Breast Cancer Research and Treatment (2005) 94: 265–272
DOI 10.1007/s10549-005-9013-y

Improvement of breast cancer relapse prediction in high risk intervals using artificial neural networks

J.M. Jerez[1], L. Franco[1,2], E. Alba[3], A. Llombart-Cussac[4], A. Lluch[5], N. Ribelles[3], B. Munárriz[6], and M. Martín[7]

[1]*Departamento de Lenguajes y Ciencias de la Computación, Universidad de Málaga, Málaga, Spain;* [2]*Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom;* [3]*Servicio de Oncología Médica, Hospital Clínico Universitario Virgen de la Victoria, Málaga, Spain;* [4]*Servicio de Oncología Médica, Instituto Valenciano de Oncología, Valencia, Spain;* [5]*Servicio de Oncología Médica, Hospital General Universitario de Valencia, Valencia, Spain;* [6]*Servicio de Oncología Médica, Hospital Universitario La Fe, Valencia, Spain;* [7]*Servicio de Oncología Médica, Hospital Clínico Universitario San Carlos, Madrid, Spain*

Key words: breast cancer prognosis, Cox proportional hazard model, high peaks of relapse, neural networks, survival analysis

Summary

The objective of this study is to compare the predictive accuracy of a neural network (NN) model versus the standard Cox proportional hazard model. Data about the 3811 patients included in this study were collected within the 'El Álamo' Project, the largest dataset on breast cancer (BC) in Spain. The best prognostic model generated by the NN contains as covariates age, tumour size, lymph node status, tumour grade and type of treatment. These same variables were considered as having prognostic significance within the Cox model analysis. Nevertheless, the predictions made by the NN were statistically significant more accurate than those from the Cox model ($p < 0.0001$). Seven different time intervals were also analyzed to find that the NN predictions were much more accurate than those from the Cox model in particular in the early intervals between 1–10 and 11–20 months, and in the later one considered from 61 months to maximum follow-up time (MFT). Interestingly, these intervals contain regions of high relapse risk that have been observed in different studies and that are also present in the analyzed dataset.

Introduction

Decisions about how to treat breast cancer patients after surgery have been contingent on the accuracy of estimating the behaviour and outcome of the disease. Histological, biochemical and clinical information have demonstrated to have prognostic utility, however predicting the disease outcome for an individual patient remains a challenging task. Regression models have traditionally been used to perform studies on the relationship between survival status and prognostic factors generally based on dichotomized variables and also by *a priori* assumptions on a linear relationship between the logarithm of the hazard and the covariates [1].

In survival analysis the outcomes are characterized by the time to an event occurrence, and thus the prediction of the clinical events requires a strategy to address the censored data (cases for which the event of interest does not occur), since simple exclusion of these observations would limit the amount of data and may lead to significant biases in event predictions. The Cox proportional hazard analysis [2] is an accepted solution to the problem of analysing censored data and has become the standard survival analysis statistical technique. However, the Cox model is mainly used to study the importance of covariates for survival, and only few authors in the literature apply the Cox model to make individual predictions of survival times [3,4]. The Cox model is a multivariate regression semi-parametric model that allows modelling of continuous covariates, and it involves that some assumptions need to be made: (1) The effect of covariates are additive and linear on the log risk scale, (2) if covariates interact with each other the regression model should include interactions terms, and (3) the relative risk between the hazard rate for two subjects is constant over time, in other words, there is no time interaction with the covariates. However, it is not well-known if these assumptions are true from a biological point of view. Moreover, some of these assumptions are not thoroughly checked in many different clinical situations where the Cox regression model is applied to survival data modelling.

Besides, a more informative analysis of the dynamics of breast cancer recurrence is actually demanded by

medical experts and patients, and this interest has leaded the scientific community to adopt non-linear modelling alternatives able to identify previously unknown prognostic relationships. In this situation, neural networks (NNs) are a well-known alternative to traditional logistic regression methods [3–5].

Several NN approaches have been proposed to model data with censorship. In some cases (see for example [6–10]) the prognostic covariates have been used as inputs to the neural system while the time to relapse is the output of the NN. These previous approaches implement a separation between the dependence on time and on the patient data resulting in non-linear proportional hazards models. A more efficient representation of time is to include it as a covariate, and in this case the output of the system becomes an indicator of relapse or not at a given time. This kind of approach (also refereed as time-coded models) has been proposed by several authors [11–16] and can be interpreted as the discrete time implementation of the proportional hazards model [3]. Moreover, this kind of NN model for survival prediction has proven to be very stable in monthly studies over follow-up periods of several years [17]. The present work uses time as an input variable and the output could be interpreted as the cumulative probability of relapse for individual patients [8].

