Reviewer's report

Title: Predict model: KI67 validation

Version: 3  Date: 11 November 2014

Reviewer: Chi-Cheng Huang

Reviewer's report:

Despite some disagreements with my comments, the authors have revised the manuscript with essential and critical corrections. Here is my opinions to the revised manuscript.

1. The distribution of ER status in the discovery cohort for KI67 is added. As the authors stated, most regression weights remained the same as those in V1/V2 so they did not repeat these descriptions in this manuscript.

2. The proportion of missing values of each clinical features in added in Table 1.

3. It's interesting that both HER2 (v2) and KI67 (v3) are "external estimates" in additional the the original PREDICT model so the regression weights of other prognostic factors could be fixed with the values derived from the ECRIC study. I still suggest that more concise descriptions of the PREDICT model could enhance the readability and completeness of this manuscript for readers who are not familiar with PREDICT and save a lot of time to find answers from the three antecedent published papers.

4. No comment.

5. In the first BCR paper, the effect of adjuvant therapy (hormone and chemotherapy coefficient) could be evaluated in two ways, either model-derived or overview estimates. The authors should state clearly whether the effect of adjuvant therapy is derived from the ECRIC study or from external estimates such as literature review. Enhanced model performance with better predictive power of breast cancer mortality did not necessarily indicate a .predictive power of chemo/hormone therapy response of individual patient, while prognostic (breast cancer mortality) and predictive (adjuvant therapy response) values are claimed for PREDICT; if I am not misunderstood.

6. No more comment.

7. In terms of machine learning algorithm, calibration and discrimination evaluate the classification performance for binary responses. The clinical validity of Oncotype comes from the survival discrepancy between defined risk groups. And they present the results with distinct survival curves, but not with the two-by-two table as PREDICT did. I agree with the authors that expensive multi-gene assays such as Oncotype may not provide additional prognostic value than the readily available PREDICT but the best way to estimate the clinical validity is to initial a prospective clinical trial (not retrospective historical cohort), such as TAILORx, and maybe in the near future we could see the results of clinical decision-making
based on PREDICT.
8. The issue of KI67 threshold is partially resolved and supported by the newly added Ref. 24.

Minor:
1,2,3,5. No more comment.
4. The word of "quintile" is so critical that we could understand the determination of d.f.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
I declared no competing interests during reviewing this manuscript.