Adaptive Conditional Hazard Regression Modeling of Multiple Event Times

George J. Knafl

School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Email: gknafl@unc.edu

Abstract

Recurrent event time data and more general multiple event time data are commonly analyzed using extensions of Cox regression, or proportional hazards regression, as used with single event time data. These methods treat covariates, either time-invariant or time-varying, as having multiplicative effects while general dependence on time is left un-estimated. An adaptive approach is formulated for analyzing multiple event time data. Conditional hazard rates are modeled in terms of dependence on both time and covariates using fractional polynomials restricted so that the conditional hazard rates are positive-valued and so that excess time probability functions (generalizing survival functions for single event times) are decreasing. Maximum likelihood is used to estimate parameters adjusting for right censored event times. Likelihood cross-validation (LCV) scores are used to compare models. Adaptive searches through alternate conditional hazard rate models are controlled by LCV scores combined with tolerance parameters. These searches identify effective models for the underlying multiple event time data. Conditional hazard regression is demonstrated using data on times between tumor recurrence for bladder cancer patients. Analyses of theory-based models for these data using extensions of Cox regression provide conflicting results on effects to treatment group and the initial number of tumors. On the other hand, fractional polynomial analyses of these theory-based models provide consistent results identifying significant effects to treatment group and initial number of tumors using both model-based and robust empirical tests. Adaptive analyses further identify distinct moderation by group of the effect of tumor order and an additive effect to group after controlling for nonlinear effects to initial number of tumors and tumor order. Results of example analyses indicate that adaptive conditional hazard rate modeling can generate useful insights into multiple event time data.

Keywords

Adaptive Regression, Fractional Polynomials, Hazard Rate, Multiple Event

Times, Recurrent Events

1. Introduction

Cox regression, also called proportional hazard regression, is the commonly used approach for modeling single survival time, failure time, or time-to-event data (see [1] for an extensive overview). The hazard function is assumed to equal a product of an unspecified function of time and the exponent of a linear function of available covariates with estimation addressing only the slope parameters for the covariates while ignoring the underlying dependence of the hazard function on time. Extensions of Cox regression have been developed to handle multiple event times. For the case of recurrent events, Andersen and Gill [2] model the hazard rate (or intensity function) in terms of possibly time-varying covariates while Lin *et al.* [3] model proportional mean and rate functions for the underlying counting process in terms of possibly time-varying covariates. For the more general case of multiple event times, Prentice *et al.* [4] use marginal Cox regression models, one for each possible event, to model possibly time-varying covariates. Wei *et al.* [5] propose an alternative marginal Cox regression approach. What these approaches have in common is that covariates have multiplicative effects while general dependence on time is left un-estimated. The purpose of this article is to formulate and demonstrate an adaptive approach for conditional hazard regression modeling of multiple event times data based on fractional polynomials [6] [7] using an extension of the approach of Knafl [8] for modeling single event time data. This approach estimates the dependence of the conditional hazard rate on current time values, prior time values, and event time order as well as on possibly time-varying covariates.

2. Methods

2.1. Multiple Event Times

Let $T_m > 0$ for $1 \le m \le M$ denote M continuous, possibly correlated random event time variables. Also, let C_m for $1 \le m \le M$ denote associated random indicators for whether T_m is right censored, that is, only a lower bound for T_m is observable ($C_m = 1$), or whether T_m is not censored, that is, an actual value for T_m is observable ($C_m = 0$). Furthermore, let $\boldsymbol{Z}_m$ for $1 \le m \le M$ be random $r \times 1$ covariate vectors consisting of random covariates $Z_{i,m}$ for $1 \le i \le r$ and $1 \le m \le M$. This formulation allows for time-varying covariates. In practice, many covariates will be time-invariant with $Z_{i,m} = Z_i$ for $1 \le m \le M$. One natural time-varying covariate is event order with values equal to the event time index m.

