

# Generalized Logit-Based Proportional Hazards Models and Their Applications in Survival and Reliability Analyses

N. Balakrishnan, M. C. Pardo and M. L. Avendaño

**Abstract** We introduce a flexible family of generalized logit-based regression models for survival and reliability analyses. We present its parametric as well as its semiparametric versions. The method of maximum likelihood and the partial likelihood approach are applied to estimate the parameters of the parametric and semiparametric models, respectively. This new family of models is illustrated with male laryngeal cancer data and compared with Cox regression.

## 1 Introduction

Data arising from survival and reliability analyses often consist of a response variable that measures the duration of time until the occurrence of a specific event and a set of variables (covariates) thought to be associated with the event-time variable. These data arise in a number of applied fields, such as medicine, biology, public health, epidemiology, engineering, economics, and demography, and they have some features that pose difficulties to traditional statistical methods. The first is that the data are generally asymmetrically distributed, while the second feature is that lifetimes are frequently censored (the end-point of interest has not been observed for that individual). Regression models for survival and reliability data have traditionally been based on the proportional hazards model of Cox [3] which is defined through the hazard function  $h(t | \mathbf{x})$  of the form

$$h(t | \mathbf{x}) = h_0(t) \exp(\mathbf{x}'\boldsymbol{\beta}),$$

---

N. Balakrishnan (✉)  
Department of Mathematics and Statistics, McMaster University,  
Hamilton, Ontario, Canada  
e-mail: bala@univmail.cis.mcmaster.ca

M.C. Pardo · M.L. Avendaño  
Department of Statistics and Operational Research I,  
Faculty of Mathematics, Complutense University of Madrid, Madrid, Spain

where  $h_0(t)$  is an arbitrary function of time called baseline hazard function,  $\mathbf{x}' = (x_1, \dots, x_p)$  is a vector of covariates for the individual at time  $t$ , and  $\beta' = (\beta_1, \dots, \beta_p)$  is a vector of unknown parameters to be estimated. In the case when the baseline hazard function is treated nonparametrically, then this model becomes a semiparametric model. Instead, if we assume that the baseline hazard function is specified up to a few unknown parameters, which is usually accomplished with a specific parametric distribution such as Weibull distribution, we obtain a parametric proportional hazards model.

Some recent research has focused on developing extended regression models that include Cox model as a special case. In this line, we can find the model introduced by Etezadi-Amoli and Ciampi [4] of the form

$$h(t | \alpha, \beta, \mathbf{x}) = \exp(\mathbf{x}'\alpha) h_0(t \exp(\mathbf{x}'\beta)),$$

where  $\alpha$  and  $\beta$  are vectors of regression parameters. For  $\beta = \mathbf{0}$ , we deduce the Cox model, while for  $\alpha = \beta$  we obtain the accelerated failure time (AFT) model which is also a popular model in the analysis of survival and reliability data. These authors then show that a better fit is obtained with this new model than with the Cox and AFT models in two examples based on artificial and real data. Nevertheless, the main emphasis of their work is on a spline approximation for the baseline hazard function.

A different family of models with smooth background hazard or survival functions have been proposed by Younes and Lachin [10] and Royston and Parmar [9], which includes the proportional hazards and proportional odds models as special cases. The class of these models is based on transformation of the survival function by a link function  $g(\cdot)$  of the form

$$g(S(t | \beta, \mathbf{x})) = g(S_0(t)) + \mathbf{x}'\beta,$$

where  $S_0(t) = S(t | \mathbf{0}, \mathbf{x})$  is the baseline survival function. The former tackled the estimation problem by using B-splines to estimate the baseline hazard function while the latter utilized natural cubic splines to model  $g(S_0(t))$ . Here, again the main focus of the work was to check the advantage of a smooth modeling of the background hazard or survival functions, respectively.

An alternative model to the Cox model is based on the hazard function

$$h(t | \beta, \mathbf{x}) = h_0(t) \frac{\exp(\mathbf{x}'\beta)}{1 + \exp(\mathbf{x}'\beta)}, \quad (1)$$

where the covariate effects are modeled on the logarithmic scale rather than on the log odds scale. In spite of the simplicity of this model, it has not been studied much in the literature. Recently, MacKenzie [7, 8] has considered this logit link-based model with a constant baseline survival function and nonproportional hazards and displayed its applicability, which is given by

$$h(t | \alpha, \beta, \mathbf{x}) = h_0 \frac{\exp(t\alpha + \mathbf{x}'\beta)}{1 + \exp(t\alpha + \mathbf{x}'\beta)}.$$

The flexibility shown by MacKenzie's model gives us an impetus to extend the Cox model in a similar manner. First of all, to assume that the baseline hazard model is a constant is to limit the flexibility the model. In fact, the aim of the previous papers was to estimate in a proper way the baseline hazard function. Secondly, to measure the influence of the unknown parameters on a generalized log-odds scale instead of a log-odds scale. Therefore, this model is a particular case of our models introduced in Sect. 2 without time-dependence. The reason for not considering time-dependence is to start with a very general family of models but then focus on its simplest form. We hope to consider in our future study time-dependence and also to estimate the background hazard with splines.

