Development and External Validation of Prediction

Models for 10-Year Survival of Invasive Breast Cancer.

Comparison with PREDICT and CancerMath.

Solon Karapanagiotis[1][**], Paul D. P. Pharoah[2], Christopher H. Jackson[1][*], Paul J. Newcombe[1][*]

[1] MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[2] Department of Oncology, University of Cambridge, Cambridge, UK

[*] Joint senior authors

[**] Corresponding author: Solon Karapanagiotis, MRC Biostatistics Unit, University of Cambridge School of Clinical Medicine, Cambridge Institute of Public Health Forgive Site, Robinson Way Cambridge Biomedical Campus,

Cambridge CB2 0SR; solon.karapanagiotis@mrc-bsu.cam.ac.uk

Abstract

Background: To compare PREDICT and CancerMath, two widely used prognostic models for invasive breast cancer, taking into account their clinical utility. Furthermore, it is unclear whether these models could be improved.

Methods: A dataset of 5729 women was used for model development. A Bayesian variable selection algorithm was implemented to stochastically search for important interaction terms among the predictors. The derived models were then compared in three independent datasets (n = 5534). We examined calibration, discrimination and performed decision curve analysis.

Results: CancerMath demonstrated worse calibration performance compared to PREDICT in oestrogen receptor (ER)-positive and ER-negative tumours. The decline in discrimination performance was -4.27% (-6.39 – -2.03) and -3.21% (-5.9 – -0.48) for ER-positive and ER-negative tumours, respectively. Our new models matched the performance of PREDICT in terms of calibration and discrimination, but offered no improvement. Decision curve analysis showed predictions for all models were clinically useful for treatment decisions made at risk thresholds between 5% and 55% for ER-positive tumours and at thresholds of 15% to 60% for ER-negative tumours. Within these threshold ranges, CancerMath provided the lowest clinical utility amongst all the models.

Conclusions: Survival probabilities from PREDICT offer both improved accuracy and discrimination over CancerMath. Using PREDICT to make treatment decisions offers greater clinical utility than CancerMath over a range of risk thresholds. Our new models performed as well as PREDICT, but no better, suggesting that, in this setting, including further interaction terms offers no predictive benefit.

Keywords: Bayesian variable selection; breast cancer; CancerMath; decision curve analysis; external validation; prediction model; PREDICT.

Introduction

Accurate prognostic predictions are crucial in clinical decision making around adjuvant therapy following surgery for early breast cancer. Adjuvant therapies, such as chemotherapy, are associated with serious side-effects and so are typically only in the patient's interest when there is non-negligible risk of mortality. Inaccurate risk predictions can lead to over-prescription of chemotherapy to patients who do not require it, and unnecessarily increased mortality among patients whose risk is underestimated. Over the years, several predictive models have been proposed, that can aid in the decision to treat patients with systemic therapy, based on clinical and pathological factors (1).

One such tool, PREDICT, was proposed in 2010 (www.predict.nhs.uk). It provides predictions of 5- and 10-year breast cancer survival and the benefits of hormonal therapy, chemotherapy and trastuzumab (2). The model utilizes data on factors such as patient age, tumour size, tumour grade, number of positive nodes, oestrogen receptor (ER) status, Ki67 status, mode of detection and human epidermal growth factor receptor 2 (HER2) status (3). PREDICT has been validated in many cohorts (4–6). An updated version improving performance has been published recently (7). Another prognostic tool, CancerMath, was proposed in 2009 (http://cancer.lifemath.net/). CancerMath models risk predictions as a function of tumour size, number of nodes and various other prognostic factors (8), and is also able to predict the impact of various adjuvant therapies. CancerMath has been validated using data from patients in the United States (9), as well as a South East population (10).

Although both PREDICT and CancerMath are widely used for cancer prognosis, to date, there is limited data on the concordance of these two prediction models. Only, Laas et al., (11) compared the overall performance of CancerMath and PREDICT using information from 965 women with ER-positive and HER2-negative early breast cancer from the United States and