The goal of this work is to compare the performance of NNs architectures versus traditional Cox regression analysis, in predicting the probability of breast cancer relapse (the event of interest in this study) and on the task of identifying relevant prognostic factors in breast cancer patients after surgical intervention. We have also carried a detailed analysis of the importance of including (or not) different clinical markers as inputs in the neural architectures. The analysis was carried both for the whole dataset but also limited interval of times were considered in order to assess the time dependency of the prognostic covariates and as a way to improve the accuracy of the system by selecting on each interval the most appropriate variables.

It is important to outline that real data sets with significant large numbers of cases have not been extensively analysed using NNs, since most of the studies reported in the literature use relatively small data sets [18]. In this sense, the present work results in one of the most extensive research studies carried out to predict the probability of breast cancer relapse on individual patients.

Materials and methods

Patients

Data were collected from the 'El Álamo' Project, the largest database on breast cancer in Spain. The dataset analyzed in this study includes demographics, therapeutic and recurrence-survival information from 3811 women patients with operable invasive breast cancer diagnosed in 32 different hospitals belonging to the Spanish Breast Cancer Research Group (GEICAM) between the years 1990 and 1993. The analysis was restricted to patients with follow-up time of at least 1 month, and 34% of patients relapsed in the period of study. The median follow-up (i.e. the time elapsed from the date of surgery to the last updating of the patient record) was 76 months (range 1–128 months). The original dataset contained 3811. As the data contained missing values, and these were not considered in the analysis, the size of the datasets used on the different models tested depends on the covariates analyzed in each particular case. Four prognostic variables (covariates) considered by the physicians as very significant prognostic markers were included in all models. When these four covariates (Age, Tumour Size, Number of Axillary lymph nodes and Histological Grade) are considered, the dataset results in 2004 patient records. The main characteristics of this dataset are reflected in Table 1.

Models

Artificial neural networks

Feed-forward NNs can be seen as analogous to regression models, in which covariates are called *inputs*, coefficients are called *synaptic weights*, and the outcome variable is called *output*. In this work, a three-layer NN model (an input layer, with each input node corresponding to a prognostic factor plus one node for the coded time; a hidden layer; and an output layer) was constructed with *ad hoc* software developed by the authors in C++ and R languages. The hidden layer of neurons allows the model to fit non-linear relationships between inputs and outputs, and can be thought to have a similar role to that of interactions terms in regression analysis. A training algorithm must be chosen to modify the weights in the network so that the network output is as close as possible to the desired output (target) when a patient data is presented to the network. The number of neurons in the hidden layer was determined using a constructive process, in which different architectures were considered with a number of neurons ranging from 5 to 30 neurons. Sixteen different combinations of covariables were analyzed in order to identify the best set of prognostic factors in terms of the accuracy in the prediction using a validation set. The variables age (A), tumour size (T), number of axillary lymph nodes (N) and grade of tumour (G), considered as very significant prognostic factors in clinical standard practice [19–21] were incorporated to every dataset. Starting from this data subset, menopausal status (M), histological type (H), estrogen receptors (ER) and type of treatment (Tr) were then included (or not) in the sixteen combinations, considered that were different in size due to the existence of missing data (Table 1).

The neural approach adopted in this work lies within those known as time-coded models, in which the time of

Table 1. Main characteristics of the dataset containing data from 2004 patients

Prognostic variables (mnemonic)	No. of patients	%
Age, years (A)		
Mean (Std)	56.21	
Range	25–90	
Tumour size (*T*)		
Mean (Std)	2.8	
Range	0.2–13	
# Axillary lymph nodes (N)		
Positive	1038	51.8
Negative	966	48.2
Histological grade (G)[a]		
1 – G(1)	390	19.4
2 – G(2)	1129	56.3
3 – G(reference)	485	24.3
Histological type (H)		
Ductal	1738	86.7
Lobular	131	6.5
Special	132	6.7
Missing	3	0.1
Hormonal receptor status (ER)		
Combined ER− and PR−	355	17.7
Combined ER+ and/or PR+	956	47.7
Missing	693	34.6
Menopausal state (M)		
Pre-menopausal	647	32.3
Post-menopausal	1319	65.8
Missing	38	1.9
Type of treatment (Tr)[b]		
No treatment – Tr(reference)	120	6.0
Radiotherapy – Tr(1)	103	5.1
Hormonetherapy – Tr(2)	469	23.4
Chemotherapy – Tr(3)	210	10.5
Hormone and radiotherapy – Tr(4)	406	20.2
Chemo and radiotherapy – Tr(5)	255	12.7
Chemo and hormonetherapy – Tr(6)	210	10.4
All three adjuvant therapies – Tr(7)	207	10.3
Intensive chemotherapy – Tr(8)	24	1.4

