

INTELLIGENT INTERFACE FOR ADJUVANT TREATMENT PLANNING IN BREAST CANCER

I.H. Jarman, T.A. Etchells and P.J.G. Lisboa
Liverpool John Moores University, UK

Abstract

The primary course of therapy for breast cancer patients, following surgery, depends on the expected prognosis together with the key clinical indicators. An interface for use by clinical oncologists is proposed, which addresses three fundamental questions, namely; evidence that the currently used Nottingham Prognostic Index can be enhanced by additional clinical features, prognostic inference for individual patients with quantified confidence levels, and visualisation of the patient database by clinical indicator of adjuvant treatment. This interface is underpinned by detailed prognostic analysis validated through longitudinal cohort studies of mortality with 931 TNM stage I/II patients recruited between 1990 and 1993 at Christie Hospital, Wilmslow. The data shown in the interface are Kaplan-Meier curves from prognostic risk groups inferred by cross validation.

1. Introduction

The starting point for this paper is the commonly used clinical prognostic index for breast cancer, the Nottingham Prognostic Index (NPI) [1]. While widely applied to inform the choice of adjuvant therapy, advances in therapy, detection technologies and health policy, such as the introduction of breast cancer screening for women aged 50 and over in the UK, has skewed the patient population and has added potential prognostic indicators. This study proposes an interface for clinical oncologists to show the added value of expanding the covariate basis for prognostic inference. It is important to note that the basis of our approach is to keep NPI and expand rather than replace current practice.

Furthermore, there is now an interest in predictive inference of prognosis for individual patients, witness the web-based prognostic interface www.Adjuvantonline.com [2]. This model is gaining clinical support in part because it infers the potential effect of different treatment choices. It also points towards a visualisation format that appears to be readily accepted by practicing clinicians. However, the predictions made do not include confidence estimates, yet are likely subject to substantial uncertainties for particular groups of patients, notably in NPI group 3, which is known to be

heterogeneous in its composition. Moreover, it is not clear that the development of this interface has followed the recommended staged process of the continuum of evidence, which is modelled on the development of medicinal drugs and is intended to assure the accuracy and generalisability of clinical inferences [3-4].

In addition to prognostic inference, a previous study of the prevalence of adjuvant treatment, typically hormone therapy e.g. tamoxifen, chemotherapy, or both in combination, showed that the different treatments are clustered primarily by key clinical important indicators of the likely response to treatment, namely oestrogen receptor count, lymph nodes affected and menopausal status [5]. The proposed interface switches between survival modelling and treatment allocation profiles.

In summary, there are two aspects of novelty presented, firstly a methodology for an individual prediction of survival with confidence intervals using neural networks and Monte Carlo methods. Secondly, implementing an interface that shows the added value of a new prognostic model over the current clinical standard prognostic model supported by a personal prognosis and data-based rules for treatment allocation.

The next section explains what the NPI is and how it was extended by modelling with additional covariates using Cox regression with the proportionality of hazards' assumption. This leads to the derivation of a cross-matching framework to discriminate between the survival of patients in each NPI risk group. Section 3 summarises the derivation of prognostic models with confidence intervals for individual patients, using Monte Carlo methods. This is followed, in section 4, by a brief overview of the rule extraction algorithm used to explain treatment allocation. Finally, the interface is described in its entirety.

2. Extended prognostic indices of survival

Survival analysis is an important field in medical statistics where the proportional hazard model [6], also known as Cox regression, is the most widely used method.

The form of the Cox regression model is:

$$h_0(t) \exp \left[\sum_{i=1}^n b_i x_i \right] \quad (1)$$

where $h_o(t)$ is the baseline hazard function. $h_o(t)$ is called the baseline function because when all the x variables are equal to zero the formula reduces to this form, hence the 'baseline' of the model. β are the coefficients of x , which are the explanatory variables.

Cox regression is a semi-parametric model that incorporates censored data, which arises when an individual drops out of a study for reasons other than the event of interest, death due to breast cancer in this study. Omitting these data from a survival analysis can introduce significant bias to any results [7].