Canada. Further comparison in larger populations may be necessary. Furthermore, risk models are routinely evaluated using only standard performance metrics of discrimination and calibration. Such metrics give limited information as to the clinical value of a prediction model (12). This is because they focus purely on the accuracy of the predictions from the model and do not consider the clinical consequences of decisions made using the model (13) such as who should receive adjuvant therapy. For example, a false positive prediction from the model would result in a patient being treated unnecessarily, while a false negative prediction would result in a patient not getting a treatment that she would benefit from. The relative harms or benefits of these alternative clinical outcomes should be weighted appropriately when evaluating the models. Decision curve analysis (14) is a decision-theoretic method to compare prediction models with respect to their clinical utility. It provides a calculation of net benefit: the expected utility of a decision to treat patients at some threshold, compared to an alternative policy such as treating nobody. "Utility" is defined precisely below, however, conceptually, it balances the number of patients whose lives are saved against the cost of unnecessarily treating patients who would have survived anyway. In practice, adjuvant chemotherapy is recommended as a treatment option from the Cambridge Breast Unit (Addenbrooke's Hospital, Cambridge, UK) if a patient's 10-year predicted survival probability using chemotherapy, minus their predicted probability without, is greater than 5% (15). Hence, it would be of interest to evaluate whether or not using the existing thresholds alongside the current prediction models lead to better clinical utility.

The aim of this study was three-fold. First, to perform the first direct comparison of CancerMath and PREDICT in a large European validation cohort, second, to explore whether modelling interactions among the risk factors used in PREDICT and CancerMath can lead to more accurate prognostic models, and finally, to exemplify the use of decision curve analysis as a means of comparing prognostic models.

Methods

Study population

Training data

The training data used for model development was based on data from 5729 patients with invasive breast cancer diagnosed in East Anglia, UK between 1999 and 2003 identified by the Eastern Cancer Registration and Information Centre (ECRIC). All analyses used data censored on 31 December 2012. For more details on the source of data and eligibility criteria see (2).

Information obtained included age at diagnosis (years), ER status (positive or negative), tumour grade (I, II, III), tumour size (mm), number of positive lymph nodes, presentation (screening vs. clinical), and type of adjuvant therapy (chemotherapy, endocrine therapy, or both). Separate models were derived for ER-positive and ER-negative breast cancer. The outcome modelled was 10-year breast cancer specific mortality (BCSM). BCSM is defined as time to death from breast cancer.

Validation data

The validation data consisted of three cohorts, the Nottingham/Tenovus Breast Cancer Study (1944 subjects) (16), the Breast Cancer Outcome Study of Mutation Carriers (981 subjects) (17), and the Prospective Study of Outcomes in Sporadic and Hereditary Breast Cancer (2609 subjects) (18). The same predictors as in the training set were available. The outcome measure was again 10-year BCSM status. If BCSM status was missing, the patient's data were excluded from the analysis. CancerMath was validated on a subset of this validation dataset (3431 ER-positive and 1418 ER-negative subjects) whose tumour size and/or number of positive nodes did not exceed the values against which it was originally validated, 50mm and 10 nodes, respectively.

Development of novel prognostic models

We developed novel prognostic models by building upon the existing PREDICT and CancerMath frameworks to incorporate interaction terms between the covariates. To this end, we used a Bayesian variable selection algorithm, Reversible Jump Markov Chain Monte Carlo (RJMCMC) to stochastically search for important interaction terms (19). We present two implementations of the algorithm. The first using a Weibull model for the baseline survival time (denoted Weibull RJMCMC hereafter), and the second with a logistic regression model for the binary outcome of 10-year survival (denoted Logistic RJMCMC hereafter). For both models, all predictors were transformed as in the latest version of PREDICT (7). The RJMCMC method searched over models with all combinations of pairwise interactions between the predictors listed, while including the main effects of all predictors in both models. For further details on the RJMCMC set-up see Supplementary Materials.

Both PREDICT and CancerMath are web-based interfaces, implemented in Javascript. The Javascript code, was downloaded in April 2017, and translated into R (20) and subsequently cross-checked with a random subset of 20 patients, to verify the accuracy of the R translation. Some of the predictors used in PREDICT and CancerMath were not available in our data sets, such as the chemotherapy generation. We considered that all the patients received second generation chemotherapy, as the data were older than 10 years. All other missing predictors were treated as unknown.

Comparison of models in terms of predictive performance and decision curve analysis

We assessed the predictive performance of the models on the validation data by examining measures of calibration, discrimination and clinical net benefit. 95% confidence intervals (CI) were obtained by bootstrapping using 1000 re-samples of the validation set.

Calibration

Calibration was assessed graphically, by plotting the predicted 10-year mortality risks (x axis) against the observed probabilities (y axis), calculated using smoothing techniques (21). Ideally, if predicted and observed 10-year survival probabilities agree over the whole range of probabilities, the plots show a 45-degree line.

