THE PARTIAL LOGISTIC ARTIFICIAL NEURAL NETWORK (PLANN): A TOOL FOR THE FLEXIBLE MODELLING OF CENSORED SURVIVAL DATA

E. M. Biganzoli
Milan Cancer Institute, Italy

P. Boracchi
University of Milan, Italy

ABSTRACT

Linear and non-linear flexible regression analysis techniques, such as those based on splines and feed forward artificial neural networks (FFANN), have been proposed for the statistical analysis of censored survival time data, to account for the presence of non linear effects of predictors. Among survival functions, the hazard has a biological interest for the study of the disease dynamics, moreover it allows for the estimation of cumulative incidence functions for predicting outcome probabilities over follow-up. Therefore, specific error functions and data representation have been introduced for FFANN extensions of generalized linear models, in the perspective of modelling the hazard function of censored survival data. These techniques can be applied to account for the prognostic contribution of new biomarkers in addition to the traditional ones.

Keywords: Survival Analysis, Generalised Linear Models, Artificial Neural Networks.

1. INTRODUCTION

In biomedicine, modelling the time to occurrence of a specific event (failure time) is relevant to build decision support systems for the prediction of patients outcome with the aim of treatment planning. As an example, in clinical oncology, the study of prognostic factors deserves a growing interest to identify groups of patients with different risk of unfavourable events (death, distant metastases, local recurrences, etc.). Due to the growing number of putative prognostic factors to be investigated and to the presence of possibly complex prognostic relationships (non linear and non additive effects), feed forward artificial neural networks (FFANN) have been advocated for outcome analysis in oncology, Baxt (2). In such a context, multilayer perceptrons (MLP) and radial basis functions (RBF) networks are flexible alternative to traditional statistical tools. The benefits of the integration of FFANN with traditional methods for outcome prediction are still to be exploited; however, some reviews, Ripley and Ripley (25), Schwarzer et al. (26), Lisboa (20), have pointed out advantages and possible limitations caused by the adoption of heuristic approaches, without a full account of the specific features of censored failure time data.
Papers merging survival analysis theory with artificial neural networks methodology in a discrete multiple classification framework have been firstly published by Liestol et al. (19), whereas a FFANN extension of the Cox's model has been proposed by Faraggi and Simon (15) and applied by Mariani et al. (22) after some modifications.
Among the different survival distributions, the hazard function (i.e. the instantaneous relative failure rate) has a key role in investigating the disease dynamics. Estimation of the hazard as a conditional function of time and covariates is a difficult problem, possibly characterized by non-monotonic behaviours and high order interactions between time and covariates; for this problem, FFANNs could provide substantial advantages with respect to linear methodologies. Starting from the relationships between generalized linear models (GLM) with Poisson or binomial errors and piecewise parametric or grouped time models, respectively, see Aitkin et al., (1), Efron, (13), their extension with MLPs and RBFs has been recently proposed for flexible modelling of the hazard function, allowing for non-linear and non-proportional effects of covariates, Biganzoli et al., (3), (4); Boracchi and Biganzoli, (11).
The aim of this paper is to present a general framework to define FFANN models on survival data, based on the relationship between GLM and piecewise parametric and grouped time models. To this aim the peculiar features of survival data are presented and the extension of GLM with MLPs and RBFs for survival data processing is subsequently illustrated. An application example is discussed and final considerations are provided. To ensure conformity and uniformity of appearance it is essential that these instructions are followed. Sample pages are enclosed with these notes to show you how the finished paper should look.

2. FAILURE TIME DATA

Let Z the random variable (r.v.) time elapsed by the beginning of the observation (e.g. date of the surgical intervention) to the appearance of a specific event (e.g. death or relapse of the disease). The followings functions of Z can be defined

- *survival*

$$S(z) = P(Z > z) = 1 - \int_0^z f(u)du = \int_z^\infty f(u)du$$

where $f(u)$ is the *probability density function.*

- *hazard*

$$h(z) = \lim_{\Delta z \to 0^+} \frac{P(z < Z \le z + \Delta z \mid Z > z)}{\Delta z}$$

then

$$h(z)=\frac{f(z)}{S(z)}$$

A feature of failure time (survival) data is the possible incomplete observation. The time of the considered event may be unknown for the (i=1,2,…,N) experimental unit. This situation is known as censorship, and it may happen because of time limits in follow-up or other possible restrictions that depend from the nature of the study. Different types of censorship are possible and they have to be considered in appropriate way in the statistic models, Klein and Moeschberger, (16). In the present note it will only be considered the right censoring which verifies if the event is not observed before the term of the study or of a competitive event that causes the interruption of the individual sequence of visits (follow-up). For the ith subject, who will be characterized by the vector of covariates x_i, the existence of a time of event (Z_i) and one of censors (C_i) are assumed; then the time observed t_i is the realization of the r.v. T_i=min(Z_i, C_i). Right censored data are represented by a couple of r.v. (T_i, δ_i); where δ_i has actual value d_i=1 if the event is observed ($T_i=Z_i$) and d_i=0 if ($T_i=C_i$).