In this chapter, we not only study the logit link-based model in (1), but also generalize it to a flexible parametric family of proportional hazards model based on a generalization of the logistic distribution (see Balakrishnan [1]) called Type-I generalized logistic model. The formulation of the model and estimation methods for parametric and semiparametric models are then discussed in Sect. 2. Measures of fitting this model are discussed in Sect. 3. Next, an illustrative example is presented in Sect. 4. Finally, some concluding remarks are made in Sect. 5.

## 2 The Generalized Logit Link Proportional Hazards Model

The logit link-based model in (1) can be generalized by replacing the logistic distribution function in (1) by a generalization of the logistic distribution called Type-I generalized logistic which is given by

$$F(y) = \frac{1}{(1 + e^{-y})^a}, \quad -\infty < y < \infty, \quad a > 0;$$

see Balakrishnan [1].

By utilizing this form, we propose a proportional hazards model defined through the hazard function

$$h(t | \beta, a, \mathbf{x}) = h_0(t) K(\beta, a, \mathbf{x}), \quad (2)$$

with

$$K(\beta, a, \mathbf{x}) = \frac{1}{\{1 + \exp(-\mathbf{x}'\beta)\}^a}, \quad a > 0.$$

For two covariate profiles  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , the hazards are proportional and the relative risk does not depend on  $t$  as

$$\begin{aligned} \rho(t | \beta, a, \mathbf{x}_i, \mathbf{x}_j) &= \frac{h(t | \beta, a, \mathbf{x}_i)}{h(t | \beta, a, \mathbf{x}_j)} \\ &= \left( \frac{1 + \exp(-\mathbf{x}_j' \beta)}{1 + \exp(-\mathbf{x}_i' \beta)} \right)^a. \end{aligned}$$

Note that in the special case when  $a = 1$ , we deduce the proportional hazards model with a logit link function in (1).

The survival function corresponding to the hazard model in (2) is

$$S(t | \beta, a, \mathbf{x}) = \exp\{-K(\beta, a, \mathbf{x}) H_0(t)\}, \quad (3)$$

where  $H_0(t) = \int_0^t h_0(u) du$  is the baseline cumulative hazard function.

Equation (2) characterizes the generalized logit link proportional hazards model with density given by

$$f(t | \beta, a, \mathbf{x}) = \exp\{-H(t | \beta, a, \mathbf{x})\} h(t | \beta, a, \mathbf{x}),$$

where  $H(t | \beta, a, \mathbf{x})$  is the cumulative hazard function.

At this point, we have not made any assumption about the baseline hazard function  $h_0(t)$ , so that the model is parametric only for the covariate effect, and consequently the model is semiparametric. Instead, if we assume a parametric form for the function  $h_0(t)$ , the model becomes parametric. Now, we will describe the statistical inferential methods for both these cases.

## 2.1 Parametric Model

We may assume that the baseline hazard function is specified up to a few unknown parameters. This is usually accomplished with a specific parametric distribution such as the Weibull distribution. In this case, we get a Weibull generalized logit link proportional hazards model WGLPH with hazard function

$$h(t | \lambda, \gamma, \beta, a, \mathbf{x}) = \lambda \gamma t^{\gamma-1} K(\beta, a, \mathbf{x}) \quad (4)$$

which is fully parametric in form. This model contains as a special case the generalized logit link exponential proportional hazards model for the case when  $\gamma = 1$ .

Then, the cumulative hazard is given by

$$H(t | \lambda, \gamma, \beta, a, \mathbf{x}) = \lambda K(\beta, a, \mathbf{x}) t^\gamma$$

and Eq. (4) characterizes the WGLPH with density given by

$$f(t | \lambda, \gamma, \beta, a, \mathbf{x}) = \lambda K(\beta, a, \mathbf{x}) \gamma t^{\gamma-1} \exp\{-\lambda K(\beta, a, \mathbf{x}) t^\gamma\}.$$

Note that this is a Weibull density function with parameters  $\gamma$  and  $\lambda K(\beta, a, \mathbf{x})$ .

