



**HAL**  
open science

## Direct comparison of logistic regression and recursive partitioning to predict chemotherapy response of breast cancer based on clinical pathological variables

Roman Rouzier, Charles Coutant, Bénédicte Lesieur, Chafika Mazouni,  
Roberto Incitti, René Natowicz, Lajos Pusztai

### ► To cite this version:

Roman Rouzier, Charles Coutant, Bénédicte Lesieur, Chafika Mazouni, Roberto Incitti, et al.. Direct comparison of logistic regression and recursive partitioning to predict chemotherapy response of breast cancer based on clinical pathological variables. *Breast Cancer Research and Treatment*, 2009, 117 (2), pp.325-331. 10.1007/s10549-009-0308-2 . hal-00535333

**HAL Id: hal-00535333**

**<https://hal.science/hal-00535333>**

Submitted on 11 Nov 2010

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Direct comparison of logistic regression and recursive partitioning to predict chemotherapy response of breast cancer based on clinical pathological variables

Roman Rouzier · Charles Coutant ·  
Bénédicte Lesieur · Chafika Mazouni ·  
Roberto Incitti · René Natowicz · Lajos Pusztai

Received: 2 January 2009 / Accepted: 5 January 2009 / Published online: 17 January 2009  
© Springer Science+Business Media, LLC. 2009

**Abstract** The purpose was to compare logistic regression model (LRM) and recursive partitioning (RP) to predict pathologic complete response to preoperative chemotherapy in patients with breast cancer. The two models were built in a same training set of 496 patients and validated in a same validation set of 337 patients. Model performance was quantified with respect to discrimination (evaluated by the areas under the receiver operating characteristics curves (AUC)) and calibration. In the training set, AUC were similar for LRM and RP models (0.77 (95% confidence interval, 0.74–0.80) and 0.75 (95% CI, 0.74–0.79), respectively) while LRM outperformed RP in the validation set (0.78 (95% CI, 0.74–0.82) versus 0.64 (95% CI, 0.60–0.67). LRM model also outperformed RP model in term of

calibration. In these real datasets, LRM model outperformed RP model. It is therefore more suitable for clinical use.

**Keywords** Breast cancer · Prediction · Pathological complete response · Logistic regression model · Recursive partitioning model

## Introduction

The use of adjuvant chemotherapy is influenced by three important considerations. These include estimation of risk of relapse (prognosis), probability of sensitivity to therapy (response prediction) and risk of adverse events from treatment. Several prognostic risk prediction models exist that combine various clinical and pathological variables into a prognostic score [1–13]. Adjuvant online is perhaps the most commonly employed and the most user friendly among these tools ([www.adjuvantonline.org](http://www.adjuvantonline.org)) [8]. We previously developed a similar clinical pathological variable based chemotherapy response prediction model [9]. This model was built by using data from patients who received preoperative, neoadjuvant chemotherapy and therefore tumor response to treatment could be directly measured. Our clinical response prediction nomogram was subsequently validated on two sets of independent cases from two different institutions and proved to be rather accurate [9]. However, there are several different mathematical ways to combine variables into a multivariable predictor. The most often used model is logistic regression that is a form of generalized linear models and allows one to predict a discrete outcome, such as group membership (i.e. response versus no response or recurrence versus no recurrence), from a set of dependent variables that may be continuous, discrete, dichotomous, or a mix of any of these. Variables are

---

R. Rouzier (✉) · C. Coutant · B. Lesieur  
Department of Obstetrics and Gynecology, Hôpital Tenon,  
Assistance Publique des Hôpitaux de Paris, Pierre et Marie Curie  
University – Paris 6, 4 rue de la Chine, 75020 Paris Cedex,  
France  
e-mail: roman.rouzier@tnn.aphp.fr

R. Rouzier · C. Mazouni  
Breast Cancer Unit, Institut Gustave Roussy, 39 rue Desmoulins,  
94805 Villejuif Cedex, France

R. Rouzier · C. Coutant · C. Mazouni · L. Pusztai  
Department of Breast Medical Oncology, Unit 1354,  
The University of Texas, M.D. Anderson Cancer Center,  
P.O. Box 301439, Houston, TX, USA