The general expression of the likelihood function for survival data in presence of right censoring, conditionally to x_i is given by

$$L=\prod_{i=1}^{N} f(t_i,x_i)^{d_i}\cdot S(t_i,x_i)^{1-d_i}=\prod_{i=1}^{N} h(t_i,x_i)^{d_i}\cdot S(t_i,x_i)$$

3. PARTITION ANG ROUPING OF FAILURE TIMES

Given the continuous random variable T, piecewise models derive from the partition of the time axis into l=1,2,…,L disjoint intervals $A_l=(\tau_{l-1},\tau_l]$; with τ_0=0. For the ith subject and the lth interval, the density probability functions $f_l(t,x_i)$, the survival function

$$S_l(t,x_i)=\int_t^{\infty} f_l(u,x_i)du$$

and the hazard function

$$h_l(t,x_i)=\frac{f_l(t,x_i)}{S_l(t,x_i)}$$

are defined. Since the density functions are allowed to change in the different intervals, the *piecewise survival* function on l intervals is

$$S(t)=\left[\prod_{\nu=1}^{l-1}\frac{S_\nu(\tau_\nu,x_i)}{S_\nu(\tau_{\nu-1},x_i)}\right]\cdot\frac{S_l(t,x_i)}{S_l(\tau_{l-1},x_i)} \qquad (1)$$

where $\frac{S_\nu(\tau_\nu,x_i)}{S_\nu(\tau_{\nu-1},x_i)}$ is the conditional probability of surviving till the time τ_ν, given no event at time $\tau_{\nu-1}$. For simplicity in the (1) and in the following calculations concerning products on the intervals it is assumed that the quantity [·] is equal to 1 for l < 2.

Discrete time models are obtained by grouping the observed times on a point of each interval, for example τ_l. In this context the

i) survival function

$$S(\tau_l,x_i)=P(Z>\tau_l,x_i)=\sum_{\nu>l}\tilde{f}_\nu(x_i)$$

where $\tilde{f}_\nu(x_i)$ is the *event probability* in the ν th interval

ii) conditional event probability (*discrete hazard*)

$$\tilde{h}_l(x_i)=P(Z\in A_l\mid Z>\tau_{l-1},x_i)=\frac{\tilde{f}_l(x_i)}{S(\tau_{l-1},x_i)} \qquad (2)$$

It follows that

$$S(\tau_l,x_i)=\prod_{\nu=1}^{l}\left[1-\tilde{h}_\nu(x_i)\right] \qquad (3)$$

The general expression of the likelihood function for the piecewise model for continuous time data with right censorship is given by

$$L_C=\prod_{i=1}^{N}\left\{h_{l_i}(t_i,x_i)^{d_i}\cdot\left[\prod_{\nu=1}^{l_i-1}\frac{S_\nu(\tau_\nu,x_i)}{S_\nu(\tau_{\nu-1},x_i)}\right]\cdot\frac{S_{l_i}(t_i,x_i)}{S_{l_i}(\tau_{l_i-1},x_i)}\right\} \qquad (4)$$

where l_i is the last interval in which the ith subject is observed. For grouped data, the equations (2) and (3) allow to express the likelihood function as

$$L_G=\prod_{i=1}^{N}\left\{\tilde{h}_{l_i}^{\,d_i}\cdot\left(1-\tilde{h}_{l_i}\right)^{1-d_i}\cdot\left[\prod_{\nu=1}^{l_i-1}\left(1-\tilde{h}_\nu\right)\right]\right\} \qquad (5)$$

3.1 Piecewises and Grouped Series Parametric Models

Piecewise and grouped times parametric models provide a flexible alternative to the traditional ones, Aitkin et al. (1). Concerning the continuous time model if $h_l(t,x_i)$ is assumed constant in each interval l, then Z will follow the exponential distribution with parameter $h_l(x_i)$. According to the (1), the piecewise exponential model will have

$$S(t,x_i)=\left[\prod_{\nu=1}^{l-1}\exp\left(-h_\nu(x_i)\cdot(\tau_\nu-\tau_{\nu-1})\right)\right]\cdot\exp\left[-h_l(x_i)\cdot(t-\tau_{l-1})\right]$$