Discrimination

Discrimination is the ability of the risk score to differentiate between patients who do and do not experience the event during the study period. Model discrimination was assessed by calculating the area under the receiver-operator characteristic curve (AUC). The AUC is the probability that the predicted mortality of a randomly selected patient who died is higher than that of a randomly selected patient who survived; the higher the AUC, the better the model is at identifying patients with a worse survival.

Decision curve analysis

We used decision curve analysis to describe and compare the clinical consequences of the models (14). The net benefit of a model is the difference between the proportion of true positives (TP) and the proportion of false positives (FP) weighted by the threshold R (expressed on the odds scale) at which an individual is classified as "high risk":

$$Net\ benefit = \frac{TP}{n} - \frac{FP}{n}\left[\frac{R}{1-R}\right].$$

Here n is the total sample size and R is the "high risk designation" threshold at which a treatment is prescribed. The risk threshold (R) used by clinicians to prescribe treatment implies a relative value for either receiving treatment if death was likely or avoiding treatment if death was not

likely. For example, if the clinician views unnecessary chemotherapy in nine women as an acceptable cost for correctly treating one high risk woman, this is equivalent to treating all women with ≤10% mortality risk. The decision curve is constructed by varying R and plotting net benefit on the y axis against alternative values of R on the x axis. We calculated the net benefit of each prediction model and two reference strategies: treat none or treat all. The model with the highest net benefit at a particular risk threshold R enables us to treat as many high risk people as possible while avoiding harm from unnecessarily treating people at low risk.

In practice, treatment for breast cancer is given if it is expected to reduce the predicted risk by some desired magnitude, or more. For instance, clinicians in the Cambridge Breast Unit (Addenbrooke's Hospital, Cambridge, UK) use the absolute 10-year survival benefit from chemotherapy to guide decision making for adjuvant chemotherapy as follows: <3% no chemotherapy; 3-5% chemotherapy discussed as a possible option; >5% chemotherapy recommended (15). Assuming that chemotherapy reduces the 10-year risk of death by 22% (22), **a target risk reduction between 3 - 5%, corresponds to risks between 14% and 23% on the decision curves.** The R software, version 3.3.3 (20) was used for data analysis.

Results

The demographic, pathologic and treatment characteristics for the training and validation sets are presented in table 1. Estimated coefficients and the probabilities of inclusion in the model for each interaction term are given in the Supplementary Materials (Tables S2-S5 and Figure S1). Subsequent performance characteristics are presented only for the validation data.

Calibration

Figure 1 shows the calibration plots of the four prognostic models. Individual plots with 95%CIs can be found in Supplementary Materials (Figure S2). CancerMath systematically underestimated the 10-year death probability, especially for the ER-negative tumours. The other models showed good calibration except for high risk individuals, whose risk was consistently overestimated by all models. Compared to PREDICT, CancerMath demonstrated worse calibration performance. PREDICT outperformed all other models.

Discrimination

Table 2 represents the percentage change in discriminative performance (PC_{AUC}) of the three models with respect to PREDICT. The PC_{AUC} was calculated as follows: $PC_{AUC}=(AUC_{model}-AUC_{PREDICT})/AUC_{PREDICT}\times100$. This allowed us to directly compare the between-model performance. The Weibull RJMCMC and Logistic RJMCMC models had superior discrimination to PREDICT for ER-positive subjects. Logistic RJMCMC also performed better than PREDICT in ER-negative tumours. However, none of these improvements were statistically significant. Again, CancerMath performed notably worse than PREDICT.

Decision curve analysis

Figure 2 demonstrates the decision curves for the four models. For risk thresholds ≤ 0.05 for ER-positive tumours and ≤ 0.15 for ER-negative tumours, no model provided benefit compared with a policy of treating all patients. For threshold probabilities around 0.55 or higher for ER-positive tumours and ≥ 0.60 for ER-negative tumours, the models also gave lower net benefit than a policy of treating no patients. Within the threshold range 0.05-0.55 for ER-positive tumours, and 0.15-0.60 for ER-negative tumours, CancerMath provided the lowest net benefit for both ER-positive and ER-negative tumours. All the other curves were very similar over these ranges, indicating that

the net-benefit from using any of these models to prescribe chemotherapy, compared to treating nobody, could be as high as 0.1, which is of equivalent value to treating an additional 10% of patients who would have died without treatment.

As already mentioned, risks between 14% and 23% correspond to absolute risk reductions from chemotherapy treatment, in the range 3-5%. In this range adjuvant chemotherapy is discussed as a treatment option (15). Risk >23% correspond to absolute risk reductions of 5% or more where chemotherapy is firmly recommended. We see that in both thresholds ranges PREDICT outperformed CancerMath. Individual decision curve plots with 95% CIs can be found in Supplementary Materials (Figure S3).