R. Incitti  
Faculté de Médecine, Institut Mondor de Médecine Moléculaire  
(IFR10), University Paris 12, Créteil 94000, France

R. Natowicz  
Computer Sciences Department, University of Paris – Est.  
ESIEE-Paris, Cité Descartes, BP.99, 93162 Noisy-le-Grand,  
France

combined in a linear manner. However, non-linear models may be more pertinent for some situations. Recursive partitioning is an example of nonlinear predictors. In recursive partitioning, data are divided on the basis of the possession of specified attributes. In contrast to logistic regression, recursive partitioning is a nonparametric type of analysis that repeatedly subdivides data into smaller and smaller subgroups based on characteristics that predict the desired endpoint. The goal is to construct subgroups that, ideally, consist entirely of subjects with one endpoint category or another. Recursive partitioning, unlike logistic regression is nonlinear in its parameters, and would have an advantage if the true relationship between the variables and the outcome of interest is nonlinear. However, if the true relationship is linear, then recursive partitioning may be inferior to linear models because it is based on assumptions that are too general. Several studies have compared the performance of linear and non-linear predictors in model building, however few studies compared the performance of distinct models built from the same data and tested on independent validation sets [4, 5, 11, 12]. Comparing various strategies for multivariable model development may allow a better understanding of the strengths and limitations of these models.

The goal of the current work, was to compare our previously reported logistic regression model using clinical variables to predict pathologic complete response to pre-operative chemotherapy with a non-linear recursive partitioning model that was built from the same discovery set ( $n = 496$ ) and tested on the same validation set ( $n = 337$ ).

## Materials and methods

In institutional clinical databases, we identified 496 patients from the Institut Gustave Roussy (IGR) in Villejuif, France, and 337 women from the University of Texas M.D. Anderson Cancer Center (MDACC) in Houston, Texas, United States, diagnosed with breast cancer and treated with an anthracycline-based primary chemotherapy. All the women gave informed written consent to therapeutic procedures and to the analysis of data related to their malignancy in accordance with Institutional Review Board institutional guidelines and the Declaration of Helsinki. The cohort treated at IGR included 496 patients who received 3 or 4 courses of anthracycline-based preoperative chemotherapy and was used as training set to develop the predictive models. The second cohort included 337 patients treated at MDACC with 4 courses of anthracycline-based preoperative chemotherapy and was used as a validation set. The clinical and histological characteristics were prospectively recorded into databases. The largest tumor dimension at clinical examination was recorded as the tumor diameter. Histologic grade, defined according to the modified Scarff, Bloom, and

Richardson system described by Contesso et al. [14], was used at IGR, and the modified Black's nuclear grade was used at MDACC. All the patients underwent axillary lymph node dissection and mastectomy or segmental mastectomy. Patients without residual invasive tumor or with only strictly in situ carcinoma were classified as having pathological complete response (pCR) in the breast.

Patient characteristics and chemotherapy modalities have been reported previously [9]. Table 1 describes the characteristics of the population, as well as tumor and histological findings on the 496 patients of the training set and on the 337 patients of the validation set. Patients had operable breast cancer: 3 (1%), 293 (59%), 161 (32%) and 39 (8%) patients had T0–1, T2, T3 and T4 tumors in the training set and 13 (4%), 109 (32%), 90 (27%) and 125 (37%) patients had T0–1, T2, T3 and T4 tumors in the validation set. Forty-three percent and 25% of patients had a negative clinical nodal status in the training set and the validation set, respectively. Twenty-nine percent and 46% of patients had estrogen receptor (ER) negative tumors in the training set and the validation set, respectively. Seven percent, 57, and 36% in the training set and 5, 40, and 55% in the validation set had grade 1, grade 2, and grade 3 tumors, respectively. Eighty-nine percent and 95% of patients had ductal tumors in the training set and the validation set, respectively.