When we assume such a fully parametric form for the distribution of survival times, the estimation of the unknown parameters of the model is by full maximum likelihood method. Consider a sample of  $n$  independent individuals with data  $(t_i, \mathbf{x}_i, \delta_i)$ , where  $\delta_i = 1$  for an event and 0 otherwise, for  $i = 1, \dots, n$ . Accordingly, under the assumption that the censoring mechanism is non-informative, the full likelihood for a random sample of  $n$  individuals is given by

$$L(\beta, a) = \prod_{i=1}^n \{h(t_i | \beta, a, \mathbf{x}_i)\}^{\delta_i} \{S(t_i | \beta, a, \mathbf{x}_i)\}.$$

For the Weibull baseline hazard, the log-likelihood function simply becomes

$$\begin{aligned} l(\lambda, \gamma, \beta, a) &= \ln(L(\lambda, \gamma, \beta, a)) \\ &= \sum_{i=1}^n \left[ \delta_i \ln \left\{ \lambda \gamma t_i^{\gamma-1} K(\beta, a, \mathbf{x}_i) \right\} - \lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma \right]. \end{aligned} \quad (5)$$

To obtain the maximum likelihood estimates, the log-likelihood function in (5) must be maximized numerically by using a procedure for constrained optimization. In order to maximize (5), we obtain its first derivatives with respect to all the parameters which are presented in Appendix A.

Observe that in this case, the corresponding survival function can then be estimated as

$$S(t | \hat{\lambda}, \hat{\gamma}, \hat{\beta}, \hat{a}, \mathbf{x}) = \exp \left\{ -K(\hat{\beta}, \hat{a}, \mathbf{x}) \hat{\lambda} t^{\hat{\gamma}} \right\}.$$

## 2.2 Semiparametric Model

On the other hand, when we assume an unknown functional form for the baseline survival function, the estimation of the unknown parameters of the model is done by maximum partial likelihood method. Consider a sample of  $n$  independent individuals with data  $(t_i, \mathbf{x}_i, \delta_i)$  as before and when the censoring mechanism is non-informative. In this case, the partial likelihood for a random sample of  $n$  individuals can be written as

$$L(\beta, a) = \prod_{i=1}^n \left[ \frac{K(\beta, a, \mathbf{x}_i)}{\sum_{l \in R(t_i)} K(\beta, a, \mathbf{x}_l)} \right]^{\delta_i},$$

where  $R(t_i)$  is the risk set at time  $t_i$ .

Then, the partial log-likelihood function is given by

$$\begin{aligned} l(\beta, a) &= \ln(L(\beta, a)) \\ &= \sum_{i=1}^n \delta_i \left[ \ln(K(\beta, a, \mathbf{x}_i)) - \ln \left( \sum_{l \in R(t_i)} K(\beta, a, \mathbf{x}_l) \right) \right]. \end{aligned} \quad (6)$$

For the purpose of maximizing (6) and obtaining the partial maximum likelihood estimates, we use numerical methods for carrying out the required constrained optimization. Its first derivatives with respect to all the parameters are presented in Appendix B.

Once we have fitted a generalized likelihood proportional hazards model, it may be of interest to estimate the survival probability. The estimator of the survival function is based on Breslow's estimator of the baseline cumulative hazard rate, which proceeds as follows:

Let the full likelihood function be

$$\begin{aligned} L(\beta, a, h_0(t)) &= \prod_{i=1}^n \{h_0(t_i) K(\beta, a, \mathbf{x}_i)\}^{\delta_i} \{\exp[-K(\beta, a, \mathbf{x}_i) H_0(t_i)]\} \\ &= \prod_{j=1}^D \{h_0(t_j) K(\beta, a, \mathbf{x}_j)\} \prod_{i=1}^n \{\exp[-K(\beta, a, \mathbf{x}_i) H_0(t_i)]\}, \end{aligned}$$

where  $j = 1, \dots, D$  correspond to the times without censoring. We then obtain

$$L(\beta, a, h_0(t)) = \prod_{j=1}^D h_0(t_j) K(\beta, a, \mathbf{x}_j) \exp \left\{ - \sum_{i=1}^n K(\beta, a, \mathbf{x}_i) H_0(t_i) \right\}.$$

Defining  $H_0(t) = \sum_{t^* < t} h_0(t^*)$ , and supposing that  $\beta$  and  $a$  are fixed, we have

$$L(h_0(t)) = \prod_{j=1}^D h_0(t_j) K(\beta, a, \mathbf{x}_j) \exp \left\{ - \sum_{i=1}^n K(\beta, a, \mathbf{x}_i) \left( \sum_{t_i^* < t_i} h_0(t_i^*) \right) \right\}.$$

Taking  $h_0(t) = 0$  when the event is censored, then we get

$$\begin{aligned}
L(h_0(t)) &= \left[ \prod_{j=1}^D h_0(t_j) K(\beta, a, \mathbf{x}_j) \right] \exp \left\{ - \sum_{j=1}^D h_0(t_j) \sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i) \right\} \\
&= \prod_{j=1}^D h_0(t_j) K(\beta, a, \mathbf{x}_j) \exp \left\{ -h_0(t_j) \sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i) \right\},
\end{aligned}$$

so that

$$L(h_0(t_1), \dots, h_0(t_D)) \propto \prod_{j=1}^D h_0(t_j) \exp \left\{ -h_0(t_j) \sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i) \right\}.$$

