off values adjusted by the maximum positive likelihood ratio in predicting the risk categories derived from multigene assays.

|                    | <i>n</i> | Low     | Intermediate <sup>a</sup> | High    | LR+ <sup>b</sup> | Kappa <sup>c</sup> |
|--------------------|----------|---------|---------------------------|---------|------------------|--------------------|
| Total              | 71       | 42 (59) | 9 (13) <sup>d</sup>       | 20 (28) |                  |                    |
| IHC4 score         |          |         |                           |         |                  |                    |
| Ki67 multiplier 4  |          |         |                           |         |                  |                    |
| ≤−43.6             | 24       | 21 (88) | 2 (8)                     | 1 (4)   | 9.7              |                    |
| −43.5–8.5          | 37       | 21 (57) | 6 (16) <sup>d</sup>       | 10 (27) |                  |                    |
| ≥8.6               | 10       | 0 (0)   | 1 (10)                    | 9 (90)  | 21.4             | 0.482              |
| Ki67 multiplier 10 |          |         |                           |         |                  |                    |
| ≤−22.9             | 24       | 21 (88) | 2 (8)                     | 1 (4)   | 9.7              |                    |
| −22.8–29.1         | 37       | 21 (57) | 6 (16) <sup>d</sup>       | 10 (27) |                  |                    |
| ≥29.2              | 10       | 0 (0)   | 1 (10)                    | 9 (90)  | 21.4             | 0.482              |
| Magee equation     |          |         |                           |         |                  |                    |
| Magee 1            |          |         |                           |         |                  |                    |
| ≤17.1              | 30       | 25 (83) | 4 (13)                    | 1 (3)   | 12.2             | 0.417              |
| 17.2–23.7          | 33       | 17 (52) | 4 (12) <sup>d</sup>       | 12 (36) |                  |                    |
| ≥23.8              | 8        | 0 (0)   | 1 (13)                    | 7 (88)  | 16.7             |                    |
| Magee 2            |          |         |                           |         |                  |                    |
| ≤15.5              | 25       | 21 (84) | 3 (12)                    | 1 (4)   | 10.1             | 0.321              |
| 15.6–24.5          | 42       | 21 (50) | 5 (12) <sup>d</sup>       | 16 (38) |                  |                    |
| ≥24.6              | 4        | 0 (0)   | 1 (25)                    | 3 (75)  | 7.1              |                    |
| Magee 3            |          |         |                           |         |                  |                    |
| ≤16.2              | 25       | 22 (88) | 2 (8)                     | 1 (4)   | 10.1             |                    |
| 16.3–23.2          | 37       | 20 (54) | 6 (16) <sup>d</sup>       | 11 (30) |                  |                    |
| ≥23.3              | 9        | 0 (0)   | 1 (11)                    | 8 (89)  | 19.0             | 0.433              |

Data are presented as *n* (%).

IHC = immunohistochemistry; LR+ = positive likelihood ratio; RS = recurrence scores.

<sup>a</sup> Eight cases with 21-gene RS < 25 (range, 18–21) were regarded as low risk, and one case with RS ≥ 25 was regarded as high risk in the analyses.

<sup>b</sup> Positive likelihood ratio in predicting the low- or high-risk group.

<sup>c</sup> Kappa statistics of two-tier risk estimation using the cut-off value.

<sup>d</sup> Including one case with RS = 26.

results of multigene assay, and significantly higher prognostic value than by using 14% or 20% (both  $p < 0.001$ ) in the cohort of 642 cases (Table S6).

#### 4. Discussion

In this study, we confirmed that IHC-based prognostic models provided inexpensive risk assessments, but their cut-off values required adjustment. On average, 23% of cases got different results of risk assessment after adjustment. The cut-off values refined by survival outcomes could get better prognostic values and predict more differences in survival among the risk groups. However, the cut-off values refined by survival did not match with those correlated to multigene assays. Magee Eq. 1 was the best of the prognostic models evaluated. It had the highest prognostic values with regard to the value calculated by the equation (Table S3) and the risk categories classified by the adjusted cut-off values (Table 2 and Table 4, respectively). Also, its cut-off values refined by survival (17.5 and 24.5, respectively) were very close to those adjusted by multigene assays (17.1 and 23.8, respectively). Replacing the cut-off of Ki67 (20%) by the median (25%) of Ki67 for our cases got higher prognostic values and better

concordance with multigene assay in distinguishing the low-risk from the high-risk luminal-type cancers, but the positive likelihood ratio of predicting the low-risk group decreased.

