Research Article

A special case of Reduced Rank Models for identification and modelling of Time Varying Effects in Survival Analysis

Aris Perperoglou*

Flexible survival models are in need when modelling data from long term follow up studies. In many cases, the assumption of proportionality imposed by a Cox model will not be valid. Instead, a model that can identify time varying effects of fixed covariates can be used. Although there are several approaches that deal with this problem, it is not always straightforward how to choose which covariates should be modelled having time varying effects and which not. At the same time, it is up to the researcher to define appropriate time functions that describe the dynamic pattern of the effects. In this work we suggest a model that can deal with both fixed and time varying effects, and uses simple hypotheses tests to distinguish which covariates do have dynamic effects. The model is an extension of the parsimonious reduced rank model of rank 1. As such, the number of parameters is kept low and thus a flexible set of time functions, such as b-splines, can be used. The basic theory is illustrated along with an efficient fitting algorithm. The proposed method is applied to a dataset of breast cancer patients and compared with a multivariate fractional polynomials approach for modelling time-varying effects.
Copyright © 2014 John Wiley & Sons, Ltd.

1. Introduction

Fitting survival models in long term follow up studies presents challenges. In addition to the two interrelated problems of selecting which variables should be included in the model, and the choice of their functional forms, in survival models one also needs to investigate whether some of the covariates have an effect that changes during the follow-up period. An example of such a dynamic behaviour can be the effect of a surgical treatment that might fade away as time passes by [1]. To address the problem of modelling time varying effects of covariates several approaches have been introduced such as additive models [2], hazard regression [3], tree-based models [4] or Bayesian models [5]. See also the book of Martinussen and Scheike [6] for a review and discussion on dynamic models.

The emphasis of this paper is given however on models that are simple extensions of Cox's proportional hazards model (PH) [7]. In many approaches, some or all of the covariates are allowed to interact with time functions. As such, there are two specific issues that need to be addressed. One is how to select variables that do have time varying effects (TVEs) and the other is how to choose an appropriate set of time functions for describing a time varying behaviour.

Regression splines have always been a popular choice, with many different alternatives based on the type of spline bases [8, 9, 10]. To deal with the problem of choosing an optimal number of knots, or overfitting the data, penalized splines have also been suggested [11, 12]. Penalized approaches aim to reduce the effective dimensions of the model and thus correct any extreme behaviour of TVEs. An earlier version of a penalized model was introduced by Verweij and van Houwelingen where time varying effects were overparametrised but then controlled by a first order penalty on the coefficients [13]. On the other side of the spectrum, parsimonious models can be found where fractional polynomials [14] are used for modelling time varying effects [15, 16]. Parsimony can also be achieved by the use of reduced rank models [17] where the number of parameters is controlled by the rank of the matrix of coefficients.

*Correspondence to: Aris Perperoglou, Department of Mathematical Sciences, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK

Choosing which of covariates effects should be modelled as dynamic is not an easy task. In some applications there might be biological reasoning that will indicate which of the covariates should have dynamic effects. However, in other cases the model needs to fit the data without any previous assumptions. Reduced rank models assume that all covariates have time varying effects and have been criticised for this feature [18]. Sauerbrei, Royston and Look [16] proposed an algorithm to identify which of the covariates should have time varying effects and which not. Another common approach is a test on Schoenfeld residuals [19] that can identify departures from proportionality.

In this work, an approach that belongs to the family of reduced rank models is investigated. In Section 2 we will provide a basic summary of reduced rank models used for modelling non proportional hazards. In Section 3 we will show how a rank=1 model, can be extended to a Hybrid Reduced Rank model, an in-between model of the rank=1 and rank=2 models. The model has been first suggested in van Houwelingen and Putter [20], page 100, but never applied before. Here, we investigate the suggestion and show how it can be fitted using a parametrisation based on a rank=2 model. We will show how this new parametrisation can lead to a flexible model that allows for covariate effects to be modelled in a proportional and a non-proportional hazards part of the model, simultaneously. The extra number of parameters can serve in two ways; either use the estimated parameters to perform hypothesis testing and identify which of the covariates do need time varying effects, or fit a full model that will utilise the extra parameters to approximate the true nature of covariate effects, even when not correctly specified. In Section 4 the method will be applied to a data set of Breast cancer patients. To evaluate the performance of the Hybrid model we will compare it with models that use fractional polynomials for the identification and modelling of TVEs. We will also illustrate a modelling strategy when it is required to choose the variables with dynamic effects. Finally, the performance of Hybrid models will be investigated under a simulation study in Section 5. The paper closes with a discussion.

2. Reduced Rank hazards regression

Consider time to event data given in the usual notation (d_i, t_i), where t_i denotes time until an event of interest for subject i, d_i is a status indicator (1=event, 0=censored) and i spans over n subjects at risk. Also, assume information on p covariates that form the X matrix of $n \times p$ dimensions. An extension of the proportional hazards model that allows for time varying effects of covariates can be achieved by adding an interaction of X covariates with a time function $f(t)$. In a simple form the model can be written as:

$$h(t|X) = h_0(t) \exp(X\beta f(t))$$

where $h_0(t)$ is a baseline hazard function and β are the risk coefficients to be estimated.

If more than one time functions are used, it is advantageous to use a different notation. Let $F_{n \times q}$ be a matrix of q time functions, where the i-th row and the j-th column, F_{ij}, contains the the j-th time function evaluated at time point i. Under the restriction of the first column being constant, a model with time varying effects of the covariates is given by:

$$h(t|X) = h_0(t) \exp(X\Theta F') \tag{1}$$

where the coefficients are in a matrix $\Theta_{p \times q}$.

Starting from model (1) one may factorize Θ as a product of two submatrices, such that $\Theta = B\Gamma'$ with B and Γ of $p \times r$ and $q \times r$ dimensions, respectively. The resulting Θ matrix is of full rank when $r = \min(p, q)$ and the number of dimensions for the model is given by: $r(p + q - r)$. Alternatively, the model can also be written as:

$$h(t|X) = h_0(t) \exp\left\{ X \sum_{k=1}^{r} \beta_k \sum_{j=1}^{q} \gamma_{jk} f_j(t) \right\} \tag{2}$$

where β_k is the k-th column of B and γ_k the k-th column of Γ. Depending on the rank r of the matrix data can be fitted using a full rank model, or parsimony can be achieved by reducing the rank of the matrix and thus the dimensions of the model. For example, when $r = 2$, the model becomes:

$$h(t|X) = h_0(t) \exp\left\{ X\beta_1 \sum_{j=1}^{q} \gamma_{j1} f_j(t) + X\beta_2 \sum_{j=1}^{q} \gamma_{j2} f_j(t) \right\} \tag{3}$$

In the original paper [17], Perperoglou et al. suggested the use of a matrix with flexible time functions, such as b-splines, and then let the rank of the model take care of overfitting or wild behaviour of the effects. The authors suggested fitting all possible reduced rank models and then selecting the one with the minimum AIC.