For the determination of the maximum likelihood estimate, we take the derivative with respect to  $h_0(t_j)$  which is given by

$$\begin{aligned}
\frac{\partial L}{\partial h_0(t_j)} &= \exp \left\{ -h_0(t_j) \sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i) \right\} \\
&\quad + h_0(t_j) \exp \left\{ -h_0(t_j) \sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i) \right\} \left\{ - \sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i) \right\}.
\end{aligned}$$

Upon equating this to zero, we obtain the maximum likelihood estimate to be

$$\hat{h}_0(t_j) = \frac{1}{\sum_{i \in R(t_j)} K(\beta, a, \mathbf{x}_i)},$$

and consequently

$$\hat{H}_0(t) = \sum_{t^* < t} \frac{1}{w(t^*)},$$

where  $w(t^*) = \sum_{l \in R(t^*)} K(\beta, a, \mathbf{x}_l)$ .

Since

$$\hat{S}_0(t) = \exp \left[ -\hat{H}_0(t) \right]$$

is the estimator of the survival function of an individual with covariate vector  $\mathbf{x} = \mathbf{0}$ , for estimating the survival function of an individual with covariate vector  $\mathbf{x} = \mathbf{x}^*$ , we use the estimator

$$\hat{S}(t | \hat{\beta}, \hat{a}, \mathbf{x} = \mathbf{x}^*) = \left\{ \hat{S}_0(t) \right\}^{K(\hat{\beta}, \hat{a}, \mathbf{x} = \mathbf{x}^*)}. \quad (7)$$

In both cases of parametric and semiparametric setting, the variance of the estimated parameters  $\hat{\theta} = (\hat{\lambda}, \hat{\gamma}, \hat{\beta}, \hat{a})$  and  $\hat{\theta} = (\hat{\beta}, \hat{a})$  obtained by maximizing Eqs. (5) and (6), respectively, can be estimated as

$$\text{var}(\hat{\theta}) = \text{diag}(I^{-1}(\hat{\theta})),$$

where  $I$  is the observed information matrix.

### 3 Measures of Fit

After fitting several possible models for a given data, we will need to compare the fit of each model for selecting the best one. When we fit several non-nested models, we may use the Akaike information criterion (AIC) to choose the best one among them. The AIC is defined as

$$-2l(\hat{\theta}) + 2(\text{number of model parameters}).$$

Essentially, we compare the AIC scores for different models and then select the one with the smallest AIC score.

Another popular criterion for model selection among parametric models is the Bayesian information criterion (BIC). The BIC is given by

$$-2l(\hat{\theta}) + (\text{number of model parameters}) \log(\text{sample size}),$$

and in the same way as with AIC scores, we select the one with the smallest BIC value.

On the other hand, to describe how well a model fits the observed data, we can do tests of goodness-of-fit for the estimated survival function. Such tests summarize the discrepancy between observed values and the expected values for the survival function under the model. We will use two well-known statistics for this purpose, the first one is the lack of fit sum of squares (SS) given by

$$\sum_i (\text{observed value}_i - \text{fitted value}_i)^2,$$

and the second is the Kolmogorov–Smirnov statistic (KS) defined as

$$\max_i |\text{observed value}_i - \text{fitted value}_i|.$$



## 4 Numerical Illustration

We illustrate the use of the family of proposed models by analyzing *death times of male laryngeal cancer patients*. Kardaun [5] reported data on 90 males diagnosed with cancer of the larynx during the period 1970–1978 at a Dutch hospital. Times recorded were the intervals (in years) between the first treatment and either death or the end of the study. Also recorded were the patient’s age at the time of diagnosis and the stage of the patient’s cancer, wherein the stage is a factor of four levels. The larynx data have been used by Klein and Moeschberger [6] to illustrate some techniques in survival analysis. The larynx data can be obtained from the `MKSurv` Package of the R software package.

### 4.1 Fit of a Fully Parametric Model

First, we fit the fully parametric proportional hazards model by means of three specific models, namely, the Weibull proportional hazards model (WPH), the Weibull logit link proportional hazards model (WLPH), and the WGLPH. To get a good fit of these models, we do a grid  $\{0.5, 1.0, \dots, 4.5, 5.0\} \times \{0.5, 1.0, \dots, 4.5, 5.0\}$  for the initial values of the parameters  $\lambda$  and  $\gamma$  for the required maximization of the three specified models. Moreover,  $a = 1$  is used as the initial value for the parameter  $a$  of the WGLPH model.