This forms the basis for a prognostic index that is clinically widely accepted, the Nottingham Prognostic Index (NPI), which uses 3 variables identified as being significant in the prediction of survival, namely; *pathological size of tumour, histological grade of tumour and the number of axillary nodes affected* and requires a calculation in the form of a simple equation, which for a clinician makes it easy to use and understand. In the case of NPI:

$$0.2 \times \text{pathological size} + \text{histological grade} + \text{nodes involved} \quad (2)$$

From this index, using the log-rank statistic, patients are allocated into 4 prognostic risk groups, ranging from very good to poor, at cut-off points < 2.41 , < 3.41 , < 5.41 and ≥ 5.41 respectively.

A further Cox regression model using six variables; *age, clinical stage nodes, histology, node ratio, pathological size and ER status* has been developed from 917 patients and validated on 931 patients from Christie Hospital near Manchester referred between 1983 – 89 and 1990-93 respectively. The latter dataset, the validation group, showed that the NPI and the new Cox model separated the patient profiles into prognostic groups with similar mean survival but with different risk group allocation, where NPI could be calculated (559 patients).

By cross-matching the two prognostic indices we are able to examine survival for patient groups within each matrix cell using Kaplan-Meier (KM) estimated survival curves in figure 1, in order to discover heterogeneity in estimated survival for any of the models prognostic groups. These differences in survival are an indication of the added value of cross-matching NPI with another survival model that uses additional variables and are providing supplementary information for prognosis.

This same idea of cross-tabulation can be extended to a scatter-plot of the prognostic indices, which allows the patient to be identified within this framework figure 2 and therefore identify how borderline a patient may be to adjacent prognostic groups or cells.

3. Individual Prognostic Predictions with Confidence Intervals

In addition to the detailed analysis of the group in which a particular patient belongs, there is interest in predictive inference of prognosis for that individual patient. The website <http://www.adjuvantonline.com/> presents such

information but without confidence intervals, so the uncertainties inherent in the prediction cannot be assessed. We present a method, using hazard predictions from a Partial Logistic Artificial Neural Network with Automatic Relevance Determination (PLANN-ARD) [8] and Monte Carlo methods, that give prognostic predictions with confidence intervals for individual patients.

The PLANN-ARD model provides a prediction of smooth estimates of the discrete time hazard. It is implemented as a direct extension of the Multi-Layer Perceptron (MLP) neural network applied as a discrete model of the hazard function. Using this MLP structure with time as an input we have

$$\frac{h_p(x_p, t_k)}{1 - h_p(x_p, t_k)} = \exp \left(\sum_{h=1}^{N_h} w_h g \left(\sum_{i=1}^{N_i} w_{ih} x_{pi} + w t_k + b_h \right) + b \right) \quad (3)$$

Estimating the weights requires a likelihood term for the status of one patient at time t_k , by using an indicator label 0 if a patient is alive at time t_k and a label 1 for the event of interest. This generic non-linear model is called the Partial Likelihood Artificial Neural Network (PLANN) [9]. In contrast to a proportional hazards model [6], PLANN does not require proportionality of the hazards over time and predicts a smooth hazard function.

At time t_i the estimated summed weights to each output unit has a Gaussian distribution $N(a_i, \mathbf{S}_i^2)$ [10]. The estimated hazard is calculated by the sigmoidal activation:

$$h(t_i) = g(a_i) = \frac{1}{1 + \exp(-a)} \quad (4)$$

Once the network weights are estimated, the survivorship is calculated from the estimated discrete time hazard by multiplying the conditionals for survival over successive time intervals treated as independent events, this gives:

$$S(t_k) = \prod_{l=1}^k (1 - h(t_l)) \quad (5)$$

Estimating an individual prognosis for patient x we use Monte Carlo methods by taking a random sample \tilde{a}_i from $N(a_i, \mathbf{S}_i^2)$, calculate $\tilde{h}_i = g(\tilde{a}_i)$ and finally estimate survival $\tilde{S}(t_k)$. Repeat these steps n times, enough to build up a distribution of survival estimates, as shown in figure 3. The personalised prognosis is the mean survival of the distribution with 95% confidence intervals determined by omitting the upper and lower 2.5% of the sample estimates.