2.1. The rank=1 model

A rank=1 model emerges from equation (2), when $r = 1$, such that:

$$h(t|X) = h_0(t) \exp\left(X\beta \sum_{j=1}^{q} \gamma_j f_j(t) \right) \tag{4}$$

where the time functions are restricted to $f_1(t) \equiv 1$ and $f_j(0) \equiv 0$ for all j not equal to 1. Then, the β vector of coefficients corresponds to covariate effects right at the start of the follow up period, when time $t = 0$. As such, they can be regarded as the initial effect of the covariate on the hazard. Each covariate in matrix X has a different initial effect, but the time varying behaviour, described by $\gamma_j f_j(t)$ is common for all variables. Therefore, a rank=1 model is proportional on the time varying effects of the covariates.

Estimation of the model can be done via an alternating least squares algorithm. Estimates for βs are obtained for some initial values of γ, then estimates for the γs are updated and the process alternates until no significant change to the likelihood occurs.

It should be noted that the factorisation in equation (4) is not unique. Without restrictions, β and γ are not identifiable, however, their linear combination in matrix Θ (see equation 1) is. For more theoretical details and an efficient algorithm to fit the models refer to [17] and [21], respectively. More information is also available on the book by Yee [22], page 170.

2.2. Time varying and time fixed effects

Reduced rank models were extended [23] to include both time fixed and time varying effects. In these models, two covariate matrices were formed, one for covariates that are only allowed to have time fixed effects, denoted by X_1, and a second one with covariates that are allowed to change their effect dynamically in time, given as X_2. A model with both time varying and time fixed covariate effects may be written as:

$$h(t|X) = h_0(t) \exp\left(X_1\xi + X_2\beta \sum_{j=1}^{q} \gamma_j f_j(t) \right) \tag{5}$$

where ξ is a vector of coefficients for the fixed effects.

3. Hybrid Reduced Rank model

When fitting survival data an important question is how to distinguish which covariates belong to matrix X_1 and which to matrix X_2. Consider a special case of model (5) with rank=1, and let $X = X_1 = X_2$, i.e. let all covariates in the fixed effects X_1 matrix also belong in the X_2 matrix of time varying effects. This is the *Hybrid Reduced Rank* model (HyR). Starting from model (5) and rewriting ξ as β_1 and β as β_2 then the Hybrid model is given by:

$$h(t|X) = h_0(t) \exp\left(X\beta_1 + X\beta_2 \sum_{j=1}^{q} \gamma_j f_j(t) \right) \tag{6}$$

where, as before, the time functions are restricted to $f_1(t) \equiv 1$ and $f_j(0) \equiv 0$ for all j not equal to 1. A simple rank=1 model has a few number of parameters to be estimated which depend on the number of covariates and the number of time functions $p + q - 1$. The Hybrid model requires twice the number of fixed coefficients, making the total number of parameters $2p + q - 1$.

Although a Hybrid model seems non-identifiable, the redundancy of the model can be remedied by using a rank=2 type parametrisation with special restrictions on the parameter estimates. In a rank=2 model, the matrix of Θ coefficients is parametrised as the product of $B_{p \times 2}$ with $\Gamma_{q \times 2}$, where B would contain the two columns β_1 and β_2, thus resulting in the model given by equation (3). However, when $\gamma_{11} = 1$, $\gamma_{j1} = 0$ for $j \neq 1$ and γ_{j2} have the same restrictions as a rank=1 model, then equation (3) reduces to the Hybrid model given in (6). That ensures that $\beta_1 + \beta_2$ and $\beta_2 + \gamma$ as given in (6) are identifiable. A detailed simple illustration is presented in Section 4.5.

The alternating algorithm [21] to fit reduced rank models, can be adapted to fit the special case of the Hybrid model. Define as vec(Θ) a function that vectorises Θ, per column, and let vec(Θ) = Θ_ν, vec(Γ) = Γ_ν and vec(B) = B_ν. Let also:

$$\frac{\partial \Theta_\nu}{\partial B_\nu} = T_B(\Gamma) = \Gamma \otimes I_p$$

Statist. Med. **2014**, 00 1–14
Prepared using **simauth.cls**

Copyright © 2014 John Wiley & Sons, Ltd.

www.sim.org   3

a $(p \times q) \times p$ matrix, with $\otimes$ denoting the kronecker product.

In the first step of fitting a rank=r model, β estimates are updated by:

$$\left[(\Gamma \otimes I_p)^T H (\Gamma \otimes I_p) \right]^{-1} (\Gamma \otimes I_p)^T \beta_\nu$$

where H is the hessian matrix, I_p is a diagonal matrix of ones with p dimensions. Here, we need to extend $T_B(\Gamma)$ into a matrix that has the following form:

$$T_B^*(\Gamma) = \begin{bmatrix} \Gamma \otimes I_p & \vdots & \mathbf{0} \\ \cdots & \cdots & \cdots \\ \mathbf{0} & \vdots & I_p \end{bmatrix}$$

Similarly, in the next step we have: $\frac{\partial \Theta_\nu}{\partial \Gamma_\nu} = T_\Gamma(B) = I_q \otimes B$
with

$$T_\Gamma^*(B) = \begin{bmatrix} T_\Gamma(B) \\ \cdots \\ \gamma_2 \otimes \beta_2 \end{bmatrix}$$

Then γ estimates are updated by:

$$\left[T_\Gamma^*(B)^T H T_\Gamma^*(B) \right]^{-1} T_\Gamma^*(B)^T \gamma_\nu$$

The algorithm then alternates between the two steps, until the estimates stabilise.

3.1. Standard errors and hypothesis testing

In any reduced rank model the Hessian matrix obtained at the end of the iterative procedure is also restricted on the rank of the model. As such, the covariance matrix is singular. To obtain standard errors for the estimates a delta method can be used, as illustrated in [17, 21]. The covariance of vector Θ_ν is given as: $\text{cov}(\Theta_\nu) = D(D'HD)^{-1}D'$, where

$$D = \begin{pmatrix} \frac{\partial \Theta_\nu}{\partial B_\nu} \\ \frac{\partial \Theta_\nu}{\partial \Gamma_\nu} \end{pmatrix}$$

Since $(D'I(\Theta)D)^{-1}$ is also singular, the matrix is inverted by using a Moore-Penrose inverse.

To take full advantage of the rank=1 model, it would also be interesting to test whether any of the coefficients in vector β_2 is zero. The hypotheses of interest are $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 \neq 0$. Failing to reject the null would cancel out the totality of a time varying effect, not just an interaction with a specific time function. Since β_2 is only identifiable here when specific restrictions are posed on the γs and also depend on β_1, we can only obtain conditional standard errors for the test. Thus, standard errors for the beta coefficients have to be estimated at the intermediate step by using the variance covariance matrix $\text{cov}(\beta) = (T_B^*(\Gamma)'HT_B^*(\Gamma))^{-1}$.