In order to develop a well-calibrated logistic regression-based nomogram and a recursive partitioning model, we built these models in a training cohort and validated them in an independent validation cohort. We tested clinicopathological characteristics (age, tumor size, nodal status, histologic type and grade, ER status, multifocality and number of courses of preoperative chemotherapy) in a multivariate analysis for association with pCR. The performance of the multivariable model was assessed with respect to discrimination and calibration.

Discrimination (i.e. whether the relative ranking of individual predictions was in the correct order) was quantified with the the area under (AUC) the receiver-operating characteristic curve (ROC) which can range from 0 to 1 (1 indicating perfect concordance, 0.5 indicating no association and 0 indicating perfect discordance) [15].

Calibration corresponds to the agreement between observed outcome frequencies and predicted probabilities. Results are displayed as a calibration curve that shows the relationship between the observed outcome frequencies and the predicted probabilities for groups of patients defined by quartiles in the logistic regression model or by the final leaf in the recursive partitioning model. A calibration curve can be approximated by a regression line with intercept  $\alpha$  and slope  $\beta$ . These parameters can be estimated in an LRM with the event as outcome and the linear predictor as the only covariate. Well-calibrated models have  $\alpha = 0$  and  $\beta = 1$ . Therefore, a sensible measure of calibration is a likelihood

**Table 1** Clinical and pathological data for the 496 patients of the training set and the 337 patients of the validation set

Clinical and pathological data	Training set ( <i>n</i> = 496 patients)	Validation set ( <i>n</i> = 337 patients)
Center	IGR	MDACC
<i>Age</i>		
Mean age, years (SD)	52 (10)	52 (11)
<i>T classification</i>		
T0–1	3 (1%)	13 (4%)
T2	293 (59%)	109 (32%)
T3	161 (32%)	90 (27%)
T4	32 (8%)	125 (37%)
<i>Clinical nodal status</i>		
Negative	213 (43%)	85 (25%)
Positive	2,837 (57%)	250 (75%)
<i>Histology</i>		
Invasive ductal carcinoma/others	440 (89%)	321 (95%)
Invasive lobular carcinoma/mixed	56 (11%)	16 (5%)
<i>Tumor grade</i>		
Well differentiated (grade 1)	37 (7%)	17 (5%)
Moderately differentiated (grade 2)	282 (57%)	136 (40%)
Poorly differentiated (grade 3)	177 (36%)	184 (55%)
<i>Estrogen receptor status</i>		
Negative	144 (29%)	155 (46%)
Positive	353 (71%)	182 (54%)

IGR Institut Gustave Roussy,  
MDACC M.D. Anderson Cancer  
Center, SD standard deviation

ratio statistic testing the null hypothesis that  $\alpha = 0$  and  $\beta = 1$ . The statistic has a  $\chi^2$  distribution with 2 degrees of freedom (unreliability [*U*]-statistic). Individual predictions were either calculated from nomograms or were obtained from the original data for the RP model. We also evaluated average error (for quartiles) [*E* average ( $E_{\text{aver}}$ )] and maximal error (for individual predictions) [*E* maximal ( $E_{\text{max}}$ )] between predictions and observations obtained from a calibration curve. This gives an idea of model performance when extrapolated to new patient populations. This is of particular importance for clinical practice because probabilities are announced to patients without a confidence interval. To indicate a probability,  $\pm$ maximal error or average error is more appropriate than providing only a probability.

All analyses were performed using the R package with the Design, Hmisc, Rpart and Verification libraries (<http://lib.stat.cmu.edu/R/CRAN/>).

## Results

### Models development

Both models were developed from the same data set (IGR) including 496 patients. Forty-five (9%) patients had pathological complete response (pCR) and 451 (91%) had residual disease in this training set. In multivariate logistic regression analysis, initial *T* stage (TNM), ER status, grade,

number of course of preoperative chemotherapy were independently associated with pCR. We added age to the final model because it improved the calibration of the logistic regression model. The equation of probability of achieving pCR was  $P = 1/(1 + \exp(-X))$ . Where  $X = -5.85766 + 3 * V1 - 0.02247 * V2 - 0.46632 * V3 + 0.91987 * V4 - 0.85084 * V5$  and V1 was the number of preoperative course (3 or 4), V2 was age in years, V3 was *T* of TNM classification (0, 1, 2, 3 or 4), V4 was histological grade (1, 2 or 3), and V5 was estrogen receptor status (0 or 1).