The survival estimate can be presented as a simple colour coded green, amber and red bar representing probabilities of survival, with 95% confidence interval

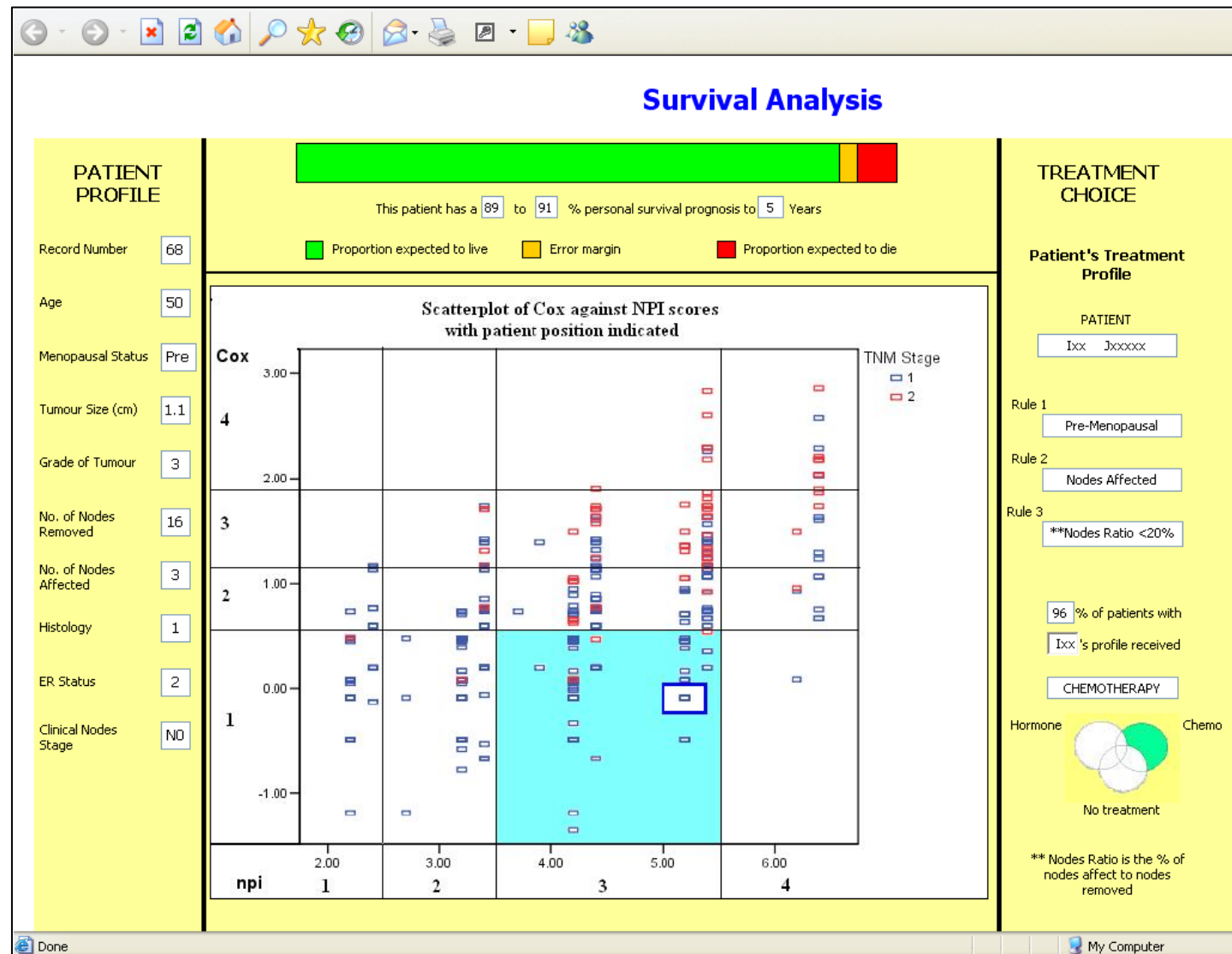


Figure 2. Proposed intelligent interface for treatment. To the left is the patient profile containing the significant variables that affect both survival and treatment choice. The centre graphic represents a scatterplot of the Cox and NPI scores, the patient position and cell within the cross-tabulation matrix is highlighted. To the right there are empirical rules describing a similar group of patients stating the proportion that received the recommended course of treatment

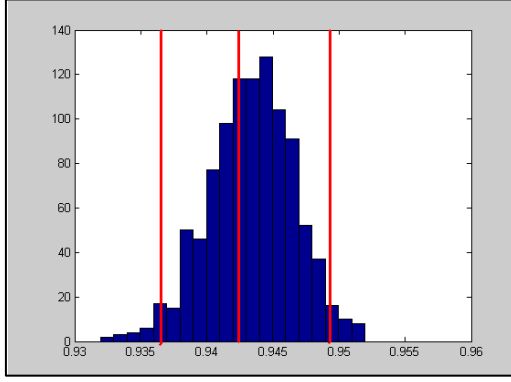


Figure 3. Distribution from 1000 iterations of estimated survival for an individual patient with mean survival and 95% confidence intervals indicated in red

and death respectively to a particular time period of interest, 5 years in this study figure 1.