Starting from (4), after defining the:

i) indicator variable d_{li} which is equal to 1 if, for the ith subject, the event of interest occurs in the interval A_l; or 0 otherwise and

ii) the risk exposure time $U_{il}=I(t_i>\tau_{l-1})\cdot\left[\min(t_i,\tau_l)-\tau_{l-1}\right]$, the

conditional likelihood function can be expressed as

$$L_P = \prod_{i=1}^{N}\prod_{l=1}^{l_i}\left\{h_l(x_i)^{\gamma_{il}}\cdot\exp\left(-h_l(x_i)\right)\right\}^{U_{il}} \qquad (6)$$

where $\gamma_{il} = d_{il}/U_{il}$ the (6) is proportional to the likelihood of $N\cdot l_i$ Poisson random variables, with proportionality constant $\prod_{i=1}^{N}\prod_{l=1}^{l_i}\left\{U_{il}^{\gamma_{il}\cdot U_{il}}\cdot\left[\left(\gamma_{il}\cdot U_{il}\right)!\right]^{-1}\right\}$

If the subjects can be grouped into *K* cells having equal covariate vectors *x*, a further version of the (6) is given by

$$\prod_{k=1}^{K}\prod_{l=1}^{L}\left\{h_l(x_k)^{\gamma_{kl}}\cdot\exp\left[-h_l(x_k)\right]\right\}^{U_{kl}} \qquad (7)$$

where $\gamma_{kl} = d_{kl}/U_{kl}$ (*empirical rates*), $d_{kl} = \sum_{i\in k} d_{il}$ and $U_{kl} = \sum_{i\in k} U_{il}$. The (7) is proportional to the likelihood of *K·L* Poisson r.v., with proportionality constant $\prod_{k=1}^{K}\prod_{l=1}^{L}\left\{U_{kl}^{\gamma_{kl}\cdot U_{kl}}\cdot\left[\left(\gamma_{kl}\cdot U_{kl}\right)!\right]^{-1}\right\}$.

Concerning grouped data, the conditional likelihood function can be obtained from the (5) as

$$L_G = \prod_{i=1}^{N}\prod_{l=1}^{l_i}\left\{\tilde{h}_l(x_i)^{d_{il}}\cdot\left[1-\tilde{h}_l(x_i)\right]^{1-d_{il}}\right\} \qquad (8)$$

this latter results from the product of Bernoulli likelihoods, one for each *i*th individual *i* in the *l*th interval in which he/she is observed. Grouping over *K* cells, the likelihood

$$L_G = \prod_{k=1}^{K}\prod_{l=1}^{L}\left\{\tilde{h}_l(x_k)^{p_{kl}}\cdot\left[1-\tilde{h}_l(x_k)\right]^{1-p_{kl}}\right\}^{n_{kl}} \qquad (9)$$

is obtained, with $p_{kl} = d_{kl}/n_{kl}$ (*empirical risks*), d_{kl} and n_{kl} are the number of events and observed subjects in the *k*th cell, respectively. The equation (9) is proportional to the likelihood of *K·L* independent binomial distributions, with proportionality constant $\prod_{k=1}^{K}\prod_{l=1}^{L} n_{kl}^{d_{kl}}$.

The piecewise exponential and grouped time models, considering a single event of interest, can be implemented on the basis of the above likelihood functions.

4. COMPETING RISKS

Quite often the clinical course of a disease is characterized by different possible events that represent the "failure causes" of the therapeutic intervention. In oncology, typically local-regional relapses, distant metastases, new primary tumours or death, may occur. From a statistical point of view, model suited to assess the dependence of the risk of each event (*cause specific hazard*, CSH) from the measured covariates are necessary, Marubini and Valsecchi (23).

Flexible linear approaches based on spline functions has been proposed for competing risks estimation, through the extension of generalized linear models with Poisson error, Boracchi et al., (10). However, when the a priori knowledge is limited, linear models may be more difficult to implement, being at risk of a possible overparameterisation. As previously mentioned, an alternative is represented by artificial neural networks (ANNs) models, which implicitly account for non-linear and non-additives effects of covariates, Biganzoli et al., (3).