[a] Histological grade and between brackets the codification used by the Cox model for categorical variables.
[b] Type of treatment and between brackets the codification used by the Cox model for categorical variables.

follow-up is included as an additional covariate. Seven different non-overlapping intervals were studied (in months, 1–10, 11–20, 21–30, 31–40, 41–50, 51–60, 61–Maximum Follow-up Time (MFT)), and the output of the network represents directly the cumulative relapse probability for a given patient [8]. This approach has also the advantage that if needed a single NN can be used to obtain predictions for every time interval.

The NN architectures were trained by backpropagation with momentum with learning rate $\epsilon = 0.05$ and momentum constant equal to 0.35, combined with a weight decay procedure (to prevent overfitting) with parameter λ chosen within the range (1.0e-5, 7.5e-1).

Cox regression analysis

Cox regression analysis, a standard statistical tool in survival analysis, was used as a comparison for the performance of the NN approach. The relationships between different prognostic factors and patient survival, as well as the calculation of the prediction of the patient outcome, were assessed by applying Cox proportional hazards regression using the COXPH and PREDICT procedures in R [22,23] . Variable selection was done using backward and forward stepwise selection processes (the significance level of entry and permanence of a given variable in the model was $p < 0.05$). Tied event times were handled by the Breslow method [24] and the estimation of the survivor function at event times were performed using the BASEHAZ statement in R. The assumption of hazard proportionality for the model was tested using the ZPH procedure in R, which performs a test for a non-slope in a generalized linear regression of the scaled Schoenfeld residuals on functions of time [23]. A p-value < 0.05 in the test for zero-slope, or a non-constant value for the parameters β over time when the scaled Schoenfeld residuals are plotted, will indicate a violation of the proportional hazard assumptions.

Validation of the prognostic models

In order to estimate the classification accuracy for the NN models a standard technique of stratified ninefold cross-validation was used [25]. Firstly, data was divided into 10 subsets of approximately equal size, and one of them was reserved (as 'future' data) to test the prediction accuracy for every prognostic model. This 'Test set' was not used in any part of the training or validation procedure. Each of the left 9 random subsets of the data served as a validation set, to select an appropriate neural architecture. For the estimation of the Cox regression model all data except the test set was used. Then, the prediction accuracy for the models was tested over the 'future' data subset. Figure 1 shows a diagram of the procedure used for the splitting of the dataset in training, validation and testing sets used for training the neural architectures. To test the classification accuracy of the Cox regression model the same test set used for the NNs was used. For both models (Cox and NNs), a survival curve for each patient in the test set was generated, obtaining the cumulative probability of survival for every time interval. The predictive accuracy of different models was computed using the area under the ROC curve for censored data [26,27]. The area under the ROC curve (AUC) can be interpreted, given its equivalence to the c-index, as the proportion of all pairs of subjects whose survival time can be ordered such that the subject with higher predicted survival is the one who survived longer [28–31]. A ROC area of 0.5 is obtained for random predictions

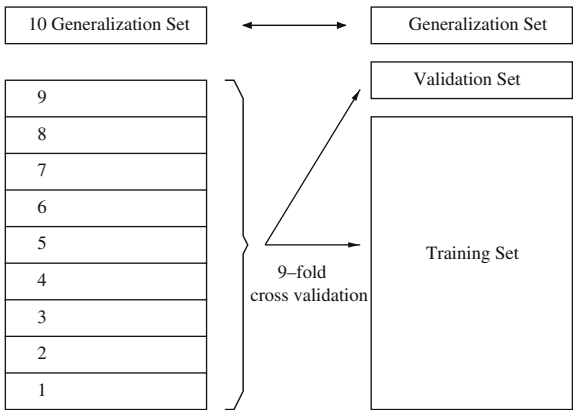

Figure 1. Scheme showing the way in which the training, validation and test sets were constructed and then used for training the neural architectures. The whole dataset was splitted in ten similar size boxes. The test set was initially randomly selected and not used for the architecture selection procedure that was implemented through a ninefold cross validation procedure.

while an area of 1 indicates a perfectly discriminating model.