For $1 \le m \le M$, let

$$\boldsymbol{W}_m = \begin{pmatrix} T_1 & \cdots & T_m & C_1 & \cdots & C_m & \boldsymbol{Z}_1^{\mathrm{T}} & \cdots & \boldsymbol{Z}_m^{\mathrm{T}} \end{pmatrix}^{\mathrm{T}}$$

be the $\left(m\cdot(2+r)\right)\times 1$ random vector of measurements up to time m where $\mathbf{Z}^{\mathrm{T}}$ denotes the transpose of $\mathbf{Z}$. Also, let t_m, c_m, and z_m denote values associated with the random variables/vectors T_m, C_m, and $\mathbf{Z}_m$ and use these to form the associated vector values

$$\boldsymbol{w}_m = \left(t_1 \quad \cdots \quad t_m \quad c_1 \quad \cdots \quad c_m \quad z_1^{\mathrm{T}} \quad \cdots \quad z_m^{\mathrm{T}} \right)^{\mathrm{T}}$$

for $\boldsymbol{W}_m$. Note that t_m is an actual value for T_m when $c_m = 0$, that is, $T_m = t_m$, and a lower bound on T_m when $c_m = 1$, that is, $T_m > t_m$.

For $1 \le m \le M$, denote the conditional distribution function

$$F\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = P\left(T_m \le t_m \mid \boldsymbol{W}_{m-1} = \boldsymbol{w}_{m-1}, \boldsymbol{Z}_m = \boldsymbol{z}_m\right)$$

For $t_m > 0$ where $\boldsymbol{W}_0 = \mathbf{0}$ and $\boldsymbol{w}_0 = \mathbf{0}$ are $\left(m\cdot(2+r)\right)\times 1$ zero-valued vectors so that

$$F\left(t_1; \boldsymbol{w}_0, \boldsymbol{z}_1\right) = P\left(T_1 \le t_1 \mid \boldsymbol{Z}_1 = \boldsymbol{z}_1\right).$$

for time-invariant covariates with $Z_{i,m} = Z_i$, the dependence of T_m on $\boldsymbol{W}_{m-1} = \boldsymbol{w}_{m-1}$ and $\boldsymbol{Z}_m = \boldsymbol{z}_m$ is simpler than in general since it depends only on the single value z_i. When $1 \le m \le M$, the associated excess time probability function is

$$S\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = 1 - F\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = P\left(T_m > t_m \mid \boldsymbol{W}_{m-1} = \boldsymbol{w}_{m-1}, \boldsymbol{Z}_m = \boldsymbol{z}_m\right)$$

(generalizing the survival function for single time event data) and the associated conditional hazard rate function is

$$\lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = \lim_{\Delta t \downarrow 0} \frac{P\left(t_m < T_m \le t_m + \Delta t \mid T_m > t_m, \boldsymbol{W}_{m-1} = \boldsymbol{w}_{m-1}, \boldsymbol{Z}_m = \boldsymbol{z}_m\right)}{\Delta t}$$
$$= \frac{f\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)}{S\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)}$$

where $f\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$ is the conditional density function for $T_m \mid \left(\boldsymbol{W}_{m-1} = \boldsymbol{w}_{m-1}, \boldsymbol{Z}_m = \boldsymbol{z}_m\right)$. Integration gives

$$\Lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = \int_0^{t_m} \lambda\left(t; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) \cdot \mathrm{d}t = -\log\left(S\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)\right)$$

so that

$$S\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = \exp\left(-\Lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)\right)$$

and

$$f\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = \lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) \cdot \exp\left(-\Lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)\right).$$

A common density function f is assumed for $T_m \mid \left(\boldsymbol{W}_{m-1} = \boldsymbol{w}_{m-1}, \boldsymbol{Z}_m = \boldsymbol{z}_m\right)$ for all m (and so also a common distribution function F, a common excess time probability function S, a common conditional hazard rate function λ, and a common integrated conditional hazard rate function Λ), but different functions for different m may be more appropriate. The above formulation assumes that such differences are accounted for using the event order covariate with values equal to the event time index m. Also, the formulation assumes that the range of possible values for T_m is not constrained by the values for $T_1, T_2, \cdots, T_{m-1}$. For

the special case of increasing event times T_m^{cum} (as for recurrent events [9]), values for T_m^{cum} are constrained to be larger than the value for T_{m-1}^{cum}, but then consecutive differences $T_m = T_m^{\text{cum}} - T_{m-1}^{\text{cum}}$ with $T_0^{\text{cum}} = 0$ can be modeled instead using the above formulation.