In the presence of *R* different types of events, data may be represented by the random variables $(T,\delta,\delta\rho)$, where ρ has observed values $r=1,\ldots,R$ and $T=\min(C,Z^1,\ldots,Z^R)$. Considering only the case that *T* is continuous, the CSH functions are the instant hazard rate for the *r*th event, in the presence of the other failure causes, given the absence of events before *t*;

$$h(r,t,x_i) = \lim_{\Delta t\to 0}\frac{P\left(t<T\le t+\Delta t,\rho=r\,|\,T>t,x_i\right)}{\Delta t} \qquad (10)$$

The survival function can be expressed as a function of the CSHs

$$S(t,x_i) = \exp\left(-\int_0^t h(.,u,x_i)du\right)$$

where

$$h(.,u,x_i) = \sum_{r=1}^{R} h(r,u,x_i) \qquad (11)$$

The (11) holds without any assumption about the independence of the risks. From the (10) the cause specific survival is defined as

$$G(r,t,x_i) = \exp\left(-\int_0^t h(r,u,x_i)du\right)$$

and, given the (11), it holds

$$S(t,x_i) = \prod_{r=1}^{R} G(r,t,x_i)$$

Concerning the piecewise exponential model for competing risks, see Larson (18), and Kramar et al., (17), the CSH functions are assumed constant in each interval: $h(r,l,x_i)$.

Introducing the indicator variable d_{ilr}, equal to 1 if the *i*th patient, in the *l*th time interval, had the *r*th event and 0 otherwise, from previous relationships the likelihood function for competing risks is

$$L_{PR} = \prod_{k=1}^{K}\prod_{l=1}^{L}\prod_{r=1}^{R}\left\{h(x_k,r,l)^{\gamma_{klr}}\cdot\exp\left[-h(x_k,r,l)\right]\right\}^{U_{kl}} \qquad (12)$$

where $\gamma_{klr} = d_{klr}/U_{kl}$. The (12) is therefore an extension of the (7) to the case of competing risks, assuming γ_{klr} as Poisson random variable.

Concerning grouped time models, defining $e_{klr} = \sum_{i\in k} d_{ilr}$, considering *K* design cells on the basis of (equal) vectors x_k, if a *R*+1 additional "at risk"

indicator $d_{il(R+1)}$ is introduced, equal to 1 in the interval A_l in which the subjects is observed without failure, and equal to 0 otherwise, $e_{kl(R+1)} = \sum_{i \in k} d_{il(R+1)} = n_{kl} - \sum_{i=1}^{R} e_{klr}$ the likelihood can be expressed as

$$L = \prod_{k=1}^{K} \prod_{l=1}^{L} \prod_{r=1}^{R+1} \left[\tilde{h}_{lr}(x_k) \right]^{e_{klr}} \qquad (13)$$

The (13) is proportional to the likelihood of $K \cdot L$ independent multinomial distributions, where $\prod_{k=1}^{K} \prod_{l=1}^{L} \frac{n_{kl}}{\prod_{r=1}^{R+1} e_{ilr}!}$ is the proportionality constant. Under the constraint

$$\sum_{r=1}^{R+1} \tilde{h}_{lr}(x_k) = 1 \qquad (14)$$

the quantity $\tilde{h}_{l(R+1)}(x_k)$ represent the conditional probability of being censored in the *l*th time interval.

5. GLMS AND FFANNS

A general framework for the development of FFANNs for survival data is that of the GLMs, McCullagh and Nelder (24). For such models, it is assumed that each component of the r.v. Y has a distribution $f(y;\theta,\phi)$ in the exponential family, whose log-likelihood function is given by

$$l(\theta,\phi;y) = \log[f(y;\theta,\phi)] = [y\theta - b(\theta)]/a(\phi) + c(y,\phi)$$

If ϕ is known, $f(y;\theta,\phi)$ is an exponential family model with the *canonical parameter* θ. It can be shown that $E(Y) = \mu = b'(\theta)$ and $Var(Y) = b''(\theta)a(\phi)$. Typically, $a(\phi) = \phi / w$, where ϕ is a constant *dispersion parameter* and w is a *prior weight* that varies from observation to observation. In a GLM, μ is related to the systematic part of the model η (*predictor)* by the *link function* $g(\mu) = \eta$. The predictor η has linear additive form $\eta = \beta_0 + \beta^T x$, with β_0 intercept and β the vector of regression coefficients. FFANNs can be considered as GLMs with a non-linear predictor, Biganzoli et al., (4), as a multi-layer perceptron (MLP)

$$\eta = \beta_0 + \sum_{h=1}^{H} \beta_h^{\alpha} \cdot \alpha_h(\beta_{0h} + \beta_h^T x) \qquad (15)$$

or a RBF network

$$\eta = \beta_0 + \sum_{h=1}^{H} \beta_h^{\varphi} \cdot \varphi_h(\|x - c_h\|) \qquad (16)$$

with *h*=1,…,*H hidden units*, α_h, φ_h *activation functions* and c_h the centre of the *h*th radial basis, Bishop (9). Expressions (15) and (16) could be called *neural predictors*.