Areas under the ROC curves (AUCs) of different models were compared by the Hanley–McNeil procedure [29–32]. It must be emphasized that differences in the AUCs merely reflect differences in the ability of the models to predict outcome in a given group of patients and not an actual difference in patient outcome. In other words, a high AUC means that outcome was predicted more accurately by a particular model, not that patient outcome was better than expected.

We computed the prediction accuracy of both NN and Cox regression models in two different ways. In both analyses the models were trained with all the available training data but were tested in different ways. First, the whole test set was used to measure the prediction accuracy independently of the interval, and second, the NN models with a different number of hidden neurons and the Cox models including interactions terms were tested with data only from the different time intervals considered. For the NNs, we used the ninefold cross-validation procedure described above, using the validation error as target for the architecture selection procedure. Different NN architectures with a single hidden layer with a number of neurons between 5 and 30 were tested. The best architecture was the one for which the validation error was the lowest, and the performance was then tested using the test set.

Hazard rate analysis

To compute the yearly hazard rate shown in Figure 3, SPSS mortality table analysis was used for patients with follow-up between 1 and 120 months in yearly intervals. To analyze the significance of the peaks the values of the hazard rate were analyzed in 6 months intervals and the significance of the peaks was contrasted against the mean value of the hazard rate (One-sample *t*-test).

Results

Using the artificial NN model, the best performance was obtained with five input variables that were age, tumour size, number of affected axillary lymph nodes, grade of tumour and type of treatment, when all data was considered together independently of the time interval. The area under the ROC curve (AUC) was 0.8497 (SE = 0.015, 95% CI 0.82–0.87) and a graph of the curve obtained is plotted in Figure 2, where also the curve obtained from the Cox model is presented.

Table 2 shows those variables that were statistically significant by applying the forward conditional stepwise selection procedure in Cox regression analysis. Hormonal receptors, menopausal status, and histological type (all with $p > 0.05$) were excluded from the full model and no interactions were found between any of

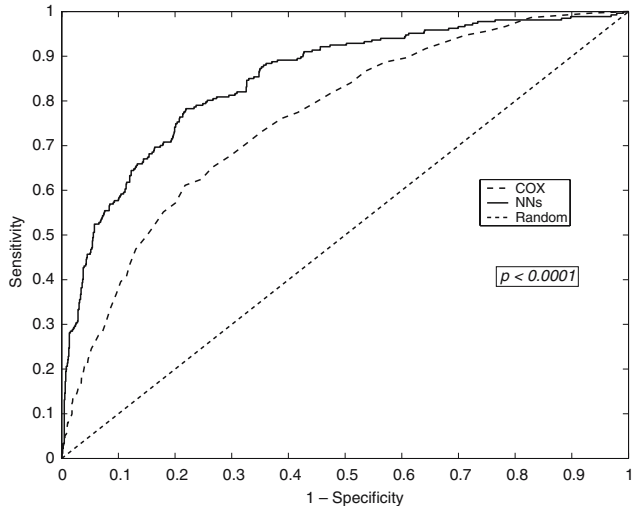

Figure 2. ROC curves obtained by using both neural networks and Cox models for the prediction of breast cancer relapse. The AUC was 0.8497 and 0.7602 for NNs and Cox model respectively and the difference was statistically significant at $p < 0.001$.

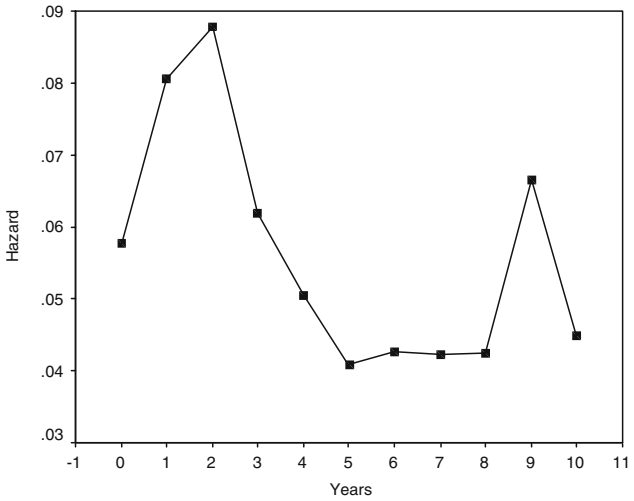

Figure 3. Yearly hazard rate function obtained from the dataset containing the records from 3811 breast cancer operated patients.

the other variables. The area of the ROC curve obtained with the Cox regression model with the same generalization dataset used for the NN was 0.7602 (SE = 0.0172, 95% CI 0.74–0.78). The difference between the two ROC areas (NNs and Cox regression) was assessed by the Hanley–McNeil procedure and was highly statistically significant (z-score = 3.8879, $p < 0.001$).