4. Application

4.1. Data Description: IASO breast cancer data

The data set contains information on 1937 breast cancer patients treated between 1978 and 2000 in Alexandra University Hospital and IASO Hospital, in Athens Greece. This is a subset of the data first published in [24] and includes patients with up to ten years of follow up, with a median of roughly four years. During this period 531 patients died due to their disease. There is also available information on clinical characteristics, such as tumour size (measured in mm), tumour grading (stage I,II or III), the age of a patient at the start of follow up, the number of lymph nodes involved and the type of treatment that patients received (chemotherapy, radiotherapy and hormonal treatment). Information about the covariates is presented in Table 1.

The event of interest is time from diagnosis to death from breast cancer. Before fitting any model all continuous covariates (age, tumour size and number of positive lymph nodes) have been standardised to have zero mean and variance one. A simple proportional hazards model on the data showed that the effects of age and radiotherapy are not significant, chemotherapy and hormonal treatment had a protective effect and all other covariates increased the risk of death from breast cancer (data not shown).

Variable	Freq (%)
Tumour Grade	
I	217 (11%)
II	1148 (59%)
III	572 (30%)
Treatment	
Chemotherapy	1149 (59%)
Radiotherapy	1414 (73%)
Hormonal	1507 (77.8%)
Median (Min, Max)	
Age	57 (24,98)
Lymph nodes	1 (0,57)
Tumour size	23 (1,120)

Table 1. Variables in the IASO breast cancer patients dataset

4.2. Modelling with Fractional Polynomials

The Multivariable Fractional Polynomials Time (MFPT) algorithm [16, 18] is a procedure that relies on the use of fractional polynomials to test first for the presence of non-linear relationships of covariates with the hazard, then identify which covariates should be included in the model and finally investigate whether some covariates have a time varying effect on the hazard function. The approach is used here to end up with an optimal Fractional Polynomial model (FPm) which will then be compared with a fit from a Hybrid model.

On the first step, two of the covariates were shown to have a more complex functional form: age, that was transformed to $(age + 2.7)^{-2}$ and the number of positive lymph nodes that was transformed to $\sqrt{nodes + 0.7}$. All other covariates were included in their original form, besides radiotherapy which had a non-significant effect. That first step involved fitting 44 models (8 fractional polynomial models of degree 1 and 36 models of degree 2) for each continuous covariate tested. On the second step, covariates that have a short term effect were included to the model. At this stage no new variable was added to the model. Next, a fractional polynomial transformation of degree 1 and 2 is applied on the time variable. Each variable is then tested for a time varying effect by testing for the significance of the interaction of the variable with the fractional polynomials. Out of the total of 44 models for each covariate, the best one is selected based on testing the deviance of each model. Table 2 presents the results of the final model.

In the final model four covariates were modelled to have time varying effects: hormonal treatment and tumour grading both interact with a first degree fractional polynomial of time (to the power 0, defined as $t^0 = \log(t)$), the number of positive lymph nodes interact with a first degree fractional polynomial of time (to the power of -2), while for chemotherapy a second degree fractional polynomial was chosen with powers, -2 and -2 (where the two time functions are defined as t^{-2} and $t^{-2} \log(t)$, respectively).

4.3. Modelling with a Hybrid Reduced Rank model

Before fitting the Hybrid Reduced Rank model (HyR) the first step of an MFPT algorithm was applied to assure that covariates with non linear functional forms will be correctly identified. As such, age and the number of positive lymph nodes where transformed in the same way as before. We choose to use a flexible matrix of several time functions to describe the time varying effects. A matrix of third degree b-spline bases was created, with boundary knots defined at the range of time and no interior knots. That produced a matrix $F = [1, f_1(t), f_2(t), f_3(t)]$ of four time functions. Having all covariates interacting with F resulted in the model given in equation (6). This is a model with $2p + q - 1 = 15$ degrees of freedom, where each covariate interacts with four time functions. An estimate of the covariates effects at time $t = 0$ is given just by adding $\hat{\beta}_1 + \hat{\beta}_2$. The effects are very similar to those given from FPm.

4.4. Model Building

To motivate a model building strategy assume a simple model where -for illustration purposes- only three covariates are considered: tumour size, tumour grading and chemotherapy. Moreover, consider a simple logarithmic transformation of the time function, i.e. $F = [f_0(t) = 1, f_1(t) = log(t)]$. Consider a full rank model like the one in equation (1) which has six parameters and denote the estimated matrix of coefficients as $\hat{\Theta}$. Also, denote as $\hat{\Theta}^*$ the estimated matrix of coefficients that is given from the Hybrid model (6) by adding $\beta_1 + \beta_2$ to get the first column. The two parametrisations will result in an identical model, with a log-likelihood of -3523.619 and the same estimated coefficients (given below as $\hat{\Theta}$). Next,

Statist. Med. **2014**, 00 1–14
Prepared using **simauth.cls**

Copyright © 2014 John Wiley & Sons, Ltd.

www.sim.org **5**

	FP model		Hybrid Model
Variable	$\hat{\beta}$	**FP(t) power**	$\hat{\beta}$ at $t = 0$
FP_1(Age)	0.0283	-	0.0251
T. Size	0.1859	-	0.2236
Horm. Treatment	-0.5297	-	-0.5186
Horm. Treatment FP_1(t)	0.1211	0	-
FP_1(Nodes)	0.6588	-	0.6023
FP_1(Nodes) FP_1(t)	-0.1248	-2	-
Grade	0.3670	-	0.4411
Grade FP_1(t)	-0.1745	0	-
Chemotherapy	-0.2303	-	-0.0712
Chemotherapy FP_1(t)	0.2834	-2	-
Chemotherapy FP_2(t)	0.1974	-2	-

Table 2. Estimated log hazard ratio $\hat{\beta}$ for variables selected in the final Fractional Polynomial model (FP) for the IASO dataset, along with the same estimates from the Hybrid Model. Outcome of interest is death from breast cancer.

	F(t):	$1 + \log(t)$	F(t):	$1 - 0.5 \times \log(t)$
	0.2042	0.0125	0.2958	-0.0133
$\hat{\Theta} =$	0.8426	-0.1813	0.8198	-0.1749
	0.7559	-0.1613	0.7313	-0.1544

Table 3. Estimates of Θ coefficients for variables tumour size, tumour grade and chemotherapy, using time varying and time fixed effects models with different parametrisations.

assume that one hopes to capture the effect of a rank=1 model by setting $F = [f_0(t) = 1, -0.5 \times f_1(t) = log(t)]$. Fitting a Hybrid model results in the same likelihood, while the estimated parameters ($\hat{\Theta}^*$) can be easily shown to be same as in $\hat{\Theta}$. The estimated coefficients of both parametrizations follow:

$$\hat{\Theta} = \left\{ \begin{array}{cc} 0.2883 & -0.0111 \\ 0.8217 & -0.1754 \\ 0.7333 & -0.1550 \end{array} \right\}, \hat{\Theta}^* = \left\{ \begin{array}{cc} 0.2883 & 0.0222 \\ 0.8217 & 0.3508 \\ 0.7333 & 0.3100 \end{array} \right\}$$

That illustrates that the γ parameter cannot be identified in such a model. The parameters that can be estimated however, are $\beta_1 + \beta_2$ and $\beta_2 + \gamma$.