5.1 Regression models for Survival Data

Following the preceding considerations, GLMs may be adopted for studying the dependence of the hazard function from the covariates *x*. Without loss of generality, models for subjects grouped into *K* cells will be considered. For fitting the piecewise exponential and grouped time models, distributions from the exponential family are adopted, namely the Poisson for the *empirical rates* γ_{kl} and the binomial for *empirical proportions* p_{kl}. To fit the two models, it is useful to minimize the distance function given by the difference between the maximum log-likelihood achieved by the perfect fit of each observation $l(y,\phi;y)$ and that achieved for the model under investigation $l(\hat{\mu},\phi;y)$, with estimates of the canonical parameter denoted by $\breve{\theta} = \theta(y)$ and $\hat{\theta} = \theta(\hat{\mu})$*;* respectively. Thus:

$$E = \sum_{k=1}^{K} \sum_{l=1}^{L} \left[y_{kl}(\breve{\theta}_{kl} - \hat{\theta}_{kl}) - b(\breve{\theta}_{kl}) + b(\hat{\theta}_{kl}) \right] \cdot w_{kl} / \phi$$

which amounts to half the *(scaled) deviance* of the model, McCullagh and Nelder (24). This statistic allows defining in general way error functions for generalized regression models to be fitted with ANNs. The corresponding error functions for the piecewise exponential and the grouped time models for a single and competing risks are respectively

$$E_P = \sum_{k=1}^{K} \sum_{l=1}^{L} \left\{ \gamma_{kl} \log\left[\frac{\gamma_{kl}}{h_l(x_k)} \right] - \gamma_{kl} + h_l(x_k) \right\} \cdot U_{kl} \qquad (17)$$

$$E_{PR} = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{r=1}^{R} \left\{ \gamma_{krl} \log\left(\frac{\gamma_{krl}}{h_{lr}(x_k)} \right) - (\gamma_{krl} - h_{rl}(x_k)) \right\} \cdot U_{kl} \qquad (18)$$

and

$$E_G = \sum_{k=1}^{K} \sum_{l=1}^{L} \left\{ p_{kl} \log\left[\frac{p_{kl}}{\tilde{h}_l(x_k)} \right] + (1 - \hat{p}_{kl}) \log\left[\frac{1 - p_{kl}}{1 - \tilde{h}_l(x_k)} \right] \right\} \cdot n_{kl} \qquad (19)$$

$$E_{GR} = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{r=1}^{R+1} \left\{ p_{klr} \log\left[\frac{p_{klr}}{\tilde{h}_{lr}(x_k)} \right] \right\} \cdot n_{kl} \qquad (20)$$

where U_{kl} and n_{kl} have the role of prior weights.

For a single risk in the discrete context, a proportional odds model with the *logit* link, canonical with respect to (8), was proposed by Cox (12) as

$$\log\left[\frac{\tilde{h}_l(x_k)}{1 - \tilde{h}_l(x_k)} \right] = \beta_0 + \beta^T x_k \qquad (21)$$

where $\beta_l^0 = \log\left[\frac{\tilde{h}_l(0)}{1-\tilde{h}_l(0)}\right]$. It is analogous to ANN models (without hidden units) for classification problems with error function (19), since the logit link is the inverse of the logistic activation function. The partial logistic artificial neural network (PLANN), proposed by Biganzoli et al. (3), follows such an approach to provide smoothed discrete hazard estimates by adopting a neural predictor (15) for model (21) and relaxing additivity constraints. The resulting MLP model is parameterized as follows

$$\tilde{h}_l(x_k) = \frac{\exp\left\{\beta_0 + \sum_{h=1}^{H} \beta_h^{\alpha} \cdot \alpha_h\left(\beta_{0h} + \beta_h^T v_{kl}\right)\right\}}{1 + \exp\left\{\beta_0 + \sum_{h=1}^{H} \beta_h^{\alpha} \cdot \alpha_h\left(\beta_{0h} + \beta_h^T v_{kl}\right)\right\}}$$

corresponding to the well known regression model for binary classification, with the logistic activation function $\alpha_h(u) = \frac{\exp(u)}{1+\exp(u)}$; the additional input for the time interval is included in $v_{kl} = (x_k, \tau_l)$. In the Poisson model for competing risks, an input vector $v_{klr} = (x_k, \tau_l, r)$ is used, with a further input for the specific cause of failure. The discrete time model for competing risks can be fitted by modelling the e_{klr} with GLM with multinomial error and the *canonical* inverse multinomial logit link.

$$\tilde{h}_{lr}(x_k) = \frac{\exp[\eta_{lr}(x_k)]}{\sum_{r=1}^{R+1} \exp[\eta_{lr}(x_k)]} \qquad (22)$$

where $\eta_{lr}(x_k)$ is the model predictor, which can be either linear or neural as

$$\eta_{lr}(x_k, \beta) = \beta_0 + \sum_{h=1}^{H} \beta_r^{\alpha} \alpha_h\left(\beta_{0h} + \beta_h^T v_{kl}\right)$$

Therefore, the single risk model has a logistic output activation function, whereas, for competing risks, the (22) is used considering *R* + 1 outputs. Such a function, called *softmax* in the ANN jargon, corresponds to the multinomial generalisation of the logistic function.