As mentioned before, we also constructed specialized neural architectures to compute predictions of survival for 7 different time intervals using neural architectures with a single hidden layer containing between 5 and 30 neurons in all cases. The results are shown in Table 3, where the time-intervals considered are indicated in the first column and the results for the Cox and NNs models are shown in columns 2 and 3 respectively (mean value ± standard error). For the case of the NN, the different prognostic factors used in each interval are shown in column 3. Finally, in the last column of Table 3 the z-scores (and related one tail p-values) of the difference between the predictions of NN and Cox regression models are included.

From these results, it is clear that the predictive accuracy of the NN models was significantly better than the one obtained by using Cox regression model (p-value < 0.0001). Five prognostic factors were found to be significant and common for both NN and Cox regression models (A, T, N, G, Tr). The model fitted by the Cox regression did verify the assumption of hazard proportionality. Despite its lower observed predictive power, one advantage of the Cox model it is the possibility of easier interpretation of the results. The categorical analysis carried out for the different type of treatments (Table 2) shows a significant effect (p-value < 0.05) in survival for patients treated with Chemotherapy (Tr 3) and with all three adjuvant therapies

Table 2. Set of clinical and pathological variables selected as potential prognostic factors by the Cox model

Variable	β	p-value	Odd ratio (exp β)	95% CI for exp β	
				Lower	Upper
Age	−0.010	0.017	0.990	0.986	0.994
Tumour size	0.132	< 0.00001	1.142	1.116	1.166
# Axillary lymph nodes	0.092	< 0.00001	1.097	1.088	1.104
Histological grade		< 0.00001			
G(1)	−0.749	< 0.00001	0.473	0.410	0.540
G(2)	−0.180	0.054	0.836	0.761	0.916
Type of Treatment		< 0.00001			
Tr(1)	0.110	0.677	1.117	0.856	1.455
Tr(2)	−0.148	0.472	0.863	0.702	1.058
Tr(3)	−0.565	0.019	0.569	0.446	0.724
Tr(4)	−0.132	0.528	0.876	0.711	1.080
Tr(5)	−0.245	0.267	0.783	0.627	0.976
Tr(6)	−0.277	0.218	0.758	0.605	0.949
Tr(7)	−0.578	0.012	0.561	0.445	0.706
Tr(8)	−2.471	0.000	0.084	0.048	0.146

For the categorical variables (Histological grade and Type of treatment) the results for the different variables according to the codification used in the Cox model (specified in Table 1) are given.

Table 3. Predictions of survival by time-intervals

Survival period (months)	Cox AUC[a] ± SE	NN AUC ± SE (Prognostic factors)[b]	Z-score[c]
1–10	0.8420 ± 0.072	0.9952 ± 0.0112 (A, T, N, G, ER, M, Tr)	2.1025 ($p < 0.05$)
11–20	0.7473 ± 0.0534	0.8908 ± 0.0451 (A, T, N, G, M)	2.0530 ($p < 0.05$)
21–30	0.7587 ± 0.0481	0.8051 ± 0.0470 (A, T, N, G, Tr)	0.6914 ($p = 0.24$)
31–40	0.7661 ± 0.0443	0.7884 ± 0.0459 (A, T, N, G, Tr)	0.3496 ($p = 0.36$)
41–50	0.7556 ± 0.0431	0.7663 ± 0.0446 (A, T, N, G, Tr)	0.1725 ($p = 0.43$)
51–60	0.7528 ± 0.0411	0.7661 ± 0.0411 (A, T, N, G, H)	0.2288 ($p = 0.40$)
61–MFT[d]	0.7120 ± 0.0407	0.8842 ± 0.0265 (A, T, N, G)	3.5456 ($p < 0.0005$)
1–MFT[d]	0.7602 ± 0.0172	0.8497 ± 0.0153	3.8879 ($p < 0.0001$)

[a] Prognostic factors selected by the Cox model were the same for all the intervals: A, T, N, G, Tr.
[b] Prognostic factors used for different intervals by the neural network model.
[c] Statistical differences between NN and Cox models AUCs.
[d] MFT: Maximum Follow-up Time.