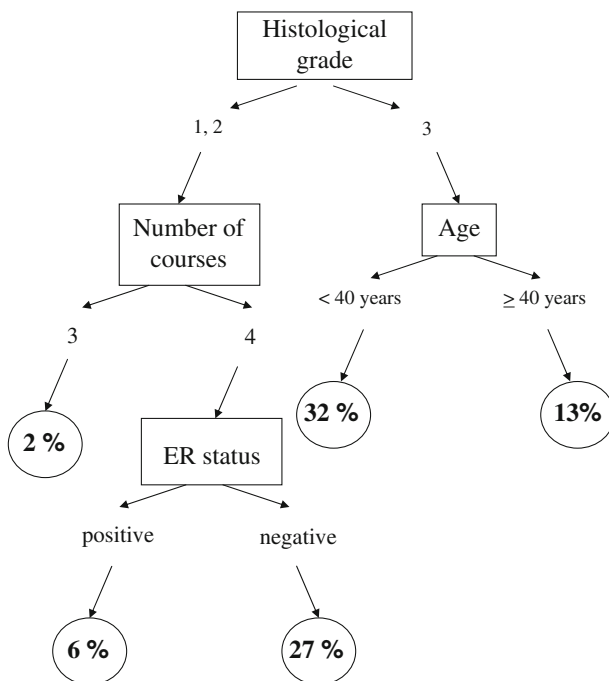
This clinical response prediction prediction model is freely accessible online [9] at [http://www.mdanderson.org/care\\_centers/breastcenter/dIndex.cfm?pn=448442B2-3EA5-4BAC-98310076A9553E63](http://www.mdanderson.org/care_centers/breastcenter/dIndex.cfm?pn=448442B2-3EA5-4BAC-98310076A9553E63).

An example of the screen is shown on Fig. 1. The website also includes a calculator to estimate probability of breast conserving surgery and probability of residual tumor less than 3 cm after neoadjuvant chemotherapy. We next developed a recursive partitioning model. Recursive partitioning analysis created a tree that used grade, ER status, age and number of cycles of preoperative chemotherapy. The tree is represented in Fig. 2.

### Comparison of the two pCR prediction models

We compared the discrimination and the calibration of the two distinct models in both the original training data and independent validation data set. The validation set included

**Fig. 1** Example of screen from the computer program to predict pathological complete response after preoperative chemotherapy for breast cancer



ER = Estrogen receptor, Number of course of chemotherapy.

**Fig. 2** Recursive partitioning model developed from our training set to predict pathological complete response after preoperative chemotherapy for breast cancer

337 patients from MDACC, the rate of pCR and residual disease were 12.7% (43 patients) and 87.3% (296 patients), respectively.

### Discrimination

In the training set, the AUC of the ROC curve obtained with the logistic regression model and with the recursive partitioning model were similar, 0.77 (95% CI, 0.74–0.80;  $P < 10^{-3}$ ) and 0.75 (95% CI, 0.74–0.79;  $P < 10^{-3}$ ), respectively (Table 2; Fig. 3). In the validation set, the AUC of the ROC curve for the logistic regression model remained good, 0.78 (95% CI, 0.74–0.82;  $P < 10^{-3}$ ), however the AUC of the ROC curve for the recursive partitioning model was lower: 0.64 (95% CI, 0.60–0.67) (Table 2; Fig. 4).