This changes if one of the variables does not belong to both X_1 and X_2. If tumour size is removed from the list the fitted model comes from equation (5) in which, as shown in Table 3, the choice of γ does matter and can be estimated by maximising the partial likelihood.

Based on this, a suggested procedure for step wise model building has the following steps. At the first step fit all univariate Cox models and rank covariates according to their significance or deviance of the model. The model for the first covariate will compare the deviance from a basic Cox model to that from Cox model with a time varying effect. To keep things simple we consider here an interaction of each covariate with the logarithm of time. For hormonal treatment the deviance of a Cox model was 7013.241 which indicated that the covariate decreased the deviance when compared to a NULL model (see Table 4). When adding a TVE the model improved the deviance by 9.184, leading to the inclusion of a time varying effect for the covariate. In the next step, the number of lymph nodes enters the model as a fixed coefficient first. The inclusion of the variable improved the model. However, a Hybrid model with the number of lymph nodes modelled as dynamic did not lead to a significant improvement. Thus, the best model includes the number of positive lymph nodes as a variable with fixed effects. The procedure is repeated for each covariate. Looking at table 5, variables that should be included both in the X_1 part and X_2 part of model (6) are hormonal treatment (HR) and tumour grading (GR). The model with just two dynamic covariates will be called an optimal Hybrid model (HyRo).

4.5. Comparison

One approach would be to fit all reduced rank models, from rank=1 to the full rank=4 model and chose the one with the optimal AIC. Table 5 presents Akaike's information criterion (AIC) for all reduced rank and Hybrid models. AIC is defined here as the log-likelihood of the model minus the number of parameters. Based on the models' AIC solely, the preferred model should have been a Hybrid model. It gives a better AIC on just 15 degrees of freedom. It has to be noted that when
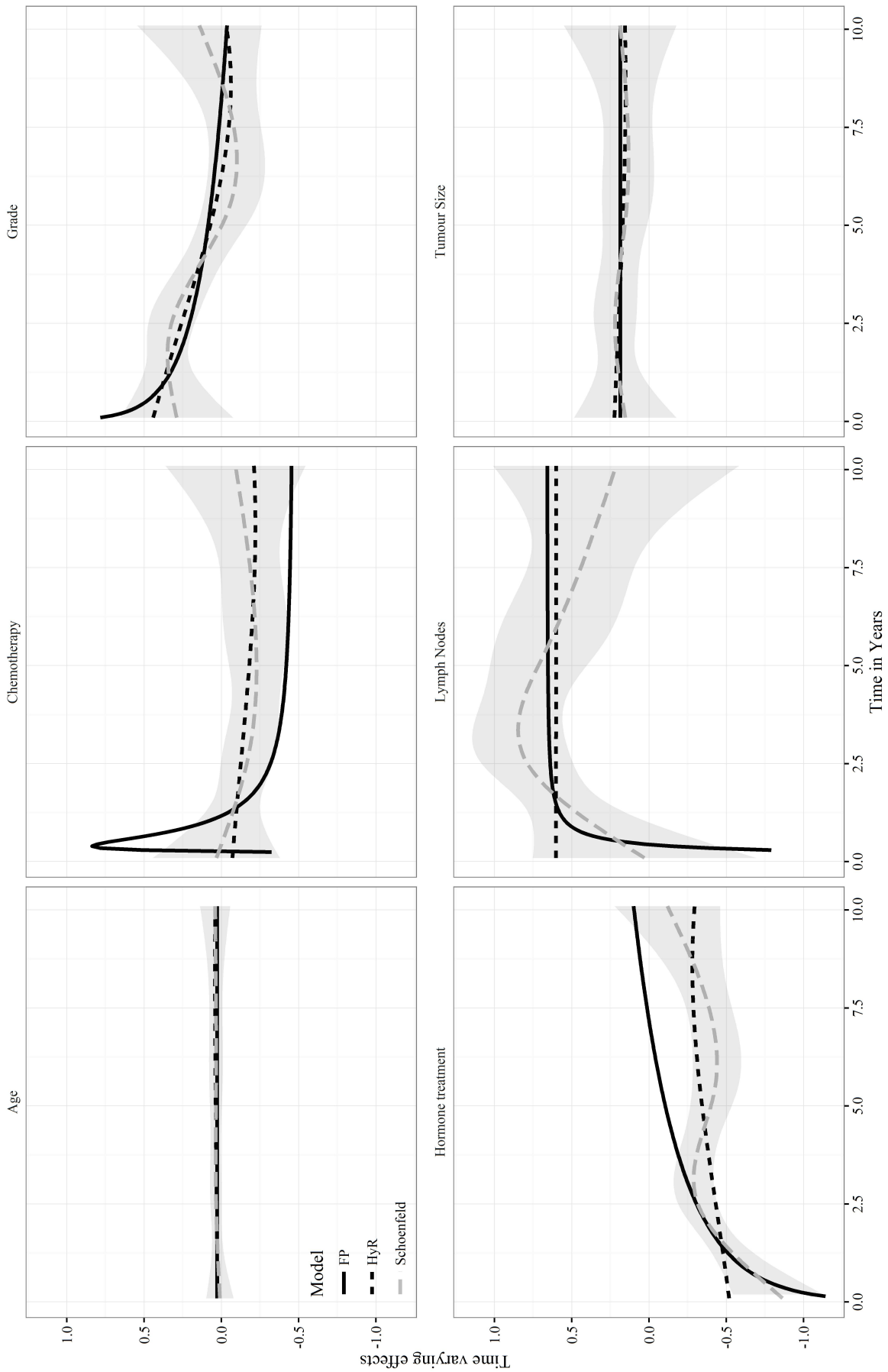
Statistics in Medicine

Figure 1. Time varying effects of all covariates in the IASO dataset, as estimated by Fractional Polynomials (FP, solid black line) and a Hybrid Model (HyR, dotted black line). The dashed grey line present a smooth of Schoenfeld residuals from a Cox Proportional Hazards model with 95% confidence bands given in the shaded gray area.

Table 4. Steps of model building procedure for a Hybrid model. At each line the deviance of the model is compared to the previously "best" model. Fixed refers to a model with the variable only included in the X_1 part of a hybrid model, and HyR to a model with the variable included both in X_1 and X_2 matrices of a Hybrid model.