It is debatable, however, to include Ki67 to distinguish the low-risk from the high-risk luminal-type cancers.<sup>5</sup> In our study, Ki67 was significantly related to the DRFS and the risk categories derived from multigene assays. The studied IHC-based prognostic models all showed prognostic significance. Magee Eq. 2 was the only one not including Ki67 in the equation, and its prognostic values were the lowest and significantly inferior to those of the other Magee equations. These findings support the theory that Ki67 scores carry important prognostic information. Although defining a single useful cut-off point may not be fully applicable to all conditions and laboratories, using a higher Ki67 index (Ki67 ≥ 35%) to indicate a poor prognosis (Table S1) and a lower Ki67 index (Ki67 < 20%) to confidently define luminal A cancer is feasible in our institute.

When the IHC-based prognostic models are used as prognostic markers, the DRFS corrected cut-off values should be the most appropriate. If the aim is to predict benefit from chemotherapy, using the result of multigene assays for external references should be a successful method. Although a threshold value has not been established, multigene assays are frequently used to assist in decisions about the inclusion of cytotoxic chemotherapy. The optimal threshold of multigene assays to define the clinical benefit should be based on the thresholds that are clinically validated against the outcomes compared between treated and untreated patients. The 21-gene assay has been shown to predict chemotherapy benefit in two analyses in Phase III clinical trial settings.<sup>19,20</sup> The low-risk group (RS < 18) did not appear to benefit from chemotherapy, and the high-risk group (RS > 30) derived major benefit from chemotherapy. The benefit in the intermediate group was unclear. Another two randomized clinical trials (TAILORx and RxPONDER trials) are currently being conducted to evaluate the benefit of chemotherapy in patients with low to intermediate risk (RS < 25). The 70-gene assay has also been reported as being predictive of chemotherapy benefit based on the results of pooled study series, and its prospective validation in a randomized clinical trial (the MINDACT trial) is ongoing.<sup>21</sup>

Whether the multigene assay is more accurate or offers more information than basic IHC is controversial.<sup>22,23</sup> The expensive multigene assays push us to refine the risk stratification for adjuvant chemotherapy for patients with hormonal receptor-positive tumors, but there is insufficient evidence to support that these assays play a role in determining ER, PR, or HER2 status in unselected patients.<sup>2–4,24</sup> ER, PR, and HER2 status determined by IHC are necessary for breast cancer patients. Using basic IHC for risk stratification has advantages in its low cost and ready availability. However, the potential for interlaboratory variation in the values of IHC remains a justifiable concern. Efforts to improve standardization and reproducibility of IHC are needed. In fact, the results of multigene assays for the same cohort of breast cancer patients can be discordant,<sup>25</sup> and agreement between the assays in one

Table 4  
The risk groups classified by IHC4 score and Magee equation using cut-off values validated by multigene assays.