(Tr 7). We note that as different treatments were applied to different groups of patients according to the clinical practice, the comparative interpretation of the benefit of the different treatments cannot be done with this data, and also the interpretation of the odd ratios has to be done very carefully.

On the other hand, the time dependent analysis shows (Table 3) that the NN outperforms significantly the Cox model mainly for the early time intervals between 1–10 and 11–20 months. For the periods between 21–30, 31–40, 41–50, and 51–60 months the difference between NN and the Cox model was not so large and in fact not statistically significant. Despite that in all intervals the NN prediction is more accurate, the lack of statistical significance might be, within others factors, a consequence of a reduced number of cases included in each of these intervals considered. In the last time interval considered (61–MFT) the NNs also outperforms the Cox model (difference between the AUC = 0.1722; p-value < 0.0005). Regarding the prognostic factors considered optimal for the prediction of the relapse using the neural architectures, it is worthwhile mention the inclusion of menopausal status (M) for the early intervals 1–10 and 1–20 months and that in all the cases the optimal NNs did not use H as a prognosis covariate.

We also analyzed the yearly hazard rate function to find that it shows two clear peaks of high risk mortality, one at 2 years after surgery and a second one at 9 years. The results of the hazard analysis are shown in Figure 3. To analyze the significance of the observed peaks, we computed the hazard rate in intervals of 6 months and perform a one sample t-test to check whether the values at the peak were significantly different that the mean of the Hazard values from all other times. For the first peak appearing around the second year in Figure 3, the values at 6, 12, 18 and 24 months were considered obtaining that all four values were significantly different from the mean (p-value < 0.05, df = 18). For the second peak shown in Figure 3 at 9 years, only the value at 108 months was statistically significantly different than the mean ($p < 0.001$), but the real existence of this peak is less clear as the analysis at this stage can be affected by its proximity to the end of the follow-up time.

Discussion

Prognostic factors in breast cancer provide information to patients about the recurrence likelihood of the disease and, more important, assist the clinicians in the selection of appropriate adjuvant treatments for individual patients. From a biological point of view, the identification of good prognostic factors supplies information about the natural history of the disease. Research on prognostic factors in operable breast cancer has focused in the identification of new factors, mainly derived from lately advances in the field of cancer molecular biology. Nevertheless, not much work has been done about the methodology addressing the identification of such prognostic factors. In this sense, Cox multivariate analysis has been accepted as the gold-standard in methods of prognostic factors identification. However, Cox multivariate analysis involves some assumptions to be made (e.g., the relative risk between the hazard rate for two subjects are constant over time), that are difficult to check in practice. In this study, using a classical set of prognostic factors an approach based on artificial NNs provides better survival predictions than those obtained by applying the Cox multivariate analysis. This better prognosis accuracy is specially relevant in the first time-interval 1–10 months, in which an AUC value of 0.99 was obtained, and also in the second interval under study 11–20 months in which the AUC was found to be 0.89.

The fact that the improvement in prognosis accuracy of NNs is statistically significant in regard to the Cox multivariate analysis has also implications from a biological point of view. Namely, the relationships among different prognostic factors and the patient outcome may well be non-linear in nature. If this is the case, traditional statistical methods that assume linear relationships among variables could be inadequate to investigate the predictive capability of different biological parameters (clinical, histological o molecular parameters). In other words, methods that assume linear relationships among variables cannot identify non-linear in biological relationships.

It is well known that patients diagnosed with breast cancer have a risk for recurrence for an extended period of time. However, Kaplan–Meier survival curves offer little information about changes in the probability for disease relapse over time. If we could define the way relapses are distributed in time, we could obtain information about the behaviour of micrometastatic foci, which could contribute to the development of new treatment strategies.