In the fitting of the WGLPH model, we found that the numerical methods, used to determining the maximum, in fact, find local maxima that can be far away from the global maximum. We therefore adopt a profile full log-likelihood method to solve this problem.

Let the profile log-likelihood function be

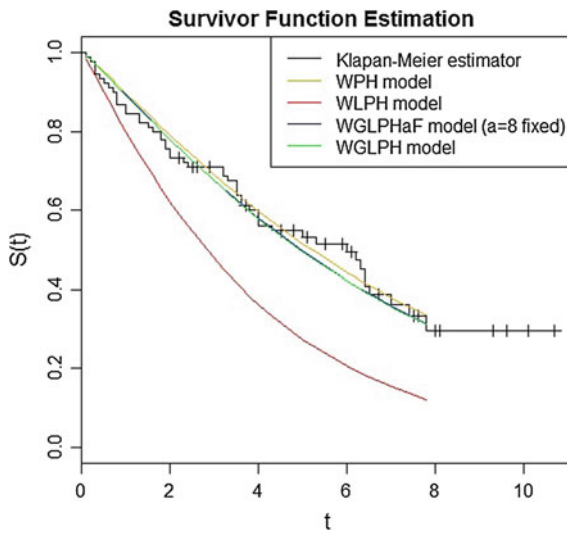
$$p l(a) = \sup_{\lambda, \gamma, \beta} l(\lambda, \gamma, \beta, a).$$

We then calculate this function for  $a \in \{5.0, 6.0, 7.0, 8.0, 9.0\}$  and the same grid for initial points of  $\lambda$  and  $\gamma$  as mentioned earlier. Finally, we found the value of  $a = 8$  to be the choice of  $a$  that maximized the profile full log-likelihood (WGLPHaF). The parameter estimates (Est) and their standard errors (SE) so obtained for the four fitted models are presented in Table 1.

Now, we compare the empirical survival function with the estimated survival function for each of the four fitted models. The empirical survival function was calculated by Kaplan–Meier estimator, while the survival function of the WLPH, WGLPHaF, and WGLPH models were estimated with the corresponding function (3) of the parametric model. In Fig. 1, we have shown a plot of these estimated survival functions. To assess the goodness-of-fit of these models, we calculated the SS and KS statistics to compare the empirical survival function with the corresponding estimated survival function for all four models (WPH, WLPH, WGLPHaF, and WGLPH).

**Table 1** Fit of the four fully parametric models

Variable	WPH		WLPH		WGLPHaF ( $a = 8$ fixed)		WGLPH	
	Est	SE	Est	SE	Est	SE	Est	SE
$\lambda$	0.01855	0.01893	0.32468	0.11235	3.16696	4.94242	2.56534	6.59244
$\gamma$	1.13020	0.13843	1.09364	0.13413	1.13038	0.13832	1.13034	0.13847
Age	0.01973	0.01420	-0.01638	0.00810	0.00740	0.00751	0.00785	0.00755
Stage II	0.16633	0.46111	-0.04115	0.60885	0.04320	0.15510	0.04406	0.16261
Stage III	0.66255	0.35545	0.75535	0.56191	0.23229	0.16294	0.24550	0.21569
Stage IV	1.74530	0.41471	11.26756	96.80031	0.73511	0.41099	0.79024	0.75908
$a$	—	—	—	—	—	—	7.74186	3.94820



**Fig. 1** Survival functions of the four fitted parametric models and the Kaplan–Meier estimator

In Table 2, the values of the SS and KS statistics obtained for these four fitted models are presented.

Also, in Table 3, two comparisons of these four models are made based on the (AIC) and the BIC.

All these results show that the WGLPH, with the parameter  $a$  determined by the profile likelihood method, provides overall the best fitting model for the considered data.

**Table 2** Goodness-of-fit statistics for the four fitted fully parametric models

	WPH	WLPH	WGLPHaF	WGLPH
SS	0.035	1.034	0.034	0.034
KS	0.057	0.289	0.071	0.073

**Table 3** AIC and BIC values for the four fitted parametric models

	WPH	WLPH	WGLPHaF	WGLPH
AIC	294.8468	300.4143	294.7641	296.7652
BIC	309.8456	315.4131	309.7630	314.2639

### 4.2 Fit of a Semiparametric Model

Now, we fit the proportional hazards model (PH), the logit link proportional hazards model (LPH), and the generalized logit link proportional hazards model (GLPH) in a semiparametric framework. In order to get the best fit for the GLPH model, we do a grid for the initial value of the parameter  $a$  in  $\{0.5, 1.0, 1.5, 2.0\}$  and we look for the maximum value of the partial log-likelihood function in (6).