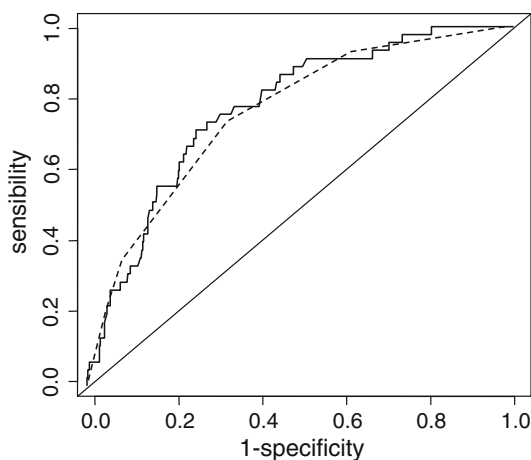
### Calibration

We evaluated the calibration of the two models on the validation set. The calibration of the logistic regression model was excellent with no statistical difference between predicted probability and observed probability ( $P = 0.24$ ). The average difference ( $E_{\text{aver}}$ ) and maximal difference ( $E_{\text{max}}$ ) in predicted and calibrated probabilities were 1.6 and  $1.6 \cdot 10^{-14}\%$ , respectively. The calibration of the recursive partitioning model was less good. However, there was also no statistical difference between predicted probability by the recursive partitioning tree and observed probability ( $P = 0.12$ ). The average difference ( $E_{\text{aver}}$ ) and maximal difference ( $E_{\text{max}}$ ) in predicted and calibrated probabilities were 6.3 and 27.5%, respectively, (Table 2; Fig. 5).

**Table 2** Comparison of logistic regression model and recursive partitioning model to predict chemotherapy response of breast cancer based on clinical pathological variables according to discrimination and calibration

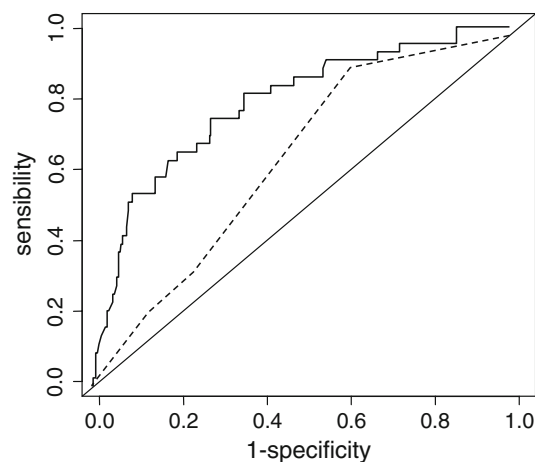
Models	Nb of patients	Discrimination AUC (95% CI)	Calibration		
			P value	$E_{\max}^a$ (%)	$E_{\text{aver}}^b$ (%)
<i>Logistic regression model</i>					
Training set (IGR cohort)	496	0.77 (0.74–0.80)	Not adequate		
Validation set (MDACC cohort)	337	0.78 (0.74–0.82)	0.24	$10^{-14}$	1.6
<i>Recursive partitioning model</i>					
Training set (IGR cohort)	496	0.75 (0.74–0.79)	Not adequate		
Validation set (MDACC cohort)	337	0.64 (0.60–0.67)	0.12	27.5	6.3

These two models were built from the same training set and tested on the same validation set. *IGR* Institut Gustave Roussy, *MDACC* M.D. Anderson cancer center, *Nb* number, *E* difference in predicted and calibrated probabilities between calibration and AUC,  $E_{\max}$  maximal error,  $E_{\text{aver}}$  average error, *AUC* Area under the ROC curve, *ROC* receiver operating characteristic, *CI* confidence interval,  $E_{\max}^a$  for individual predictions,  $E_{\text{aver}}^b$  for quartiles



Models (Training set)	AUC (95% CI)
— Logistic regression model	0.77(0.74-0.80)
- - - Recursive partitioning model	0.75 (0.74-0.79)

AUC = Area under the ROC curve, ROC = receiver operating characteristic, CI = confidence interval,

**Fig. 3** Receiver operating characteristics (ROC) curves of the logistic regression model and the recursive partitioning model for the 496 patients of the training set

Models (Validation set)	AUC (95% CI)
— Logistic regression model	0.78(0.74-0.82)
- - - Recursive partitioning model	0.64 (0.60-0.67)

AUC = Area under the ROC curve, ROC = receiver operating characteristic, CI = confidence interval,