The different risk recurrence analysis performed by Demicheli et al. [33] by using the data of more than 1100 patients treated at the Milan Cancer Institute, led him to the conclusion that breast cancer has two peaks of higher recurrence risk: one early peak of higher incidence at 18 months after surgery and a second peak around 60 months. Other studies have only been able to define one peak of higher relapse incidence in the number of relapses. The analysis of the annual relapse risk performed by Saphner et al. [34] in a population of more than 3500 patients included in different ECOG studies, showed only one peak of higher relapse risk at 18 months after surgery. The results obtained by Karrison et al. [35] from the analysis of almost 1600 patients, pointed out a unique peak of higher risk. Furthermore, Retsky has reviewed the data from nine sets of biological and clinical data, finding the same pattern of relapse [36]. On the other hand, Demicheli et al. [37] performed a comparative study of the mortality pattern in his cohort of patients treated only with surgery, compared with the historic control group from Middle-

sex [38], which included non-treated breast cancer patients. While the Milan Cancer Institute series showed a mortality pattern with two peaks of higher mortality rate (the first peak between the 3rd and 4th year after surgery, and the second peak in the 8th year), in the Middlesex series, only one peak of increased mortality (between the 4th and the 5th year after surgery) was observed. Despite the evident methodological differences in data gathering from the two series, the authors concluded that the comparison of the two curves gives some support to the hypotheses that in breast cancer, surgical removal of the primary tumour may induce changes in the growth kinetic of metastatic foci.

For the dataset considered in the present analysis the yearly hazard rate function shows two clear peaks of high risk mortality, one at 2 years after surgery and a second one at 9 years. The peaks are similar to those found in the comparative study by Demicheli et al. [39] but with the main difference that in our study the second peak is located a little bit later, occurring at 9 years after surgery.

For the first time, and according to our results, it seems that artificial NNs are especially useful in order to predict the first peak of breast cancer relapse that some authors related to surgery. Furthermore, our study confirmed the results obtained by other authors, as mentioned above, by means of a completely different methodological procedure.

Acknowledgements

Support from APOMA (Málaga, Spain) under contract No. 8.06/47/2176 is acknowledged. The authors also thank Esther Mahillo (GEICAM) for her collaboration and coordination of this project.

References

1. Hilsenbeck SG, Ravdin PM, de Moor CA: Time-dependence of hazard ratios for prognostic factors in primary breast cancer. Breast Cancer Res Treat 52: 227–237, 1998
2. Cox DR: Regression models and life tables. J R Stat Soc 34: 187–202, 1972
3. Ohno-Machado : A comparison of Cox proportional hazards and artificial neural network models for medical prognosis. Comput Biol Med 27: 55–65, 1997
4. Xiang A, Lapuerta P, Ryutov A, Buckley J, Azen S: Comparison of the performance of neural network methods and Cox regression for censored survival data. Comput Stat and Data Anal 34: 243–257, 2000
5. Burke H, Goodman P, Rosen D, Henson D, Weinstein J, Harrel F, Marks J, Winchester D, Bostwick D: Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 79: 857–862, 1997
6. Brown SF, Branford AJ, Moran W: On the use of artificial neural networks for the analysis of survival data. IEEE Trans Neural Networks 8: 1071–1077, 1997
7. Faraggi D, Simon R, Yaskil E, Kramar A: Bayesian neural network models for censored data. Biometrica J 5: 519–532, 1997
8. Ripley RM, Harris AL, Tarassenko L: Neural network models for breast cancer prognosis. Neural Comput Appl 7: 367–375, 1998
9. Lundin M, Lundin J, Burke HB, Toikkanen S, Pylkkänen Joensuu H: Artificial neural networks applied to survival prediction in breast cancer. Oncology 57: 281–286, 1999
10. Jerez JM, Gómez JA, Ramos G, Muñoz J, Alba E: A combined neural network and decision trees model for prognosis of breast cancer relapse. Artif Intell Med 27: 45–63, 2003
11. Ravdin PM, Clark GM, Hilsenbeck G, Owens MA, Vendely P, Pandian MR, McGuire WL: A demonstration that breast cancer recurrence can be predicted by neural network analysis. Breast Cancer Res Treat 21: 47–53, 1992
12. De Laurentis M, Ravdin PM: A technique for using neural network analysis to perform survival analysis of censored data. Cancer Lett 77: 127–138, 1994
13. De Laurentiis M, De Placido S, Bianco AR, Clark GM, Ravdin PM: A prognostic model that makes quantitative estimates of probability of relapse for breast cancer patients. Clin Cancer Res 5: 4133–4139, 1999
14. Boracchi P, Biganzoli E, Marubini E: Modelling cause-specific hazards with radial basis function artificial neural networks: application to 2233 breast cancer patients. Stat Med 20: 3677–3694, 2001
15. Liestol K, Andersen PK: Updating of covariates and choice of time origin in survival analysis: problems with vaguely defined disease states. Stat Med 21: 3701–3714, 2002
16. Biganzoli E, Boracchi P, Coradini D, Daidone MG, Marubini E: Prognosis in node-negative primary breast cancer: a neural network analysis of risk profiles using routinely assessed factors. Ann Oncol 14: 1484–1493, 2003
17. Biganzoli E, Boracchi P, Mariani L, Marubini E: Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Stat Med 17: 1169–1186, 1998
18. Dreiseitl S, Ohno-Machado L: Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inf 35: 352–359, 2002
19. Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, Ruby SG, O'Malley F, Simpson JF, Connolly JL, Hayes DF, Edge SB, Lichter A, Schnitt SJ: Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 124: 966, 2000
20. Kroman N, Jensen MB, Wohlfahrt J, Mouridsen HT, Andersen PK, Melbye M: Factors influencing the effect of age on prognosis in breast cancer: population based study. BMJ 320: 474–478, 2000
21. Colleoni M, Rotmensz N, Robertson C, Orlando L, Viale G, Renne G, Luini A, Veronesi P, Intra M, Orecchia R, Catalano G, Galimberti V, Nole F, Martinelli G, Goldhirsch A: Very young women (< 35 years) with operable breast cancer: features of disease at presentation. Ann Oncol 13: 273–279, 2002
22. Therneau TM, Grambsch PM: Modelling Survival Data: Extending Survival Data. Springer-Verlag, New York, NY, 2000
23. Fox : An R and S-Plus companion to applied regression. Sage Publications, Thousand Oaks, CA, 2002
24. Breslow NE: Covariance analysis of censored survival data. Biometrics 30: 89–99, 1974
25. Shao J, Tu D: The Jacknife and Bootstrap. Springer Verlag, New York, NY, 1995
26. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA: Evaluating the yield of medical tests. J Am Med Assoc 247: 2543–2546, 1982
27. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA: Regression modeling strategies for improved prognostic prediction. Stat Med 3: 143–152, 1984
28. Hanley J, Mc Neil B: The meaning and use of the area under the receiver operating characteristic (ROC) curve. Radiology 143: 29–36, 1982
29. Hanley JA, McNeil BJ: A method for comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 148: 839–843, 1983