4. Data-based Rules to Describe the Patient Allocations Made by the Analytical Risk Scores

In this paper the Orthogonal Search Rule Extraction (OSRE) algorithm [11] is used to extract rules for the treatment of the 559 patients in this study.

The OSRE algorithm finds conjunctive rules for classifications of data using a Multi-layer Perceptron (MLP), or any other smooth response surface, that has been trained to accurately predict the classifications of a dataset. A detailed account of the algorithm and the mathematical framework that underpins it can be found in chapters 3 & 4 of [12].

In essence OSRE searches for changes in response from an MLP, starting from each data point in turn in the data set and systematically searching in orthogonal directions. To demonstrate the algorithm we take a data set that has three variables and each variable has values ranging from 1 to 6. Figure 4 shows the data-space and a surface boundary that separates the in and out of class data. The arrows show the directions in which the algorithm searches for changes in the response of the surface. Notice that in the direction of the variable a_1 , there is no change in response from the surface.

The consequence of there being no change in response of the surface for a particular variable is that the variable does not feature in the set of conjunctive rules for this surface. Figure 5 shows the ‘hyper-box’ that the algorithm generates for the data-point represented in figure 4.

The rule generated from the ‘hyper-box’ is

$$(a_1 \leq 6) \text{ AND } (a_2 \leq 4) \text{ AND } (a_3 \geq 3)$$

or more simply, as a_1 takes all possible values,

$$(a_2 \leq 4) \text{ AND } (a_3 \geq 3).$$

This process is repeated for each data-point for which the surface predicts it to be in-class. A set of rules the size of the number of data predicted in-class is generated. The algorithm is enhanced with a refining method to reduce the number of explanatory rules conditional on maintaining sensitivity and specificity values above minimal acceptable thresholds [5].

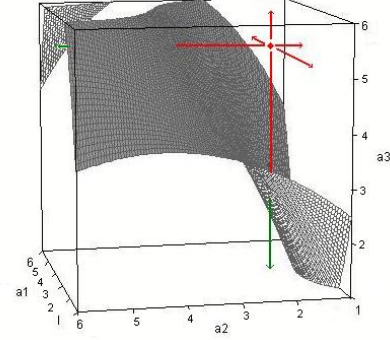


Figure 4. OSRE searching for changes in response in orthogonal directions.

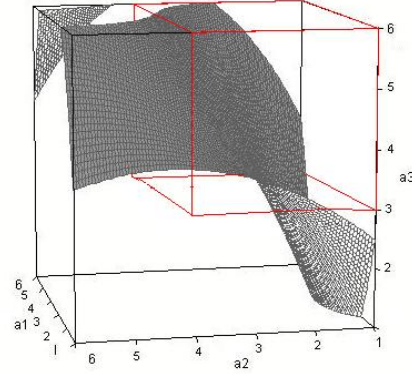


Figure 5. The OSRE algorithm generates a ‘hyper-box’ from which conjunctive rules are found.

5. Integrated Intelligent Interface for Breast Oncology

Combining all the elements described above enables us to present an integrated intelligent interface for clinicians. This is achieved with the cross-matching matrix where each column represents patients in prognostic risk groups for the current standard NPI model, the rows representing the risk groups for the new prognostic prediction. This can inform the clinician on a patient’s survival outcome (figure 1) as well as giving NPI survival estimates with extended survival predictions for sub-groups within the cross-tabulation matrix. This allows the clinician to assess heterogeneity in survival within a prognostic risk group. Presenting a new model as an extension of NPI enables clinicians to relate to their own reasoning model, where the use of the current prognostic group allocation assists as an indicator for choice of therapy [13]. In addition the bar graphic above the cross-matching matrix presents the individual prognosis with 95% confidence intervals

derived by using an Artificial Neural Network with Monte Carlo methods as described in section 3.