Discussion

Accurate prognostic tools are critical to personalised breast cancer treatment, allowing clinicians and patients to make more informed decisions about adjuvant therapy. Using several large European validation cohorts, we have performed the first detailed comparison of the two most widely used breast cancer prognostic models: PREDICT and CancerMath. We found that survival probabilities from PREDICT offer both improved accuracy and discrimination over CancerMath. Using decision curve analysis, we also demonstrate that the use of PREDICT to make treatment decisions offers greater clinical utility than CancerMath over a range of risk thresholds, including those at which adjuvant chemotherapy is currently recommended by clinicians in the UK. We also explored novel prognostic models, extending PREDICT to include interaction terms among the predictors. Our new models performed as well as PREDICT, but no better, suggesting that, in this setting, the incorporation of interaction terms offers no predictive benefit.

Although all models demonstrated good overall prediction of 10-year mortality for both ER-positive and negative tumours, they consistently overestimated risk in women with the worse

11

prognosis. This highlights the shortcoming of current prediction models, showing significant discrepancies especially in high-risk populations. Laas et al., (11) arrived at the same conclusion but using a smaller sample of ER-positive HER2-negative breast cancer subjects. This could have consequences for the individual patient if their survival probability for breast cancer is underestimated.

We further applied decision curve analysis because this offers important information beyond the standard performance metrics of discrimination and calibration (14,23). To our knowledge, this is the first study to take into account the clinical utility of treatment decisions in breast cancer risk models. The main clinical utility of breast cancer prognostic models is to facilitate the decision of whether a patient will benefit from systemic treatment or not. Treatment is often prescribed at a risk threshold, which implies the clinician or patient's relative value of unnecessary treatment vs worthwhile use. From our results, both our models and PREDICT would provide the greatest clinical utility when recommending treatment to patients with risks falling in the range of 0.05 and 0.55 for ER-positive tumours and between 0.15 and 0.60 for ER-negative tumours. It has been suggested that when the absolute 10-year survival benefit from chemotherapy is 3-5%, chemotherapy is discussed as a treatment option (15). We demonstrated that in this range PREDICT outperformed CancerMath. Note that these calculations assumed a 22% risk reduction when treated with second generation chemotherapy (22). However, it is important to emphasize that there are no studies to date that have evaluated what would be an acceptable range of threshold probabilities that would help with adjuvant therapy recommendation.

Several potential limitations of our study should be considered. First, our comparison focused on web-based models, which are readily available for use in daily practice and incorporate only a modest set of risk factors which would be measured as a matter of course. We also

acknowledge that other models may provide more accurate predictions, for example by incorporating more biomarker information, such as genetics or other "omics". Second, some data were missing from the validation cohorts, such as the chemotherapy generation, which were required in the PREDICT and CancerMath models. As an approximation, we assumed that all patients received second generation chemotherapy since the data were older than 10 years. Since this same approximation was used in the fitting and validation of all models, we think it is unlikely to add much bias to their relative performance. We also note that any patient with missing information (predictors or outcome) was excluded from the training and validation data. Overall this was a small number of patients (<5%) so is unlikely to have biased our results.

In summary, we have compared two widely used prediction models for survival in patients with early stage breast cancer, PREDICT and CancerMath. The former outperformed the latter in discriminatory accuracy, calibration and clinical utility. We also developed novel prognostic models which performed as well as PREDICT, indicating that the incorporation of interaction terms offers only marginal predictive benefit. We have shown that PREDICT offers robust predictive performance, equivalent to our more complex modelling strategy, and for the first time quantified its clinical utility in identifying high risk women that would benefit from adjuvant treatment. We did not elicit the relative benefits and harms of treatment directly from patients or clinicians, but inferred them from the risk threshold currently used to prescribe chemotherapy. Further research will be required to more precisely determine relative utilities of treatment vs no treatment, which could lead to new risk thresholds to focus decision curve analysis on.

Acknowledgments

Funding

References

1. Engelhardt EG, Garvelink MM, de Haes JHCJM, et al. Predicting and Communicating the Risk of Recurrence and Death in Women With Early-Stage Breast Cancer: A Systematic Review of Risk Prediction Models. *J Clin Oncol*. 2013;32(3):238-250. doi:10.1200/JCO.2013.50.3417.