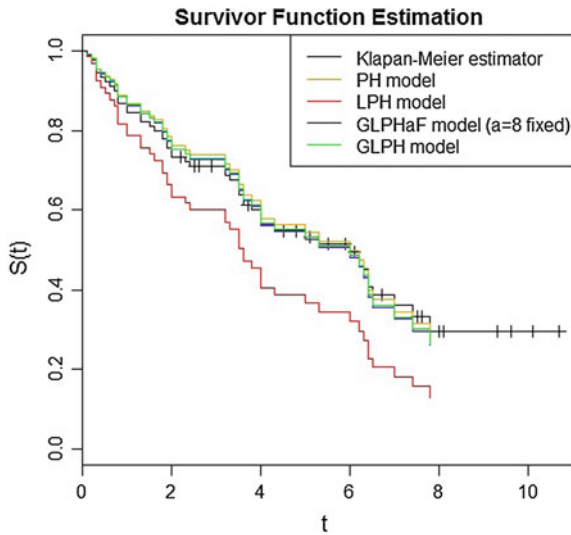
As in the parametric case, we consider a profile partial log-likelihood method maximizing the function

$$p l(a) = \sup_{\beta} l(\beta, a).$$

We calculate this function for  $a \in \{1, 2, \dots, 8, 9, 10\}$  in GLPH with  $a$  fixed and we get  $a = 8$  that maximizes the profile partial log-likelihood function (GLPHaF). In Table 4, we present the parameter estimates (Est) and their standard errors (SE) for the four fitted models. Note that we obtain similar estimation for GLPHaF ( $a = 8$  fixed) and GLPH models, but we reduce substantially the SE in GLPHaF, as the parameter  $a$  is fixed in this case. Furthermore, we compare the empirical survival function and the estimated survival function for each of the four models. The empirical survival function was calculated by Kaplan–Meier estimator, the Breslow’s estimator was used to estimate the survival function in Cox PH model, while the survival function of LPH, GLPHaF, and GLPH models were estimated with the corresponding function  $S(t)$  of the semiparametric model in (7). In Fig. 2, we have presented a plot of these estimated survival functions. In Table 5 we present the values of the SS and KS statistics to compare the empirical survival function and the estimated survival function for each of the four fitted models (PH, LPH, GLPHaF and GLPH).

**Table 4** Fit of the four semiparametric models

Variable	PH		LPH		GLPHaF ( $a = 8$ fixed)		GLPH	
	Est	SE	Est	SE	Est	SE	Est	SE
Age	0.01890	0.01425	-0.01568	0.00828	0.00570	0.00565	0.00574	0.02170
Stage II	0.13856	0.46231	-0.07464	0.61227	0.03047	0.12829	0.03060	0.14502
Stage III	0.63835	0.35608	0.75403	0.57134	0.18337	0.12140	0.18453	0.67040
Stage IV	1.69306	0.42221	7.93424	19.34429	0.55128	0.23583	0.55537	2.35490
$a$	—	—	—	—	—	—	8.95955	22.95138



**Fig. 2** Survival functions of the four fitted semiparametric models and the Kaplan–Meier estimator

**Table 5** Goodness-of-fit statistics for the four fitted semiparametric models

	PH	LPH	GLPHaF	GLPH
SS	0.014	0.517	0.008	0.011
KS	0.029	0.182	0.033	0.033

**Table 6** AIC and BIC values for the four fitted semiparametric models

	PH	LPH	GLPHaF	GLPH
AIC	384.3589	389.2912	384.3173	386.3173
BIC	394.3581	399.2905	394.3165	398.8163

In Table 6, the comparisons of the four fitted models are made based on the AIC and the BIC. From all these results, we draw the general conclusion that the generalized logit-link proportional hazards models are good competitors for the Cox model.

## 5 Final Remarks

The family of proposed models is quite flexible and seems to provide a good competitor for the Cox model. We derive the likelihood function and the partial likelihood function for the parametric and semiparametric models, respectively, for obtaining the parameter estimates and their standard errors. The estimated survival

function of a member of the proposed family of models fits the empirical survival function better than the Cox model. Furthermore, this generalized logit-based proportional hazards model is the one with minimum AIC and BIC.

There are still some unresolved issues in this regard. First of all, the asymptotic properties of the parameter estimates have to be established, and associated statistical inferential issues need to be studied in detail. Another problem of interest is to introduce time-dependence in the models. In this case, the proposed models will provide an extension of [7, 8] models in two ways, with one coming from the non-constant hazard function and the other arising from the generalized logit-link function.