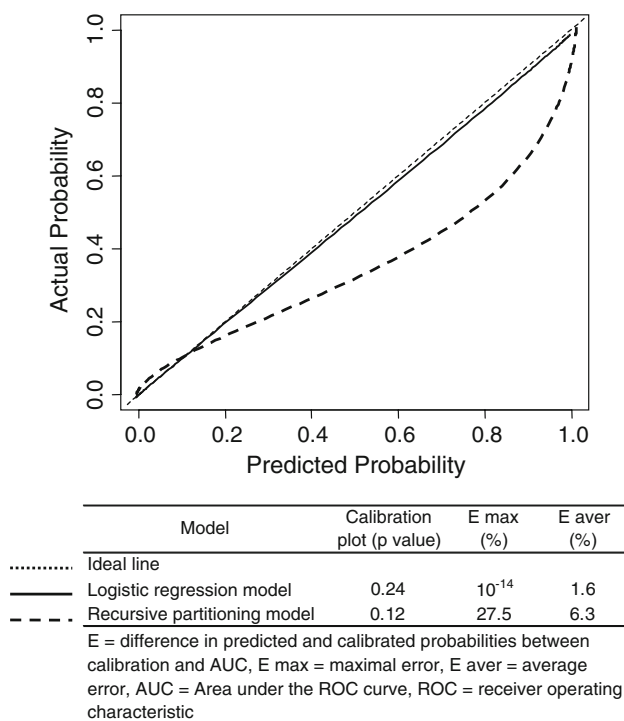
**Fig. 4** Receiver operating characteristics (ROC) curves of the logistic regression model and the recursive partitioning model for the 337 patients of the validation set

## Discussion

In these real datasets, the logistic regression model predicted breast tumor chemosensitivity to primary chemotherapy was better than recursive partitioning. This finding is of particular interest because the goal of these models is to develop a predictive model for clinical use and the highest possible accuracy should be the primary objective. Several mathematical models are available to combine variables into multivariate predictors and one method might be better than the other in particular situations [1–6, 9–13]. In his study, we compared not only discrimination but also calibration of both a logistic regression-based nomogram and a recursive

partitioning-based model. The logistic regression model that we developed is based on five input variables, including age, ER status, histologic grade, *T* stage and number of preoperative course. The recursive partitioning model includes age, ER status, grade and number of courses of preoperative of chemotherapy. Age was added in the logistic regression model even though it was non significant in univariate analysis because it improved calibration of that model. The development of logistic regression is easier than recursive partitioning. On the other hand, the tree that results from recursive partitioning allows an easy visualization of the decision steps and it is more easily interpretable in the biological context. The recursive-partitioning tree in Fig. 2





**Fig. 5** Calibration plot of the logistic regression model and the recursive partitioning model for the 337 patients of the validation set

shows for example that ER status is useful to distinguish chemosensitive low grade tumors but not high grade tumors.

In the literature, there are reports of head-to-head comparison between logistic regression and other models such as artificial neural networks or recursive partitioning when applied to training data and used for model building [4, 12, 13]. However, few studies are able to compare the performance of such competing models in independent validation sets of patients. Our results indicate that the AUC under the ROC curve of the logistic regression-based nomogram 0.77 (95% CI, 0.74–0.80) was similar to the AUC of the recursive partitioning 0.75 (95% CI, 0.74–0.79) in the training set. The small 2% difference in predictive accuracy does not imply that the use of the nomogram instead of the recursive partitioning in 1,000 patients will result in 20 additional (2%) correctly classified patients (while this is the common belief [13]). Accuracy, or the predictive ability of a model, has two major components, discrimination and calibration. The AUC is similar to the c-index, which is a measure of discrimination, or the ability to separate two patients. The c-index is independent of the prior probability of each outcome. The c-index computation requires only that the algorithm produce an ordinal-scaled relative predictive score, not a true probability. The c-statistic is the probability that, for a randomly selected pair of subjects, one with favorable outcome (pCR in our study) and the other with unfavorable outcome (residual disease), the person with favorable outcome will have the higher estimated