2. Wishart GC, Azzato EM, Greenberg DC, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res*. 2010;12(1):R1. doi:10.1186/bcr2464.

3. Wishart GC, Bajdik CD, Dicks E, et al. PREDICT Plus: development and validation of a prognostic model for early breast cancer that includes HER2. *Br J Cancer*. 2012;107(5):800-807. doi:10.1038/bjc.2012.338.

4. Wishart GC, Bajdik CD, Azzato EM, et al. A population-based validation of the prognostic model PREDICT for early breast cancer. *Eur J Surg Oncol*. 2011;37(5):411-417. doi:10.1016/j.ejso.2011.02.001.

5. De Glas NA, Bastiaannet E, Engels CC, et al. Validity of the online PREDICT tool in older patients with breast cancer: a population-based study. *Br J Cancer*. 2016;114(4):395.

6. Wong H-S, Subramaniam S, Alias Z, et al. The predictive accuracy of PREDICT: a personalized

decision-making tool for Southeast Asian women with breast cancer. *Medicine (Baltimore)*. 2015;94(8):e593. doi:10.1097/MD.0000000000000593.

7.   Candido dos Reis FJ, Wishart GC, Dicks EM, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res*. 2017;19(1):58. doi:10.1186/s13058-017-0852-3.

8.   Chen LL, Nolan ME, Silverstein MJ, et al. The impact of primary tumor size, lymph node status, and other prognostic factors on the risk of cancer death. *Cancer*. 2009;115(21):5071-5083.

9.   Michaelson JS, Chen LL, Bush D, Fong A, Smith B, Younger J. Improved web-based calculators for predicting breast carcinoma outcomes. *Breast Cancer Res Treat*. 2011;128(3):827-835.

10. Miao H, Hartman M, Verkooijen HM, et al. Validation of the CancerMath prognostic tool for breast cancer in Southeast Asia. *BMC Cancer*. 2016;16(1):820.

11. Laas E, Mallon P, Delomenie M, et al. Are we able to predict survival in ER-positive HER2-negative breast cancer? A comparison of web-based models. *Br J Cancer*. 2015;112(5):912-917.

12. Vickers AJ, Cronin AM. Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework. In: *Seminars in Oncology*. Vol 37. ; 2010:31-38.

13. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *bmj*. 2016;352:i6.

14. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decis Mak*. 2006;26(6):565-574. doi:10.1177/0272989X06295361.

15. Down SK, Lucas O, Benson JR, Wishart GC. Effect of PREDICT on chemotherapy/trastuzumab

recommendations in HER2-positive patients with early-stage breast cancer. *Oncol Lett*. 2014;8(6):2757-2761.

16. Wishart GC, Rakha E, Green A, et al. Inclusion of KI67 significantly improves performance of the PREDICT prognostication and prediction model for early breast cancer. *BMC Cancer*. 2014;14:908. doi:10.1186/1471-2407-14-908.

17. Engelhardt EG, van den Broek AJ, Linn SC, et al. Accuracy of the online prognostication tools PREDICT and Adjuvant! for early-stage breast cancer patients younger than 50 years. *Eur J Cancer*. 2017;78:37-44. doi:10.1016/j.ejca.2017.03.015.

18. Copson E, Eccles B, Maishman T, et al. Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: The POSH study. *J Natl Cancer Inst*. 2013;105(13):978-988. doi:10.1093/jnci/djt134.

19. Newcombe P, Raza Ali H, Blows F, et al. Weibull regression with Bayesian variable selection to identify prognostic tumour markers of breast cancer survival. *Stat Methods Med Res*. 2017;26(1):414-436. doi:10.1177/0962280214548748.

20. R Core Team. R: A Language and Environment for Statistical Computing. *R Found Stat Comput*. 2017:{ISBN} 3-900051-07-0, http://www.R-project.org. doi:http://www.R-project.org/.

21. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med*. 2014;33(3):517-535. doi:10.1002/sim.5941.

22. Group EBCTC, others. Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100 000 women in 123 randomised trials. *Lancet*. 2012;379(9814):432-444.

23. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for

development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931.

24. Green PJ. Reversible jump Markov chain monte carlo computation and Bayesian model determination. *Biometrika*. 1995;82(4):711-732. doi:10.1093/biomet/82.4.711.