Variable	Model	Deviance	Difference	P-value
Hormonal treatment	Fixed	7013.241	122.272	0.0000
	HyR	7004.057	9.184	0.0024
Lymph nodes	Fixed	6942.506	61.551	0.0000
	HyR	6942.240	0.266	0.6060
Tumour size	Fixed	6920.708	21.532	0.0000
	HyR	6920.070	0.638	0.4244
Grade	Fixed	6908.424	11.646	0.0006
	HyR	6899.371	9.053	0.0026
Chemotherapy	Fixed	6891.257	8.114	0.0044
	HyR	6890.851	0.406	0.5240
Age	Fixed	6886.735	4.116	0.0425
	HyR	6886.567	0.168	0.6819
Radiotherapy	Fixed	6886.201	0.366	0.5452
	HyR	6885.168	1.033	0.3095

comparing a Hybrid model to the optimal fractional polynomial model, the latter seems to have a better AIC. However, this number is computed by subtracting only eleven parameters of the final model from the likelihood. In reality, it is very difficult to see what are the real degrees of freedom of the model. The optimized HyRo model has a better AIC than the models before (-3452.36, based on 9 degrees of freedom). Note also that, when looking at the Therneau-Grambsch test of proportionality at a 5% significance level, only tumour grading out of the whole dataset is not following the proportionality assumption. The actual fit that is obtained from this optimal model is almost identical to what it is presented in Figure 1.

Model	Parameters	AIC
Rank=1	9	-3618.14
Rank=2	16	-3615.24
Rank=3	21	-3717.24
Rank=4	24	-3457.71
HyR	15	-3455.68
HyRo	9	-3452.36
FPm	11	-3444.86

Table 5. Akaikes Information Criterion on different reduced rank models, Hybrid (HyR) and optimized Hybrid models (HyRo) and a fractional polynomials model (FPm) applied to the IASO breast cancer data. Also given, the number of parameters under each model.

It has already been shown that the two models produce similar estimates for the effects of covariates at the start of the follow up. The fractional polynomials model though, assumes that four covariates have time varying effects, while the Hybrid model assumes that all covariates should be modelled dynamically. In Figure 1 the estimated effects of all covariates under the two different models are presented. As a line of reference a smoothed function of the Schoenfeld residuals as suggested by [19] with 95% confidence intervals (in gray shade) is added to the plot.

Using fractional polynomials, there was no significant TVE for age, thus, the effect remains the same over the whole time period. The Hybrid model does include a TVE for age, however, with the extra parameter for age on β_1 of model (6) the effect looks almost identical to the line fit. The same pattern repeats for the effect of tumour size, where FPm reveals a constant effect and the Hybrid model shows an almost vertical straight line for the effect as well. The two models also give similar patterns for the effect of hormonal treatment and tumour grading. For the latter, although tumour grading does have an effect of the hazard, both models show that this effect washes away towards the end of the time period, at around nine years. Important differences between the two models emerge when looking at the effects of chemotherapy and the number of lymph nodes. In both cases, fractional polynomial functions show an erratic behaviour at the first small values of time. This artefact is corrected by the b-spline functions. For example, the effect of positive nodes appears to be protective at the start of the follow up period under the FPm fit, while on the contrary, the Hybrid model is in agreement

Table 6. Time varying effects and cross-validated partial likelihood of four different models. Also given is a p-value out of a Grambsch-Therneau proportionality test.

Variable	Grambsch-Therneau	Time Varying Effects				
		Rank=1	Rank=4	HyR	HyRo	FPm
Age	0.647	x	x	x		
T. Size	0.591	x	x	x		
Grade	0.002	x	x	x	x	x
Nodes	0.604	x	x	x		x
Horm.Treatment	0.065	x	x	x	x	x
Chemotherapy	0.564	x	x	x		x
cvl		-4063.535	-4009.535	-3990.306	-3984.446	-3969.5

with biological reasoning. It estimates the effect of positive nodes as harmful and resilient over time.

4.5.1. Leave-one-out cross-validation The modelling approach suggested in this work uses extra parameters to account for time varying effects. It is possible that extending a model will explain better the data, but on the other hand, reduce the predictive value of the model as a result of over-fitting. To test how Hybrid models perform, leave-one-out cross validation [29] was used to evaluate the performance of the models. A cross-validated partial likelihood is defined as:

$$cvl = \sum_{i=1}^{n} l_i(\hat{\beta}_{-i})$$

where $l_i(\beta) = l(\beta) - l_{-i}(\beta)$ measures the contribution of observation i to the log-likelihood, and $l_{-i}(\beta)$ is the log-likelihood when observation i is left out. Leave-one out cross validation in practice can measure the ability of a model to predict a new observation, based on a model build when the observation was left out.

A proportional hazards model fitted to the IASO data, using the six covariates as time fixed has a partial likelihood $l(\beta) = -3450.9$, which gives an AIC of -3456.9, not far away from the Hybrid model (Table 5) that had an AIC of -3455.68. Verweij and van Houwelingen argue in their 1993 paper that *AIC is not easy to interpret, because the components of the partial likelihood are not independent*. A look at the cross-validated partial likelihood will show evidence of overfitting for the Hybrid model (Cox model $cvl = -3988.719$ vs Hybrid model $cvl = -3990.306$). That result should have been anticipated, since the Hybrid model used extra parameters even for variables that should be modelled as time fixed, as indicated either visually in Figure 1. However, it should be noted that cvl from a Hybrid model is bigger than what one would have achieved if using a rank=1 model ($cvl = -4063.535$) or in fact a full rank model ($cvl = -4009.448$).

In Section 4.4, the optimal Hybrid model did have a better AIC than all reduced rank models. The cross-validated partial likelihood would be $cvl = -3984.446$ for that model, which does improve on the cvl from the hybrid model. Still, a model based of fractional polynomials does have a better cvl (-3969.5).

Table 6 presents how the cross validated partial likelihood is obtained starting from a model close to rank=1 and moving to the final optimized HyRo model. Cross validation could also be useful as a model building technique. However, it is computationally expensive and not trivial to apply in advanced models.

4.5.2. Brier Scores [27] are often used in survival analysis to measure the predictive performance of a model. A Brier score at a specific time point t measures the difference between true disease outcome, given by the indicator function $I(T_i > t)$, and the predicted probability of an event given by the prediction model. A Brier score is a cross-sectional approach, which means that the measure is evaluated at a specific time point in the follow-up. When comparing different survival models, the one that has the smaller score is the one that provides better predictions. Figure 2 presents Brier scores for four different models, proportional hazards, hybrid, optimized hybrid and a model using fractional polynomial. All models produce similar results. In fact it is very hard to see any differences between different approaches.