It can be also considered that in such models, though approximate, the assumption of independence of the contribution to the likelihood for each individual across time intervals leads to reasonable results, extending arguments for the binomial model provided by Efron (13). Since the grouped time competing risk model can also be viewed as an extension of PLANN, Biganzoli et al. (3), (4), to multiple failures causes in a competing risks framework, it will be denoted by the acronym PLANNCR, Biganzoli et al. (8).

In an analogous way, the log-linear piecewise exponential models with canonical link $\log[h_l(x_i)]$ can be considered, Larson (18). A development of the approach was proposed in the competing risks framework, Boracchi and Biganzoli (11), by considering the RBF expansion (16) as a model predictor and the error function (17).

6. APPLICATIONS TO CANCER DATA

To show the general properties of the modelling approach, a problem of competing risks in breast cancer; is considered. The risk of local relapses (IBTR) and distant metastasis (DM) has been studied according to the patient age at surgery, tumour size, histological type, number of axillary metastatic lymph nodes and site of the tumor. The study includes 2233 patients hospitalized at the Istituto Nazionale per Studio e la Cura dei Tumori di Milano between 1970 and 1987. Details on the study and on the strategy adopted for the evaluation of the artificial neural net RBF models have been reported in the paper of Boracchi et al. (10). Globally, not monotonic patterns of the CSH functions in time and different covariate effects have been observed for the two considered event: IBTR and DM. The prognostic impact of age, tumour size and histology on IBTR appears more evident than that of the number of axillary metastatic lymph nodes and of tumour site. The CSH for IBTR decreases with the increase of the age (Fig. 1a in section 10), increases with the increasing of tumour size (Fig. 1b in section 10) and tends to decrease with the increase of the number of axillary metastatic lymph nodes (Fig. 1c in section 10). Concerning DM, the pattern of the risk in the time appears markedly not-monotonic and the maximum value is observed at about to two years and half of follow-up (Fig.2 in section 10). The impact of the tumour size and of the metastatic lymph nodes on DM appears to be more marked than that of the age, histology and tumour site. CSH weakly decreases as age increases, increases as tumour size and the number of axillary metastatic lymph nodes increase. The effect of the tumour size decreases with follow-up time, pointing out the possible time dependent role of this prognostic variable, in agreement with the findings in other case series.

7. CONCLUSIONS

Building a multiple regression model for outcome prediction is a problem of approximation of an unknown multivariate dependence relationship. In presence of non linear effects and/or non additives effects, an approach commonly assumed as ”natural” is that of including in the model, Schwarzer et al (26), polynomial terms and their cross-products. However, this approach is not necessarily optimal if low a priori knowledge of the phenomenon is available. FFNN and RBF models offer an alternative model parameterisation not constrained to strong assumptions on the effect of the covariates. Therefore, the use of neural network models for outcome prediction could be mostly relevant for exploratory analyses and as a benchmark for assessing the performances of other concurrent models. The PLANN(CR) model can provide relevant indications on the underlying patterns to the prognostic problem under evaluation, thus substantially contributing to the individual risk bioprofiling. Moreover the Bayesian extension of PLANN proposed by Lisboa et al. (21)

faces the aspect of the optimal control of model complexity in a principled way.

In recent research works, there was a specific interest in the application of advanced regression techniques such as neural networks for the analysis of genomic/proteomic data. Overall, a critical aspect is related to the evaluation of the model performance to assess the true gain from such approaches when applied on noisy data from microarray analyses, Biganzoli and Boracchi (4). Several applications regarded outcome prediction, but most of them failed to account for censored survival data. The latter situation is partly motivated by study design issues, Biganzoli et al (7).

Nevertheless, waiting for the refinement of the omic techniques, a critical role of neural networks techniques for censored failure data can be played now, Biganzoli et al (5) for a substantial improvement of outcome prediction strategies in cancer, based on traditional and/or new clinical and biological markers.

8. ACKNOWLEDGMENTS

This work is partly supported by EU *BIOPATTERN* project FP6-2002-IST-1 N° 508803.