30. Clark TG, Bradburn MJ, Love SB, Altman DG: Survival Analysis Part IV: Further concepts and methods in survival analysis. Br J Cancer 89: 781–786, 2003
31. Dreiseitl S, Ohno-Machado L, Harald Kittler H, Vinterbo S, Billhardt H, Binder M: A comparison of machine learning methods for the diagnosis of pigmented skin lesions. J Biomed Inform 34: 28–36, 2001
32. Greiner M, Pfeiffer D, Smith RD: Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Prev Vet Med 45: 23–41, 2000
33. Demicheli R, Abbattista A, Miceli R, Valagussa P, Bonadonna G: Time distribution of the recurrence risk for breast cancer patients undergoing mastectomy: further support about the concept of tumor dormancy. Breast Cancer Res Treat 41: 177–185, 1996
34. Saphner T, Tormey DC, Gray R: Annual hazard rates of recurrence for breast cancer after primary therapy. J Clin Oncol 14: 2738–2746, 1996
35. Karrison TG, Ferguson DJ, Meier P: Dormancy of mammary carcinoma after mastectomy. J Natl Cancer Inst 91: 80–85, 1999
36. Retsky MW, Wardwell RH, Swartzendruber DE, Headley DL: Prospective computerized simulation of breast cancer: comparison of computer predictions with nine sets of biological and clinical data. Cancer Res 47: 4982–4987, 1987
37. Demicheli R, Valagussa P, Bonadonna G: Does surgery modify growth kinetics of breast cancer micrometastases?. Br J Cancer 85: 490–492, 2001
38. Bloom HJG, Richardson WW, Harries EJ: Natural history of untreated breast cancer (1905–1933). BMJ 2: 213–221, 1962
39. Demicheli R, Miceli R, Valagussa P, Bonadonna G: Re: dormancy of mammary carcinoma after mastectomy. J Natl Cancer Inst 92: 347–348, 2000

Address for offprints and correspondence: Leonardo Franco, Depto. de Lenguajes y Ciencias de la Computación, Universidad de Málaga, Campus de Teatinos S/N, 29071, Málaga, Spain; *Tel.:* + 34-952-133304; *Fax:* + 34-952-133397; *E-mail:* Leonardo.Franco@psy.ox.ac.uk