|                       | n (%)      | 5y-DRFS (%) | Univariate <sup>a</sup> |               | Multivariate <sup>b</sup> |               |
|-----------------------|------------|-------------|-------------------------|---------------|---------------------------|---------------|
|                       |            |             | C                       | HR (p)        | C                         | HR (p)        |
| <b>IHC4 score</b>     |            |             |                         |               |                           |               |
| Ki67 multiplier 4     |            |             | 0.7296                  |               | 0.7579                    |               |
| ≤−43.6                | 180 (28.0) | 99.4        |                         | 0.13 (0.046)  |                           | 0.14 (0.062)  |
| −43.5–8.5             | 298 (46.4) | 94.8        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥8.6                  | 164 (25.6) | 87.1        |                         | 3.02 (0.002)  |                           | 2.74 (0.006)  |
| Ki67 multiplier 10    |            |             | 0.7303                  |               | 0.7584                    |               |
| ≤−22.9                | 180 (28.0) | 99.4        |                         | 0.13 (0.046)  |                           | 0.14 (0.062)  |
| −22.8–29.1            | 299 (46.6) | 94.8        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥29.2                 | 163 (25.4) | 87.0        |                         | 3.05 (0.002)  |                           | 2.76 (0.006)  |
| <b>Magee equation</b> |            |             |                         |               |                           |               |
| Magee 1               |            |             | 0.7537                  |               | 0.7676                    |               |
| ≤17.1                 | 255 (39.7) | 99.1        |                         | 0.17 (0.025)  |                           | 0.20 (0.039)  |
| 17.2–23.7             | 201 (31.3) | 94.6        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥23.8                 | 186 (29.0) | 86.6        |                         | 3.03 (0.005)  |                           | 2.84 (0.010)  |
| Magee 2               |            |             | 0.6793                  |               | 0.7080                    |               |
| ≤15.5                 | 197 (30.7) | 97.8        |                         | 0.45 (0.154)  |                           | 0.53 (0.265)  |
| 15.6–24.5             | 331 (51.6) | 94.7        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥24.6                 | 114 (17.8) | 85.7        |                         | 3.14 (0.002)  |                           | 2.78 (0.006)  |
| Magee 3               |            |             | 0.7324                  |               | 0.7523                    |               |
| ≤16.2                 | 223 (34.7) | 99.0        |                         | 0.18 (0.023)  |                           | 0.20 (0.034)  |
| 16.3–23.2             | 285 (44.4) | 94.3        |                         | 1 (Reference) |                           | 1 (Reference) |
| ≥23.3                 | 134 (20.9) | 85.5        |                         | 2.95 (0.002)  |                           | 2.67 (0.008)  |

C = Harrell C; 5y-DRFS = 5-year distant recurrence-free survival rate; HR (p) = hazard ratio (significance); IHC = immunohistochemistry.

<sup>a</sup> Univariate analyses.

<sup>b</sup> Multivariate analyses with adjustment of chemotherapy and hormonal therapy.

study was only moderate ( $Kappa = 0.527$ ).<sup>23</sup> In the present study, IHC4 scores and Magee equations using the cut-off values with maximum positive likelihood ratio reached fair to moderate agreement with those using multigene assays ( $Kappa = 0.390–0.470$ ; Table 3). As a matter of fact, we could not expect totally matched results since the principles and the targets of detection by the IHC and multigene assays were different. At least the use of IHC can reduce the number of cases requiring expensive multigene assays. If the risk estimated by Magee Eq. 1 falls clearly in the high- or low-risk category, a dramatically different result from multigene assays should not be expected.

The current study was limited by data collected in a single institute with restricted sample size and follow-up time. Despite this, the 5-year DRFS rate of our cases (5.3%) is consistent with those in the literature (4–5%).<sup>6,7</sup> Our survival refined cut-off values were close to those adjusted by multigene assays (external reference). Further validations in larger cohorts with a longer follow-up time and in different laboratories are needed for IHC-based prognostic models to be widely implemented.

In conclusion, it is necessary to adjust the cut-off values of IHC-based prognostic models to fit the purpose. The risk group was reclassified in about one fifth of our cases after adjustment. If the estimated risk from the IHC-based models is clearly high or low, the result from the multigene assays is less likely to be significantly different, and it may be reasonable to omit multigene assays in this setting when cost is a consideration.

## Acknowledgments

This study was supported by grants from Taipei Veterans General Hospital (V99C1-183 and V104C-187).

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jcma.2016.06.004>.

## References

- Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 2007;**25**: 5287–312.
- National Comprehensive Cancer Network. NCCN clinical practice guidelines in oncology: breast cancer, 2015, version 3. [http://www.nccn.org/professionals/physician\\_gls/pdf/breast.pdf](http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf). [Accessed 15 July 2015].
- Hammond MEH, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Arch Pathol Lab Med* 2010;**134**:907–22.
- Wolff AC, Hammond MEH, Hicks DG, Dowsett M, McShane LM, Allison KH, et al. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Arch Pathol Lab Med* 2014;**138**:241–56.
- Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnani M, Piccart-Gebhart M, et al. Tailoring therapies—improving the management of early breast cancer: St Gallen International Expert Consensus on the