By replacing survival estimates with a scatter-plot of prognostic scores for each prognostic model (figure 2), we can examine whether a particular patient is borderline between cells in the matrix as the cross-tabulation matrix is placed over the cut-off points for prognostic group scores. In addition, a patient's TNM stage (a commonly used prognostic model) is highlighted by colour coding the data points, this shows the wide scatter of the TNM stage across the map, thus providing another level of insight to the clinician.

This information is also supported by empirically derived rules using the rule extraction method, OSRE, described in section 4, this informs the clinician about the treatment given to similar patient groups and presents the rules derived from the data for the treatment received by this group.

With all elements combined an intelligent interface is presented to the clinician, by expanding NPI into a matrix, maintaining their current knowledge of survival expectation and treatment allocation for patient groups while showing the difference additional information has on sub-groups of patients survival prognosis. It also informs on patient cases that may be borderline between prognostic groups. Additionally, it provides an individual prognosis of survival to 5 years with 95% confidence intervals and presents a Boolean expression of group characteristics for treatment derived from evidence in historical data.

6. Conclusion

An interface for breast oncology is proposed, which shows the value of additional covariates in discriminating patients by mortality risk. The interface starts from a currently used clinical index, NPI, and extends this to include a cross-matching matrix of grouped survival curves and the position a patient resides within the matrix, complemented with individual prognostic predictions qualified with predicted confidence intervals, additionally treatment allocation is explained by data-based rules. These data in combination add significantly to the discriminatory information currently available to clinicians about prognostic risk and allocation of adjuvant treatment.

This complex information is presented in a format designed to match the clinician's own reasoning. Further work is now required to evaluate the clinical acceptance of the proposed methodology.

7. Acknowledgement

This work was supported by EPSRC research grant and the European Network of Excellence BioPattern.

8. References

1. Haybittle JL, Blamey RW, Elston CW, Johnson J, Doyle PJ, Campbell FC, Nicholson RI, Griffiths K, 'A prognostic index in primary breast cancer', *Br. J. Cancer*, 45: 3621, 1982.
2. Ravdin P.M., Siminoff L.A., Davis G.J. et al 'Computer Program to Assist in Making Decisions about Adjuvant Therapy for Women with Early Breast Cancer'. *Journal of Clinical Oncology*, 74 (4) 980-991; 2001.
3. Lisboa, P.J.G. 'A review of evidence of health benefit from artificial neural networks in medical intervention', *Neural Networks, Invited Paper*, 15(1), 9-37, 2002.
4. Campbell, M., Fitzpatrick, R., Haines, A., Kinmonth, A.L., Sandercock, P., Spiegelhalter, D. and Tryer, P. 'Framework for design and evaluation of complex interventions to improve health'. *BMJ*, 321: 694-696, 2000. www.mrc.ac.uk/complex_packages.html
5. Etchells, T.A., Jarman, I.H., Lisboa, P.J.G., 'Empirically derived rules for adjuvant chemotherapy in breast cancer treatment' *MEDSIP, Proc. 2nd International Conference*, 345-351, 2004,
6. Cox DR, 'Regression models and life tables', *Journal of the Royal Statistical Society, B*, 74: 187-220, 1972.
7. Collett D. 'Modelling Survival Data in Medical Research'. Chapman and Hall, 1994.
8. Lisboa, P.J.G., Wong, H., Harris, P. and Swindell, R. 'A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer'. *Artificial Intelligence in Medicine*. 28(1):1-25, May 2003.
9. Biganzoli E, Boracchi P, Mariani L, Marubini E, 'Feed forward neural networks for the analysis of censored survival data: A partial logistic regression approach', *Statist. Med*, 17, 1169-1186, 1998.
10. Bishop C.M. 'Neural network for pattern recognition'. Oxford; 1995: Clarendon Press.
11. Etchells, T.A. & Lisboa P.J.G. 'On rule extraction from smooth decision surfaces'. *NWSMED/CIMED, Proc. 5th International Conference*, pp 23-28, 2003.
12. Etchells, T.A., Rule Extraction from Neural Networks: A practical and efficient approach. Unpublished PhD thesis, 2003. <http://www.cms.livjm.ac.uk/etchells/phd/etchells.pdf>
13. <http://www.nice.org.uk>