## Appendix A: Derivatives of the Log-Likelihood Function in a Fully Parametric Model

To maximize (5), we obtain its first derivatives with respect to all the parameters as

$$\begin{aligned}\frac{\partial l(\lambda, \gamma, \beta, a)}{\partial \lambda} &= \sum_{i=1}^n \left[ \frac{\delta_i}{\lambda} - K(\beta, a, \mathbf{x}_i) t_i^\gamma \right], \\ \frac{\partial l(\lambda, \gamma, \beta, a)}{\partial \gamma} &= \sum_{i=1}^n \left[ \delta_i \left( \frac{1}{\gamma} + \ln(t_i) \right) - \lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma \ln(t_i) \right], \\ \frac{\partial l(\lambda, \gamma, \beta, a)}{\partial \beta_j} &= \sum_{i=1}^n \left[ \frac{a x_{ij} (\delta_i - \lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma)}{1 + \exp(\mathbf{x}'_i \beta)} \right],\end{aligned}$$

for  $j = 1, \dots, p$ , and

$$\frac{\partial l(\lambda, \gamma, \beta, a)}{\partial a} = \sum_{i=1}^n \left[ \ln(1 + \exp(-\mathbf{x}'_i \beta)) (\lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma - \delta_i) \right].$$

To obtain the corresponding information matrix  $I(\lambda, \gamma, \beta, a)$ , we need the Hessian matrix  $H(\lambda, \gamma, \beta, a)$  which is the matrix of second derivatives of the log-likelihood function in (5) with respect to its parameters.

We obtain them readily as follows:

$$\begin{aligned}\frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \lambda^2} &= -\frac{1}{\lambda^2} \left[ \sum_{i=1}^n \delta_i \right], \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \lambda \partial \gamma} &= -\sum_{i=1}^n K(\beta, a, \mathbf{x}_i) t_i^\gamma \ln(t_i),\end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \lambda \partial \beta_j} &= -a \sum_{i=1}^n \frac{K(\beta, a, \mathbf{x}_i) t_i^\gamma x_{ij}}{1 + \exp(\mathbf{x}'_i \beta)}, \quad \text{with } j = 1, \dots, p, \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \lambda \partial a} &= \sum_{i=1}^n K(\beta, a, \mathbf{x}_i) t_i^\gamma \ln(1 + \exp(-\mathbf{x}'_i \beta)), \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \gamma^2} &= \sum_{i=1}^n \left[ -\frac{\delta_i}{\gamma^2} - \lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma (\ln(t_i))^2 \right], \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \gamma \partial \beta_j} &= -a \lambda \sum_{i=1}^n \frac{K(\beta, a, \mathbf{x}_i) t_i^\gamma \ln(t_i) x_{ij}}{1 + \exp(\mathbf{x}'_i \beta)}, \quad \text{with } j = 1, \dots, p, \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \gamma \partial a} &= \lambda \sum_{i=1}^n K(\beta, a, \mathbf{x}_i) t_i^\gamma \ln(t_i) \ln(1 + \exp(-\mathbf{x}'_i \beta)), \end{aligned}$$

for  $j = 1, \dots, p$  and  $l = j, \dots, p$ ,

$$\begin{aligned} \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \beta_j \partial \beta_l} &= -a \sum_{i=1}^n \\ &\quad \times \left[ \frac{\delta_i \exp(\mathbf{x}'_i \beta) x_{ij} x_{il} + \lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma x_{ij} x_{il} (a - \exp(\mathbf{x}'_i \beta))}{(1 + \exp(\mathbf{x}'_i \beta))^2} \right], \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial \beta_j \partial a} &= \sum_{i=1}^n \left[ \frac{\delta_i x_{ij} - \lambda K(\beta, a, \mathbf{x}_i) t_i^\gamma x_{ij} \{1 - a \ln(1 + \exp(-\mathbf{x}'_i \beta))\}}{1 + \exp(\mathbf{x}'_i \beta)} \right], \\ \frac{\partial^2 l(\lambda, \gamma, \beta, a)}{\partial a^2} &= -\lambda \sum_{i=1}^n K(\beta, a, \mathbf{x}_i) t_i^\gamma \{ \ln(1 + \exp(-\mathbf{x}'_i \beta)) \}^2. \end{aligned}$$

## Appendix B: Derivatives of the Partial Log-Likelihood Function in the Semiparametric Model

To maximize (6), we obtain its first derivatives with respect to all the parameters as

$$\frac{\partial l(\beta, a)}{\partial \beta_j} = a \sum_{i=1}^n \delta_i \left[ \frac{x_{ij}}{1 + \exp(\mathbf{x}'_i \beta)} - \frac{1}{\sum_{l \in R(t_i)} K(\beta, a, \mathbf{x}_l)} \sum_{l \in R(t_i)} \frac{x_{lj} K(\beta, a, \mathbf{x}_l)}{1 + \exp(\mathbf{x}'_l \beta)} \right]$$

for  $j = 1, \dots, p$ , and

$$\frac{\partial l(\boldsymbol{\beta}, a)}{\partial a} = \sum_{i=1}^n \delta_i \left[ -\ln(1 + \exp(-\mathbf{x}'_i \boldsymbol{\beta})) + \frac{1}{\sum_{l \in R(t_i)} K(\boldsymbol{\beta}, a, \mathbf{x}_l)} \sum_{l \in R(t_i)} K(\boldsymbol{\beta}, a, \mathbf{x}_l) \ln(1 + \exp(-\mathbf{x}'_l \boldsymbol{\beta})) \right].$$