25. Miller A. *Subset Selection in Regression*. CRC Press; 2002.

26. Group EBCTC, others. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;365(9472):1687-1717.
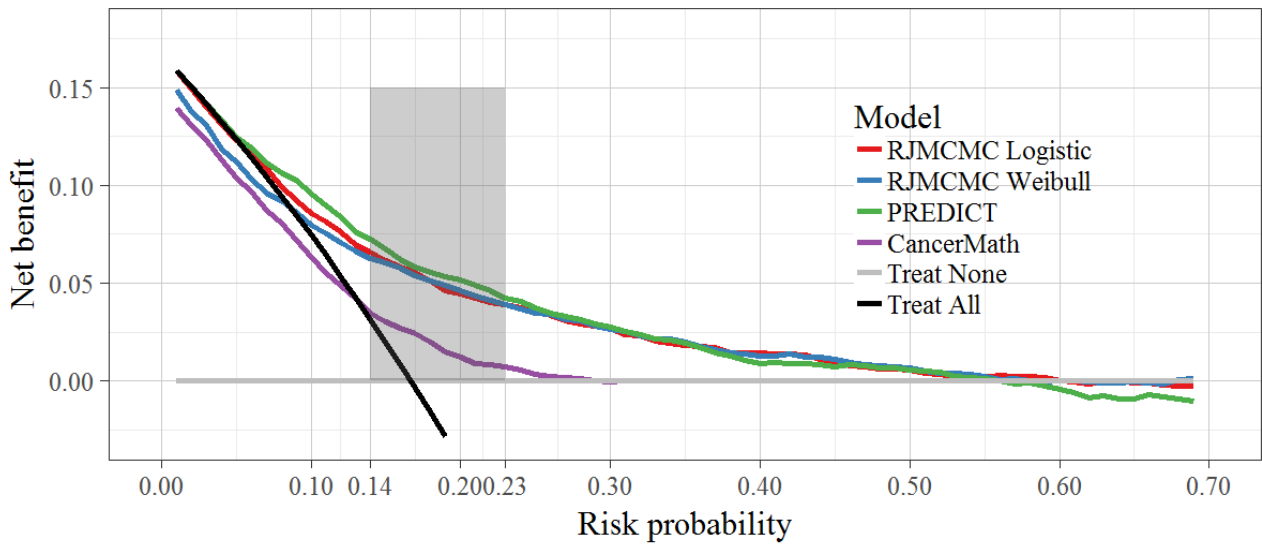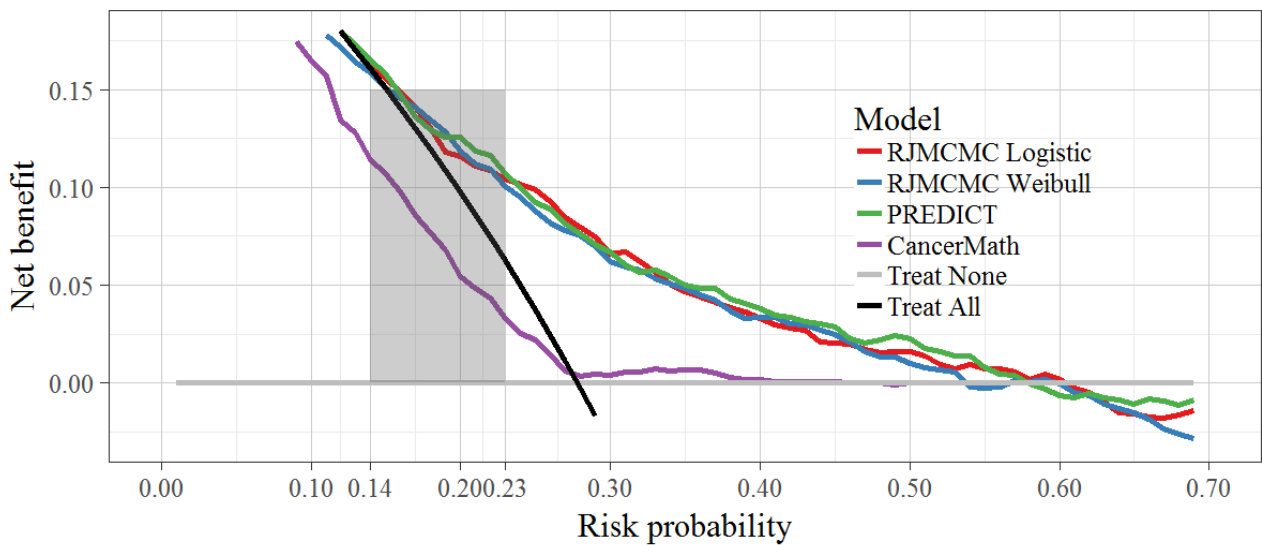
Figures
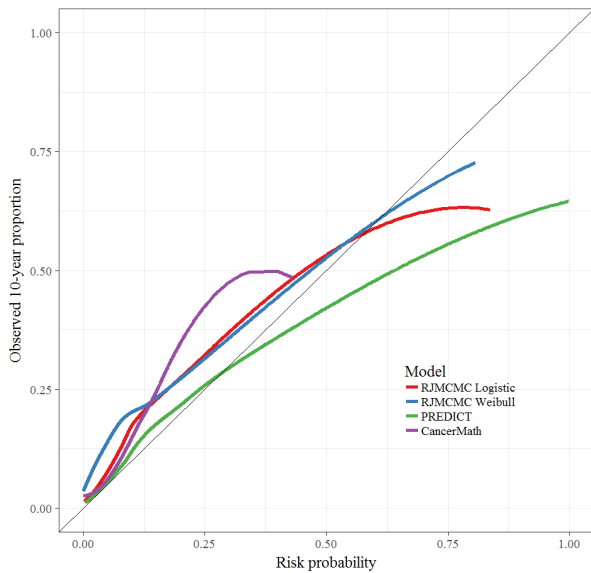
(A) ER positive
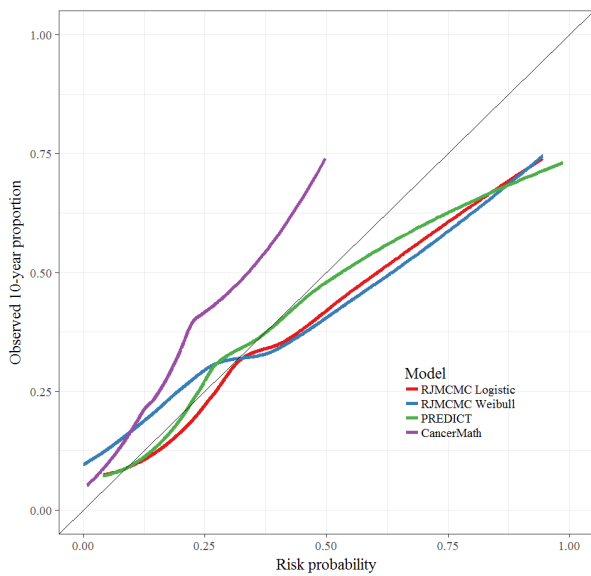
(B) ER negative