9. REFERENCES

1. Aitkin, M., Anderson, D., Francis, B., and Hinde, J., 1989, "Statistical Modelling in GLIM", Oxford University Press, New York.
2. Baxt,W.G., 1995, "Application of artificial neural networks to clinical medicine", Lancet, 346, 1135-1138.
3. Biganzoli, E., Boracchi P., Mariani L., and Marubini E., 1998, "Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach", Statistics in Medicine, 17, 1169-1186.
4. Biganzoli, E., Boracchi, P., and Marubini, E., 2002, "A general framework for neural network models on censored survival data", Neural Networks, 15, 209-218.
5. Biganzoli, E., Boracchi, P., Coradini, D., Daidone, M.G., and Marubini E., 2003, "Prognosis in node-negative primary breast cancer: a neural network analysis of risk profiles using routinely assessed factors", Ann. Oncol. 14, 1484-93.
6. Biganzoli, E., Boracchi, P., 2004, "Old and new markers for breast cancer prognosis: the need for integrated research on quantitative issues", Eur J Cancer. 40, 1803- 1806.
7. Biganzoli, E., Boracchi, P., Ambrogi, F., and Marubini, E., 2005, "Artificial neural network models for discrete cause specific hazards", *Technical Report*. Annex of deliverable D11 of the NoE Biopattern: *State-of-the-art report of artificial neural networks used in survival analysis of patients suffering from breast cancer*.
8. Biganzoli, E., Lama, N., Ambrogi, F., Antolini, L., and Boracchi, P., 2005, "Prediction of cancer outcome with microarrays. (letter)", The Lancet, 365, 1683
9. Bishop, C.M., 1995, "Neural networks for pattern recognition", Oxford University Press, New York.
10. Boracchi, P., Biganzoli, E., and Marubini E., 2001, "Modelling cause-specific hazards with radial basis function artificial neural networks: application to 2233 breast cancer patients", Statistics in Medicine, 20, 3677-3694.
11. Boracchi, P., Biganzoli, E., 2002, "Radial Basis Function Neural Networks for the Analysis of Survival Data", Metron, 60, 191-210.
12. Cox, D.R., 1972, "Regression models and life-tables (with discussion)", Journal of the Royal Statistical Society Series B, 34, 187-220.
13. Efron, B., 1988, "Logistic regression, survival analysis, and the Kaplan-Meier curve" Journal of the American Statistical Association, 83, 414-425.
14. Fahrmeir, L., 1994, "Dynamic modelling and penalized likelihood estimation for discrete time survival data", Biometrika, 81, 317-330.
15. Faraggi, D., Simon, R., 1995, "A neural network model for survival data", Statistics in Medicine, 14, 73-82.
16. Klein, J.P., Moeschberger, M.L., 1997, "Survival analysis. Techniques for censored and truncated data", Springer-Verlag, New York.
17. Kramar, A., Pejovic, M. H., and Chassagne, D., 1987, "A method of analysis taking into account competing events: application to the study of digestive complications following irradiation for cervical cancer", Statistics in Medicine, 6, 785-794
18. Larson, M.G., 1984, "Covariate analysis of competing-risks data with log-linear models", Biometrics, 40, 459-468.
19. Liestol, K., Andersen, P.K., and Andersen, U., 1994, "Survival analysis and neural nets", Statistics in Medicine, 13, 1189-1200.
20. Lisboa, P.J., 2002, "A review of evidence of health benefit from artificial neural networks in medical intervention", Neural Netw. 15, 11-39.
21. Lisboa, P.J., Wong, H., Harris, P., and Swindell, R., 2003, "A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer", Artif Intell Med,. 28, 1-25.
22. Mariani, L., Coradini, D., Biganzoli, E., Boracchi, P., Marubini, E., Pilotti, S., Salvadori, B., Silvestrini, R., Veronesi, U., Zucali, R., and Rilke, F., 1997, "Prognostic factors for metachronous contralateral breast cancer: a comparison of the linear Cox regression model and its artificial neural network extension", Breast Cancer Research & Treatment, 44, 167-178.
23. Marubini E., Valsecchi M.G., 1995, "Analysing survival data from clinical trials and observational studies", Chichester, John Wiley & Sons.
24. McCullagh, P., Nelder, J. A., 1989, "Generalized linear models", 2nd edition, Chapman & Hall, London, UK.

25. Ripley, B. D., Ripley, R. M., 1998, "Neural networks as statistical methods in survival analysis", In R. Dybowski and V. Gant (Eds.), *Artificial Neural Networks: Prospects for Medicine*, Cambridge University Press, Cambridge, UK, (In press).

26. Schwarzer, G., Vach, W., and Schumacher, M., 2000, "On the misuses of artificial neural networks for prognostic and diagnostic classification in oncology", Statistics in Medicine, 19, 541-561.