- Primary Therapy of Early Breast Cancer 2015. *Ann Oncol* 2015;**26**:1533–46.
6. Cuzick J, Dowsett M, Pineda S, Wale C, Salter J, Quinn E, et al. Prognostic value of a combined estrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 immunohistochemical score and comparison with the Genomic Health recurrence score in early breast cancer. *J Clin Oncol* 2011;**29**:4273–8.
  7. Sgroi DC, Sestak I, Cuzick J, Zhang Y, Schnabel CA, Schroeder B, et al. Prediction of late distant recurrence in patients with oestrogen-receptor-positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the Trans-ATAC study population. *Lancet Oncol* 2013;**14**:1067–76.
  8. Klein ME, Dabbs DJ, Shuai Y, Brufsky AM, Jankowitz R, Puhalla SL, et al. Prediction of the Oncotype DX recurrence score: use of pathology-generated equations derived by linear regression analysis. *Mod Pathol* 2013;**26**:658–64.
  9. Barton S, Zabaglo L, A'Hern R, Turner N, Ferguson T, O'Neill S, et al. Assessment of the contribution of the IHC4+C score to decision making in clinical practice in early breast cancer. *Br J Cancer* 2012;**106**:1760–5.
  10. Ward S, Scope A, Rafia R, Pandor A, Harnan S, Evans P, et al. Gene expression profiling and expanded immunohistochemistry tests to guide the use of adjuvant chemotherapy in breast cancer management: a systematic review and cost-effectiveness analysis. *Health Technol Assess* 2013;**17**:1–302.
  11. Gudlaugsson E, Skaland I, Janssen EA, Smaaland R, Shao Z, Malpica A, et al. Comparison of the effect of different techniques for measurement of Ki67 proliferation on reproducibility and prognosis prediction accuracy in breast cancer. *Histopathology* 2012;**61**:1134–44.
  12. Goldhirsch A, Ingle JN, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Thresholds for therapies: highlights of the St Gallen International Expert Consensus on the primary therapy of early breast cancer 2009. *Ann Oncol* 2009;**20**:1319–29.
  13. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2011;**22**:1736–47.
  14. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol* 2013;**24**:2206–23.
  15. Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes RC, Cuzick J, et al. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer Working Group. *J Natl Cancer Inst* 2011;**103**:1656–64.
  16. Hsu CY, Ho DM, Yang CF, Chiang H. Interobserver reproducibility of MIB-1 labeling index in astrocytic tumors using different counting methods. *Mod Pathol* 2003;**16**:951–7.
  17. Tuominen VJ, Ruotoistenmaki S, Viitanen A, Jumppanen M, Isola J. ImmunoRatio: a publicly available web application for quantitative image analysis of estrogen receptor (ER), progesterone receptor (PR), and Ki-67. *Breast Cancer Res* 2010;**12**:R56.
  18. Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–87.
  19. Albain KS, Barlow WE, Shak S, Hortobagyi GN, Livingston RB, Yeh IT, et al. Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 2010;**11**:55–65.
  20. Paik S, Tang G, Shak S, Kim C, Baker J, Kim W, et al. Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 2006;**24**:3726–34.
  21. Albain KS, Paik S, van't Veer L. Prediction of adjuvant chemotherapy benefit in endocrine responsive, early breast cancer using multigene assays. *Breast* 2009;**18**(Suppl 3):S141–5.
  22. Shiang C, Pusztai L. Molecular profiling contributes more than routine histology and immunohistochemistry to breast cancer diagnostics. *Breast Cancer Res* 2010;**12**(Suppl 4):S6.
  23. Weigelt B, Reis-Filho JS. Molecular profiling currently offers no more than tumour morphology and basic immunohistochemistry. *Breast Cancer Res* 2010;**12**(Suppl 4):S5.
  24. O'Connor SM, Beriwal S, Dabbs DJ, Bhargava R. Concordance between semiquantitative immunohistochemical assay and oncotype DX RT-PCR assay for estrogen and progesterone receptors. *Appl Immunohistochem Mol Morphol* 2010;**18**:268–72.
  25. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, et al. Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006;**355**:560–9.