Dynamic scores can be more interesting since they show the ability of a model to predict a conditional survival probability, highlighting the errors associated predicting survival up to some point in the future, say in $t + w$, for those alive in time t. This is a sort of a rolling window of size w prediction. The generalisation of a brier score to a dynamic brier score is given in [20], page 46. On the right side of Figure 2 dynamic brier scores are presented for the same four models. Still, is hard to spot any difference for the different approaches. A final approach is to compute the interval specific and total prediction error. Table 7 presents errors obtained for rolling windows of two years for the different models. The Cox proportional hazards models has also been added for a reference. The table shows the reduction in each interval, of the error of each non-proportional hazards model when compared with a single proportional hazards. The last line of the table,
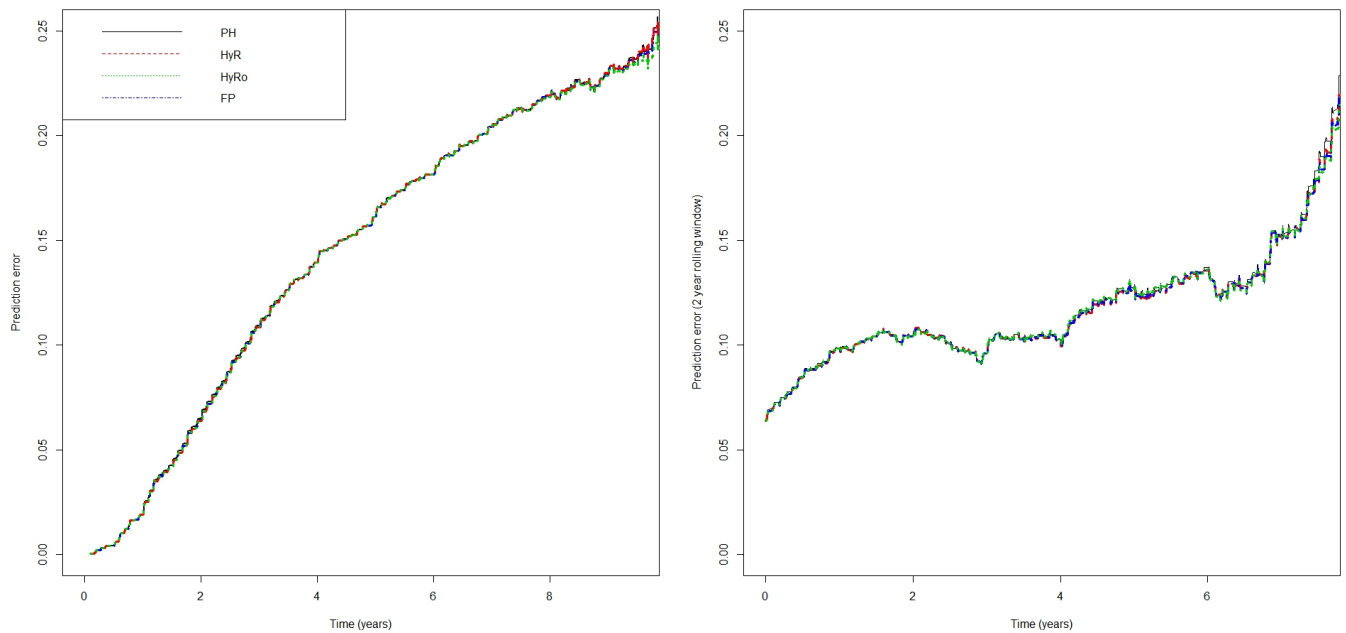
Statist. Med. **2014**, 00 1–14
Prepared using **simauth.cls**

Copyright © 2014 John Wiley & Sons, Ltd.

www.sim.org **9**

Figure 2. Brier scores (left) and dynamic Brier scores (right) for four different models.

Table 7. Interval specific Brier scores, reduction (Rd) and total prediction errors for different models (FPm=fractional polynomials), HyR =Hybrid Model and HyRo= optimal Hybrid Model with variable selection. First column gives the time period and the second shows a Kaplan-Meier survival estimate at the start of each specific interval.

Interval	KMstart	PH	FPm	Rd	HyR	Rd	HyRo	Rd
0-2	1.000	0.0647	0.0640	0.010	0.0639	0.010	0.0641	0.010
2-4	0.923	0.0647	0.0640	0.010	0.0639	0.010	0.0641	0.010
4-6	0.808	0.1040	0.1040	0.000	0.1043	0.000	0.1039	0.000
6-8	0.708	0.1028	0.1021	0.010	0.1019	0.010	0.1027	0.000
8-10	0.591	0.1369	0.1358	0.010	0.1356	0.010	0.1359	0.010
10+	0.405	0.2271	0.2248	0.010	0.2263	0.000	0.2194	0.030
Total	NA	0.4542	0.4507	0.008	0.4511	0.007	0.4492	0.011

labelled total, gives an overall estimate of the error under the model. Overall, the optimized hybrid model has the smallest error, with the biggest reduction in error occuring at the extreme right times of follow up. However, the differences are still subtle and one should be very careful in their interpretation.

5. Simulation

A small simulation study was run to investigate how well a Hybrid model can identify and describe a time varying effect. For each of 1000 simulations a dataset of 600 cases was created, from the following model:

$$h(t|X) = h_0(t) \exp(X_1 \beta_1(t) + X_2 \beta_2 + X_3 \beta_3 + X_4 \beta_4(t) + X_5 \beta_5)$$

where X_1, X_2, X_5 come from a binomial distribution with $p_1 = p_2 = 0.5$ and $p_3 = 0.7$, respectively, and X_3, X_4 come from a standard normal distribution. The fixed effects of the covariates where: $\beta_2 = 0.5, \beta_3 = 0.7$ and $\beta_5 = 0.1$, while the time varying effects where created as: $\beta_1(t) = 0.4 - 0.5 \times \log(t + 0.1)$ and $\beta_4(t) = -0.2 + 0.7 \times \log(t + 0.1)$. The simulation follows a similar design that can be found in chapter 6 of the Ph.D. thesis by A. Buchholz [25]. On each repetition, two different Hybrid models were fitted, one (HyR$_1$) where all five covariates interacted with $F = [f_0(t) = 1, f_1(t) = \log(t)]$, and a second one (HyR$_2$) where all covariates interacted with b-splines as time functions, same as the ones used in the application section. The idea was to see how well the model performs when we assume some knowledge

of the true nature of time varying effect but do not know whether it is needed on all covariates. In the second case, no previous knowledge is assumed over the data. In both cases the fitted model was (6) with $\beta_2 = [\beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{25}]$.

To test for a time varying effect, five different Wald tests were performed on the hypothesis $H_0 : \beta_2 = 0$, for each one of the coefficients in the vector. An effect was termed significant if the p-value was less than 0.01. Over a series of 1000 repetitions, HyR$_1$ correctly identified the time varying effects of X_1, X_4 in 911 and 794 times, respectively, while for covariates with fixed effects it misspecified a TVE in less than 2% of the cases for each variable. The results are comparable with those from a test of proportionality based on the Grambsch-Therneau test at the same significance level using Kaplan-Meier to transform survival time. The results indicate that the model with flexible time functions is slightly more sensitive to identify a time varying effect than the others, while still keeping the number of false positives at reasonable levels (less than 4%, for each time fixed effect).

Variable	PH	HyR$_1$	HyR$_2$
X_1	904	911	907
X_2	19	16	39
X_3	9	8	36
X_4	795	794	811
X_5	14	13	37

Table 8. Frequency of smaller than 0.01 p-values over 1000 simulated datasets, for the presence of a time varying effect using different models. Therneau-Grambsch test was used for the proportional hazards model (PH) and Wald type tests for the two Hybrid models.