10. LIST OF FIGURES

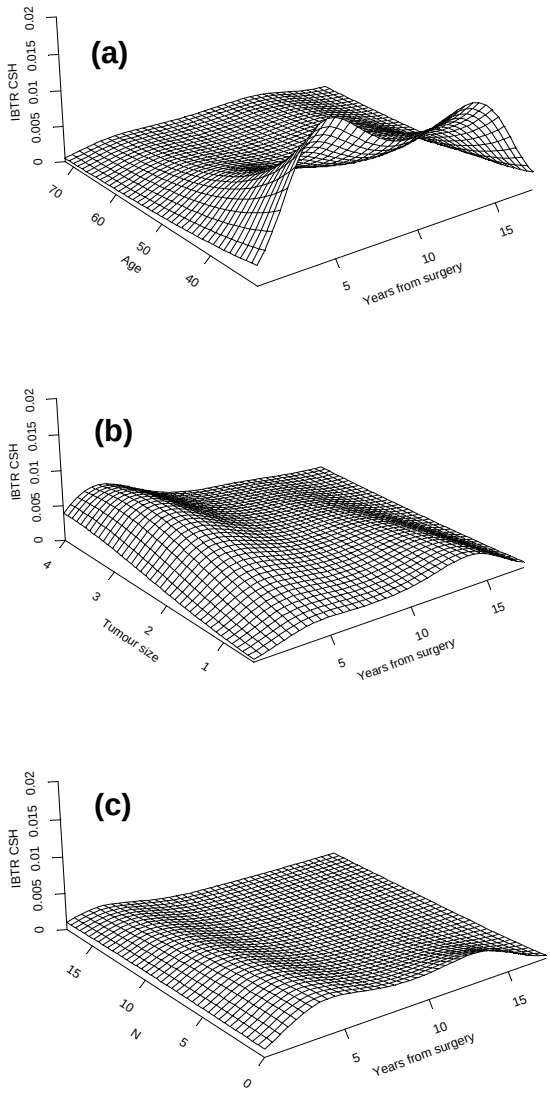

Figure 1. Graph of the conditioned surface of the CSHs for IBTR as a function of age (a), tumour size (b) number of the metastatic axillary lymph nodes (c) other covariates are fixed to the median values for age and tumour size and to modal categories for tumour site and histological type.

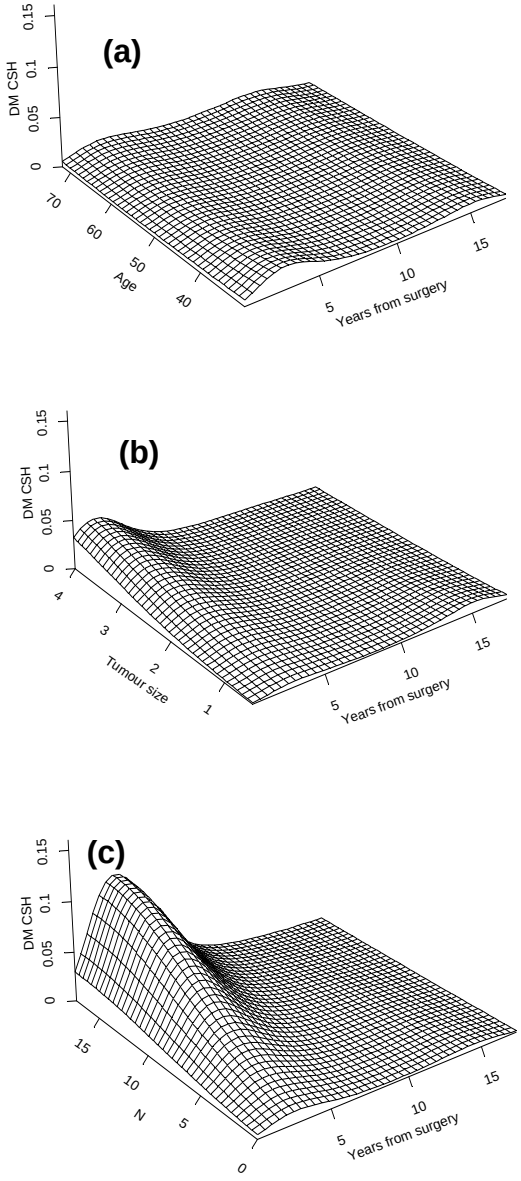

Figure 2. Graph of the conditioned surface of the CSHs for DM as a function of age (a), tumour size (b) number of the metastatic axillary lymph nodes (c) other covariates are fixed to the median values for age and tumour size and to modal categories for tumour site and histological type.