probability according to the model. Therefore, the c-index reflects how accurate is the model to correctly predict, between two patients, which one will have the best outcome. In a prospective validation, calibration, or how well the predicted probabilities reflect actual risk, is another aspect of accuracy that is not captured by the c-index. A model could discriminate well but lack even internal calibration if the fitted scores do not reflect the true probability of an event. Actually, a well-calibrated model will also be discriminant. Our results demonstrate these elements. In our data, the logistic regression model was better calibrated and was therefore a better predictor of the probability of pCR as shown in Figs. 4, 5 and Table 2. These observations are in keeping with previous studies that demonstrated that tree-based models do not provide an improvement in predictive accuracy over logistic or Cox regression model [4, 12].

In conclusion, we demonstrate that the logistic regression-based nomogram that we previously reported is slightly more accurate than recursive partitioning.

## References

- Chollet P, Amat S, Belembaogo E et al (2003) Is Nottingham prognostic index useful after induction chemotherapy in operable breast cancer? *Br J Cancer* 89:1185–1191. doi:10.1038/sj.bjc.6601258
- Diblasio CJ, Kattan MW (2003) Use of nomograms to predict the risk of disease recurrence after definitive local therapy for prostate cancer. *Urology* 62(suppl 1):9–18. doi:10.1016/j.urology.2003.09.029
- Eastham JA, Kattan MW, Scardino PT (2002) Nomograms as predictive models. *Semin Urol Oncol* 20:108–115. doi:10.1053/suro.2002.32936
- Kattan MW (2003) Comparison of Cox regression with other methods for determining prediction models and nomograms. *J Urol* 170:S6–S9. doi:10.1097/01.ju.0000094764.56269.2d (discussion S10)
- Kattan MW (2003) Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: preoperative application in prostate cancer. *Curr Opin Urol* 13:111–116. doi:10.1097/00042307-200303000-00005
- Kattan MW, Giri D, Panageas KS et al (2004) A tool for predicting breast carcinoma mortality in women who do not receive adjuvant therapy. *Cancer* 101:2509–2515. doi:10.1002/cncr.20635
- Pusztai L, Gianni L (2004) Technology insight: Emerging techniques to predict response to preoperative chemotherapy in breast cancer. *Nat Clin Pract Oncol* 1:44–50. doi:10.1038/nponc0025
- Ravdin PM, Siminoff LA, Davis GJ et al (2001) Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 19:980–991
- Rouzier R, Pusztai L, Delalage S et al (2005) Nomograms to predict pathologic complete response and metastasis-free survival after preoperative chemotherapy for breast cancer. *J Clin Oncol* 23:8331–8339. doi:10.1200/JCO.2005.01.2898
- Rouzier R, Pusztai L, Garbay JR et al (2006) Development and validation of nomograms for predicting residual tumor size and the probability of successful conservative surgery with neoadjuvant chemotherapy for breast cancer. *Cancer* 107:1459–1466. doi:10.1002/cncr.22177

11. Shariat SF, Capitanio U, Jeldres C, Karakiewicz PI (2008) Can nomograms be superior to other prediction tools? *BJU Int*
12. Shariat SF, Karakiewicz PI, Suardi N, Kattan MW (2008) Comparison of nomograms with other methods for predicting outcomes in prostate cancer: a critical analysis of the literature. *Clin Cancer Res* 14:4400–4407. doi:[10.1158/1078-0432.CCR-07-4713](https://doi.org/10.1158/1078-0432.CCR-07-4713)
13. Chun FK, Graefen M, Briganti A et al (2007) Initial biopsy outcome prediction—head-to-head comparison of a logistic regression-based nomogram versus artificial neural network. *Eur Urol* 51:1236–1240. doi:[10.1016/j.eururo.2006.07.021](https://doi.org/10.1016/j.eururo.2006.07.021) (discussion 1241–1233)
14. Contesso G, Mouriessse H, Friedman S, Genin J, Sarrazin D, Rouesse J (1987) The importance of histologic grade in long-term prognosis of breast cancer: a study of 1,010 patients, uniformly treated at the Institut Gustave-Roussy. *J Clin Oncol* 5:1378–1386
15. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36