Figure 1: Calibration plots of the four models in the validation set for (A) ER positive, and (B)

ER negative tumours. Ten-year mortality risks based on the four prognostic models are plotted

against the observed mortality rate. The 45-degree line represents the perfect calibration slope.

(A) ER positive

(B) ER negative

Figure 2: Decision curve analysis for 10-year mortality for each model for (A) ER positive, and (B)

ER negative tumours. The y axis measures net benefit. The 0.14 to 0.23 shaded area on the x axis

corresponds to 3-5% absolute risk of death reduction with and without chemotherapy assuming a

relative risk reduction of 22% if treated. These are the risk ranges where chemotherapy is

discussed as a treatment option. A model is of clinical value if it has the highest net benefit

compared with the simple strategies of treating all patients (black line) or no patients (horizontal gray line) across the full range of threshold probabilities at which a patient would choose to be treated. The units on the y axis may be interpreted as the benefit associated with one patient who would die without treatment and who receives therapy.

Tables

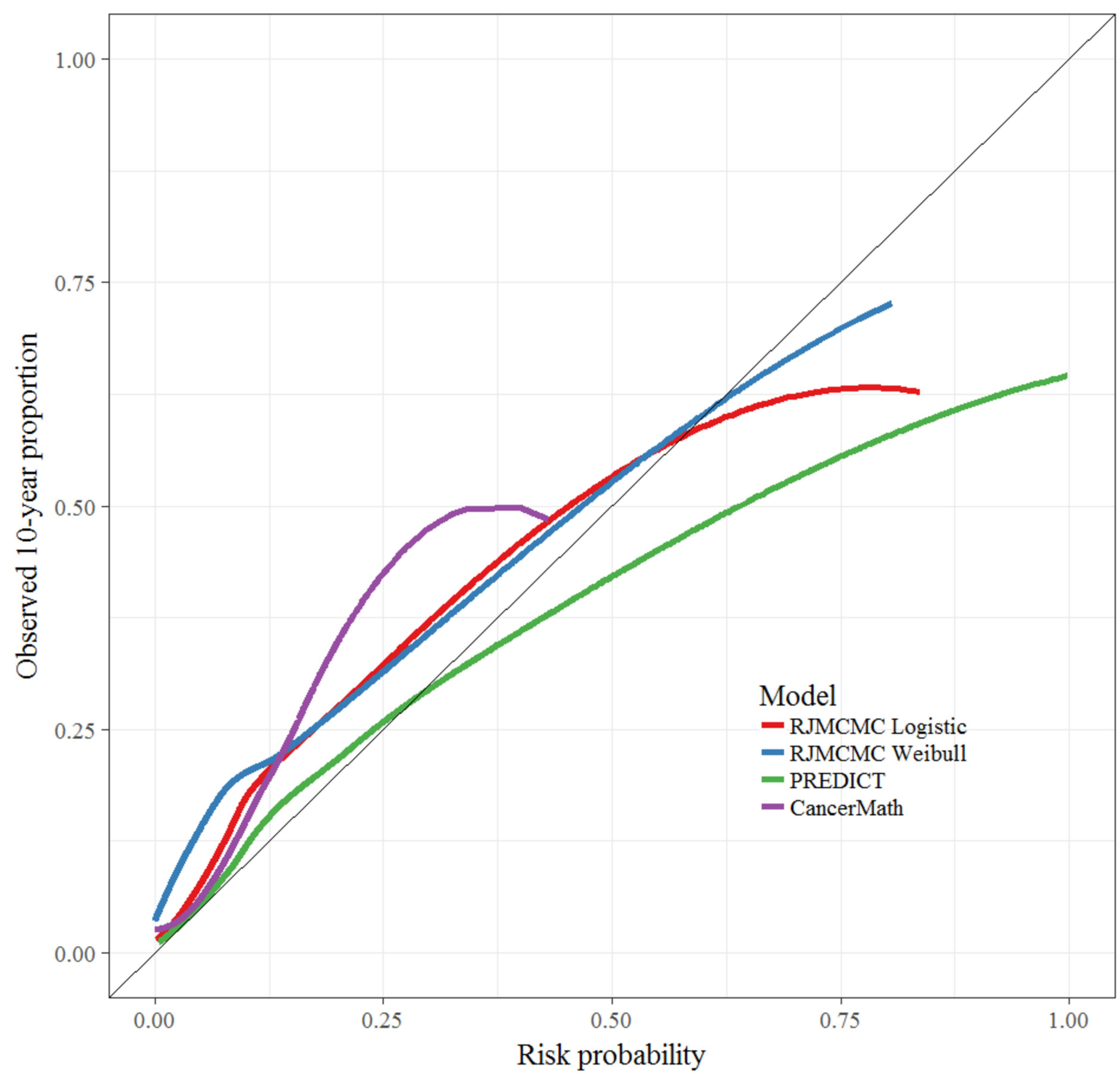Table 1: Characteristics of women in the training and validation data sets.

	Training data	Validation data
Total number of women	5729	5534
Breast cancer deaths, n (%)	994 (17)	1079 (20)
Missing breast cancer deaths, n (%)	0 (0)	118 (2)
Other deaths, n (%)	680 (12)	208 (4)
Missing other deaths, n (%)	0 (0)	118 (2)
Median age at diagnosis (IQR), years	58 (17)	40 (13)
Median tumor size (IQR), mm	19 (13)	20 (15)
Missing tumor size, n (%)	0 (0)	157 (3)
Median number of nodes (IQR)	0 (2)	0 (2)
Grade, n (%)		
I	987 (18)	702 (13)

II	2932 (51)	1918 (35)
III	1810 (32)	2914 (53)
ER status, n (%)		
Positive	4711 (82)	3879 (70)
Negative	1018 (18)	1655 (30)
Adjuvant therapy, n (%)		
None	735 (13)	1310 (24)
Chemo	909 (16)	1562 (28)
Hormone	2981 (52)	1020 (18)
Both	1104 (19)	1642 (30)

Table 2: Percentage change in AUC (PC$_{AUC}$) between PREDICT and each of the other models. The 95% confidence intervals (CI) are based on 1000 bootstrap replicates

	ER positive		ER negative	
PREDICT vs	PC$_{AUC}$ (%)	95% CI	PC$_{AUC}$ (%)	95% CI
CancerMath	-4.27	-6.39 – -2.03	-3.21	-5.9 – -0.48
Logistic RJMCMC	0.04	-1.04 – 1.12	0.19	-0.19 – 0.57
Weibull RJMCMC	0.06	-0.96 – 1.11	-0.4	-0.83 – 0.03