In Figure 2 we present the estimated effects for four of the covariates, over each of the iteration, along with the true nature of the relationship and the average fit of the model. It should be pointed out, that a Hybrid model produces a reasonable fit to the data, even when the functional form of the time varying effect is not correctly specified, or indeed is not there at all (for variables X_2, X_3). Both models provide the right average pattern of the effects, whilst the more flexible HyR$_2$ model deviates from reality towards the latter years of follow up, when there are not enough events left in the data.

6. Discussion

In their book Modelling Survival Data [26], page 142, Therneau and Grambsch state then when a test or visual inspection provide evidence of non-proportionality *the first two questions to ask is: "does it matter?" and "is it real?"*. They then argue that often the best strategy is to do nothing. However, if non-proportionality is real then they suggest ways of dealing with the problem (page 145). In this work we have shown how a Hybrid model behaves in a real and a set of simulated datasets. We argue that the model can provide the means both for identifying deviations from proportionality, and for modelling non-proportionality. The aim is not to produce a complex model that will overfit the data and complicate interpretation. Instead, a Hybrid model can be an excellent tool to produce better explanatory models, when deviations from proportionality can be assessed either by using a statistical test or a visual investigation. In the IASO data, when modelling all covariates with TVEs, visual investigation revealed that the effect was "real" in just two of them.

When fitting dynamic models, the time varying factor of the interaction dominates over the covariate values. As such one can only see what happens to the initial effect, i.e. the effect of covariates at time $t = 0$. From that point on, the time varying part is becoming more important and covariate values less important. A Hybrid Reduced Rank model accounts for that by the addition of just one parameter. The addition of a second time fixed coefficient could help to restore balance between the effect of the fixed and the TVE. Adding one extra parameter allows for more flexibility and changes completely the behaviour of the model, since covariates are not assumed to have proportional time varying effects. The extra fixed parameter, can be used also for testing whether the TVE is needed in the model. If the form is correctly specified then the extra term should not be significant. In cases where model parsimony is important, a forward selection type algorithm can be used to end up with a final model that only includes TVEs for some, not all, of covariates.

The new method comes as an extension of a reduced rank model that has less parameters than a full model but it is still flexible enough to account for complex relationships. We presented the basic theory, showed how to estimate the model and how to obtain conditional standard errors. These errors can then be used for hypothesis testing of the significance of the effects. A Hybrid model has the benefit of allowing the use of flexible time functions, such as b-splines, but reducing the number of parameters. The low rank of the model should restrict the time-functions from producing wild behaviours of the effects. At the same time, by introducing an extra parameter for each covariate, it is easy to move away from a very restricted model to a more flexible one. Testing for the significance of covariate effects can then provide a means to choose whether time varying effects are needed for each covariate.
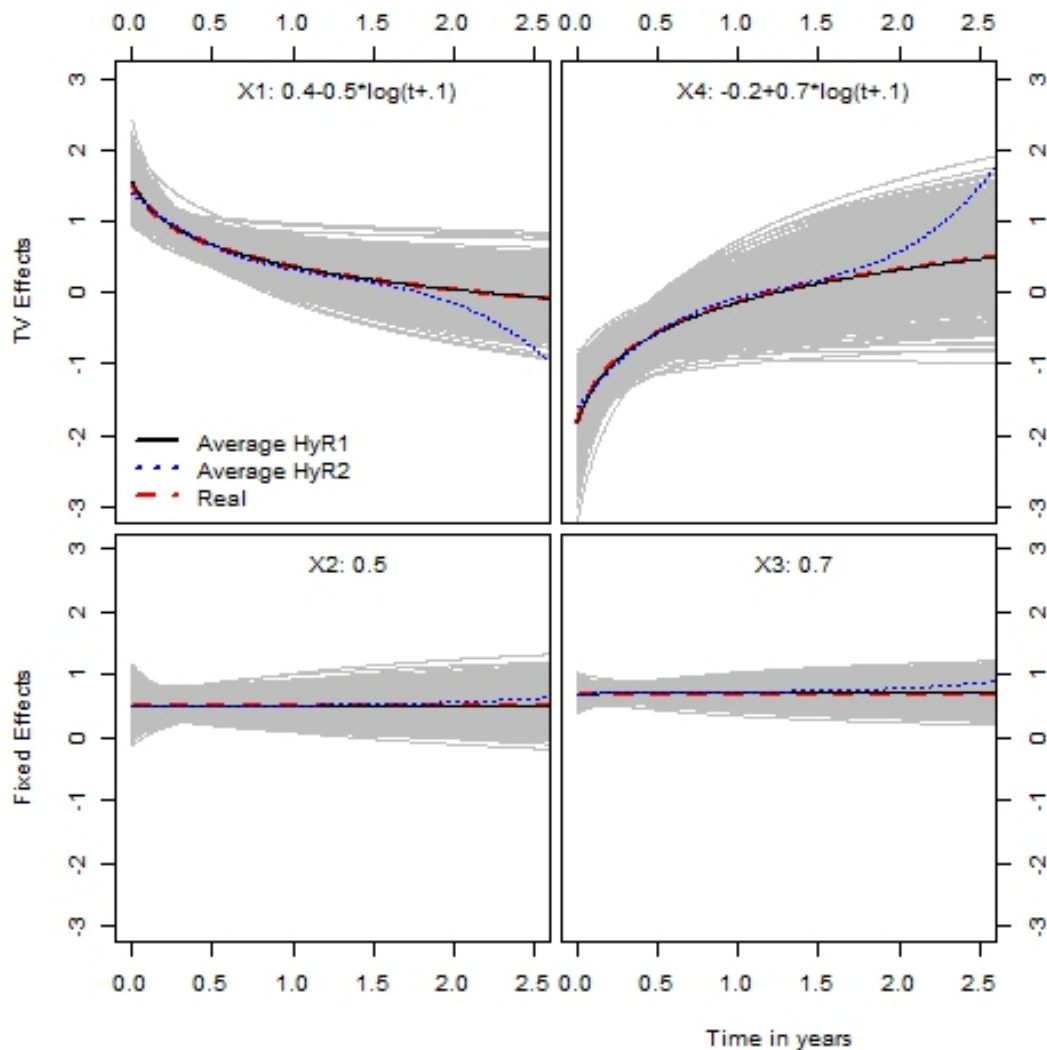
Figure 3. Estimated effects (gray lines) of a Hybrid model with log(t) as time function. The true effect is given in the dashed (red) line, solid black line provides the average fit of HyR$_1$ model, dotted (blue) line shows the average of HyR$_2$ model. Top two graphs present covariates that had a TVE and the bottom two graphs covariates that were created to have fixed effects. Results on 1000 simulated datasets.

The suggested model was applied to a dataset of breast cancer patients and the results were compared to a fit obtained from a fractional polynomial models. In a simple approach, the Hybrid model showed that the patterns that the model gives are similar to those of an "optimal" FPm model. The Hybrid model corrected the wild behaviour of fractional polynomials at times close to zero and provided results that are closer to biological reasoning. The model was also applied to a series of simulated data where it was able to identify which covariates did have TVEs and describe adequately the behaviour of the true coefficients.