Now, to obtain the information matrix, we need the second derivatives of the partial log-likelihood function (6) for  $j = 1, \dots, p$  and  $m = j, \dots, p$ , which are as follows:

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\beta}, a)}{\partial \beta_j \partial \beta_m} &= a \sum_{i=1}^n \delta_i \left[ \frac{-\exp(\mathbf{x}'_i \boldsymbol{\beta}) x_{ij} x_{im}}{(1 + \exp(\mathbf{x}'_i \boldsymbol{\beta}))^2} - \frac{a v_m(t_i) v_j(t_i) - w(t_i) y_{jm}(t_i)}{(w(t_i))^2} \right], \\ \frac{\partial^2 l(\boldsymbol{\beta}, a)}{\partial \beta_j \partial a} &= \sum_{i=1}^n \delta_i \left[ \frac{x_{ij}}{1 + \exp(\mathbf{x}'_i \boldsymbol{\beta})} - \frac{v_j(t_i)}{w(t_i)} - a \frac{v_j(t_i) u(t_i)}{(w(t_i))^2} + a \frac{z_j(t_i)}{w(t_i)} \right], \\ \frac{\partial^2 l(\boldsymbol{\beta}, a)}{\partial a^2} &= \sum_{i=1}^n \delta_i \left[ \frac{-(u(t_i))^2}{(w(t_i))^2} + \frac{\sum_{l \in R(t_i)} K(\boldsymbol{\beta}, a, \mathbf{x}_l) \{\ln(1 + \exp(-\mathbf{x}'_l \boldsymbol{\beta}))\}^2}{w(t_i)} \right], \end{aligned}$$

where

$$\begin{aligned} w(t) &= \sum_{l \in R(t)} K(\boldsymbol{\beta}, a, \mathbf{x}_l), \\ v_j(t) &= \sum_{l \in R(t)} \frac{K(\boldsymbol{\beta}, a, \mathbf{x}_l) x_{lj}}{1 + \exp(\mathbf{x}'_l \boldsymbol{\beta})}, \\ y_{jm}(t) &= \sum_{l \in R(t)} \frac{K(\boldsymbol{\beta}, a, \mathbf{x}_l) x_{lj} x_{lm} (a - \exp(\mathbf{x}'_l \boldsymbol{\beta}))}{(1 + \exp(\mathbf{x}'_l \boldsymbol{\beta}))^2}, \\ u(t) &= \sum_{l \in R(t)} K(\boldsymbol{\beta}, a, \mathbf{x}_l) \ln(1 + \exp(-\mathbf{x}'_l \boldsymbol{\beta})), \\ z_j(t) &= \sum_{l \in R(t)} \frac{K(\boldsymbol{\beta}, a, \mathbf{x}_l) x_{lj} \{\ln(1 + \exp(-\mathbf{x}'_l \boldsymbol{\beta}))\}}{1 + \exp(\mathbf{x}'_l \boldsymbol{\beta})}. \end{aligned}$$

**Acknowledgments** This work was supported by research grant MTM2009-06997 and MAEC-AECID.

## References

1. Balakrishnan N (ed) (1992) Handbook of the Logistic Distribution. Marcel Dekker, New York
2. Cheng SC, Wei LJ, Ying Z (1995) Analysis of transformation models with censored data. *Biometrika* 82:835–845
3. Cox DR (1972) Regression models and life-tables. *J Roy Stat Soc B* 34(2):187–220
4. Etezadi-Amoli J, Ciampi A (1987) Extended hazard regression for censored survival data with covariates: A spline approximation for the background hazard function. *Biometrics* 43:181–192
5. Kardaun O (1983) Statistical analysis of male larynx-cancer patients: A case study. *Stat Nederlandica* 37:103–126
6. Klein JP, Moeschberger ML (1997) Survival analysis: techniques for censored and truncated data. Springer-Verlag, New York
7. MacKenzie G (1996) Regression models for survival data: The generalized time-dependent logistic family. *Statistician* 45:21–34
8. MacKenzie G (1997) On a non-proportional hazards regression model for repeated medical random counts. *Stat Med* 16:1831–1843
9. Royston P, Parmar MKB (2002) Flexible parametric proportional hazards and proportional odds models for censored survival analysis, with application to prognostic modelling and estimation of treatment effects. *Stat Med* 21:2175–2197
10. Younes N, Lachin J (1997) Link-based models for survival data with interval and continuous time censoring. *Biometrics* 53:1199–1211