Author's response to reviews

Title: Predict model: Ki67 validation

Authors:

  Gordon C Wishart Professor (gcwishart@gmail.com)
  Emad Rakha Dr (emadrakha@yahoo.com)
  Andrew Green Dr (Andrew.Green@nottingham.ac.uk)
  Ian Ellis Professor (Ian.Ellis@nottingham.ac.uk)
  H. Raza Ali Dr (Raza.Ali@cruk.cam.ac.uk)
  Elena Provenzano Dr (Elena.provenzano@addenbrookes.nhs.uk)
  Fiona M Blows (fmb28@medschl.cam.ac.uk)
  Carlos Caldas Professor (Carlos.Caldas@cruk.cam.ac.uk)
  Paul D P Pharoah Prof (paul.pharoah@srl.cam.ac.uk)

Version: 4 Date: 14 November 2014

Author's response to reviews: see over
Referee 4

It's interesting that both HER2 (v2) and Ki67 (v3) are "external estimates" in addition to 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 antecedent published papers.

We have added a paragraph to the text in the introduction briefly describing the derivation of PREDICT. We have added a table (Table 1) that provides the hazard ratios used by PREDICT (these are as previously published and have not been derived in this study). The newly derived HR.

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.

We have clarified that PREDICT (as implemented online) used the benefits of adjuvant chemotherapy and hormone therapy from randomised trials.

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.

This is all a matter of terminology. We are very clear that we are using prediction in the sense that it is used for a predictive model – i.e. it is about prediction of a future event. Oncologists often make a contrast between prognostic and predictive markers. When used like this, the term predictive has no well-established, standard definition, and we think it best avoided. Given that predictive has no clear definition we do not make any claims for PREDICT in this sense.

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.

This is not clinical validity by any standard definition!
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.

Such a trial would be to establish clinical utility, not just clinical validity. This referee seems not to understand that the terms clinical validity and utility have well-established definitions.