To investigate which model is better, we looked into the predictive ability of each model applied. Preliminary analysis revealed that is not easy to distinguish which is the best model, when looking at Brier scores [27] or global Brier scores [20]. van Houwelingen [28] has also argued, that is very difficult to tell the difference in terms of estimating survival probabilities, between different types of flexible models when there is a long follow up and enough events in the study. Both the FPm and HyR models do have enough parameters to allow for a good prediction. A comparison of the predictive ability of the models revealed only minor differences between the models. These results are in agreement with what has been seen before ([20, 28, 24]). As one reviewer stated in an earlier version of the paper, *the fact that overall predictions are nearly the same for two models whose individual components look quite different is an important lesson*. Many researchers tend to over-interpret trends in the data which in many cases might not add anything to the model. Another view would be looking at completely different metrics for model performance, such as the use of cross-validated likelihood [29]. Leave-one-out cross validated likelihood can be used instead of looking at the p-values of each term. However, this is a

computationally expensive method and quite difficult to apply to large datasets.

At this point, it would be difficult to come up with a single recommendation. Initially, we regarded a Hybrid model as a means to identify TVEs, perform hypothesis testing and then conclude to a model that will clearly distinguish which covariates belong to the proportional and which to the non proportional hazards part of the model. A model that does not make that distinction, could lead to overfitting due to the extra parameters. Still, in terms of leave-one-out cross validation, a Hybrid model was better than all reduced rank models. Visual inspection did show which covariates could be modelled as dynamic and which not. From that point on it is up to the researcher to either inspect visually which of the covariates do have time varying effects or start with a rank=1 model and apply the suggested forward type algorithm.

As one of the reviewers pointed out, there is a danger of over-interpretation of the importance of the estimated hazard ratio. In a short discussion paper [30] Hernán noted that one should not endow a hazard ratio with a causal interpretation. In fact, throughout this manuscript the stress in on the pattern, rather than the number itself. What is important here, is the behaviour of a hazard ratio as a pattern in time. In some cases that may be an artefact of the model or the data, and a researcher should be very cautious on the interpretation of such patterns. However, building a model that allows for identification and modelling of these effects is a step towards better understanding complex biological systems. This work focussed on on time varying effects not because they are the most important part of the model building process. Survival estimates and a deep look into the data is always required to get a thorough insight into the data. The emphasis on this paper was given on time varying effects estimation as an important and a necessary block of building good explanatory models.

Acknowledgements

References

1. Putter H, Sasako M, Hartgrink HH, Velde CJH, Houwelingen HC. Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer trial. *Statistics in Medicine* 2005; 2807-2821.
2. Aalen OO. A linear regression model for the analysis of lifetimes. *Statistics in Medicine* 1989; **8**: 907-925.
3. Kooperberg C, Stone CJ, Truong YK. Hazard regression. *Journal of the American Statistical Association* 1995; **90**(429): 78-94
4. Xu R and Adak S. Survival Analysis with Time-Varying Regression Effects Using a Tree-Based Approach. *Biometrics* 2002; **58**(2): 305-315.
5. Gamerman D. Dynamic Bayesian models for survival data. *Applied Statistics* 1991; **40**: 63-79.
6. Martinussen T and Scheike TH. *Dynamic Regression Models for Survival Data*. Springer Science & Business Media: New York, 2006
7. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 1972; **34**: 187-220.
8. Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* 1992; **87**: 942-951
9. Hess KR. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine* 1994; **13**(10): 1045-1062.
10. Bolard P, et al. Assessing time-by-covariate interactions in relative survival models using restrictive cubic spline functions. *Journal of Cancer Epidemiology and Prevention* 2002; **7**(3): 113-122.
11. Kauermann G. Penalized spline smoothing in multivariable survival models with varying coefficients. *Computational Statistics & Data Analysis* 2005; **49**(1): 169-186.
12. Perperoglou A. Cox models with dynamic ridge penalties on time-varying effects of the covariates. *Statistics in Medicine* 2014; **33**(1): 170-180.
13. Verweij PJM. and van Houwelingen HC. Time-dependent effects of fixed covariates in Cox regression. *Biometrics* 1995;**51**(4): 1550-1556.
14. Royston P and Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics* 1994;**43**(3): 429-467.
15. Berger U, Schäfer J, Ulm K. Dynamic Cox modelling based on fractional polynomials: time-variations in gastric cancer prognosis. *Statistics in Medicine* 2003; **22**: 1163-1180.
16. Sauerbrei W, Royston P, Look, M. A new proposal for multivariable modelling of time varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal* 2007; **49**, 453-473.
17. Perperoglou A, le Cessie S, van Houwelingen HC. Reduced-rank hazard regression for modeling non-proportional hazards. *Statistics in Medicine* 2006; **25**(16): 2831-2845.
18. Buchholz A and Sauerbrei W. Comparison of procedures to assess non-linear and time-varying effects in multivariable models for survival data. *Biometrical Journal* 2011;**53**: 308-331.
19. Grambsch PM. & Therneau T. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81**(3): 515-526.
20. van Houwelingen HC and Putter H. *Dynamic Prediction in Clinical Survival Analysis*. *CRC Press, Inc.*, 2011.
21. Perperoglou A, le Cessie S, van Houwelingen HC. A fast routine for fitting cox models with time varying effects of the covariates. *Computer Methods and Programs in Biomedicine* 2006; **81**(2): 154-161.
22. Yee T. *Vector Generalized Linear and Additive Models*. Springer: New York, 2015
23. Perperoglou A. Reduced rank hazard regression with fixed and time-varying effects of the covariates. *Biometrical Journal* 2013; **55**(1): 38-51.

Statist. Med. **2014**, 00 1–14
Prepared using simauth.cls

Copyright © 2014 John Wiley & Sons, Ltd.

www.sim.org   **13**

24. Perperoglou A, Keramopoullos A and van Houwelingen HC. Approaches in modelling long-term survival: an application to breast cancer. *Statistics in Medicine* 2007; 26: 26662685

25. Buchholz A. *Assessment of time-varying long-term effects of therapies and prognostic factors*. PhD thesis, Technische Universität Dortmund, 2010.

26. Therneau TM and Grambsch PM. *Modeling survival data: extending the Cox model*. Springer Science & Business Media: New York, 2000

27. Graf E. *Explained variation measures for survival data*. PhD thesis, University of Freiburg 1998.

28. van Houwelingen HC. Modeling clinical long term survival data. Can we tell the difference between cure models,frailty models or time-dependent effects models?. *Proceedings of the XXth International Biometric Conference* 2000; II: 117-123.

29. Verweij PJM. and van Houwelingen HC. Cross-validation in survival analysis.*Statistics in medicine* 1993; **12**(24): 2305-2314.

30. Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010; **21**(1): 13-15