2.2. Fractional Polynomial Conditional Hazard Rate Modeling

Fractional polynomials [10] [11] are based on power transforms of primary predictors u. For modeling the conditional hazard rate $\lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$, the primary predictors u are functions of t_m and the entries of the vectors $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$. The values for $t_1, t_2, \cdots, t_m$ are positive while the values for $c_1, c_2, \cdots, c_{m-1}$ are nonnegative. For most i, the entries $z_{i,1}, z_{i,2}, \cdots, z_{i,m}$ will be nonnegative so that it is reasonable to restrict $z_{i,m}$ to be nonnegative for $1 \le i \le r$ and $1 \le m \le M$. Under this assumption, power transforms of primary predictors u can be defined as

$$g\left(u; q\right) = \begin{cases} 0, & u = 0 \\ u^q & u > 0 \end{cases}$$

for real-valued powers q.

The fractional polynomial conditional hazard rate is defined as

$$\lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = \boldsymbol{x}^{\text{T}}\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) \cdot \boldsymbol{\beta}$$

where $\boldsymbol{x}\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$ is a $p \times 1$ vector of individual terms $x_j\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$. Each term is a power transform of a single primary predictor or of a product of multiple power-transformed primary predictors (called a geometric combination, generalizing the standard interaction). The primary predictors for the fractional polynomial are a subset of t_m and the entries of $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$ while $\boldsymbol{\beta}$ is a $p \times 1$ vector of slope parameters β_j for $1 \le j \le p$. An intercept parameter can be included in the model using the unit transform $u \equiv 1$ as a primary predictor. This formulation requires that $\lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$ be constrained to take on positive values as addressed later. This restriction could be resolved by modeling instead the natural log of $\lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$ as a linear function in $\boldsymbol{\beta}$ (as in [2] [3] [4] [5]), but then the integral

$$\Lambda\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) = \int_0^{t_m} \lambda\left(t; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right) \cdot \mathrm{d}t$$

cannot always be expressed in closed form. The advantage of the above definition is that power transforms $g\left(u_j; q_j\right)$ are simply integrated with respect to t_m, which can also speed up computation of estimates for the parameter vector $\boldsymbol{\beta}$. The possible forms for $x_j\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$ are a power transform of t_m including the unit transform $g\left(t_m; 0\right) \equiv 1$, a power transform of any one entry of either $\boldsymbol{w}_{m-1}$ or $\boldsymbol{z}_m$, a power-transformed geometric combination in t_m combined with one or more entries of $\boldsymbol{w}_{m-1}$ and/or $\boldsymbol{z}_m$, and a power-transformed geometric combination in one or more entries of $\boldsymbol{w}_{m-1}$ and/or $\boldsymbol{z}_m$ but not in t_m. Consequently, $x_j\left(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m\right)$ can be expressed as the product of a function of t_m and a function of $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$, that is,

$$x_j(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = x_{j,1}(t_m) \cdot x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = g(t_m; q_j) \cdot x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$$

where $x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ is either the unit transform or depends on some subset of the entries of $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$ but not on t_m. When $q_j = 0$ for all j, none of the $x_j(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = x_j(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ depend on t_m, and so the conditional hazard rate is constant in t_m and satisfies

$$\lambda(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = \lambda(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \boldsymbol{\beta}.$$

Also, $x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ equals the unit transform when $x_j(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = x_{j,1}(t_m)$ depends only on t_m. The power q_j is either the power for a single power transform of t_m or equals a product of a power q' of t_m and a power q'' transforming a geometric combination containing the power q' of t_m. In general, $x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ can contain power transforms of the entries of $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$.

Integration gives

$$\Lambda(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = \boldsymbol{h}^{\mathrm{T}}(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \boldsymbol{\beta}$$

where $\boldsymbol{h}(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ is the $p \times 1$ vector with entries

$$h_j(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = \int_0^{t_m} x_j(t; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \mathrm{d}t = \left(\int_0^{t_m} g(t; q_j) \cdot \mathrm{d}t \right) \cdot x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m).$$

Assume that $q_j > -1$, so that

$$h_j(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = \frac{g(t_m; q_j + 1)}{q_j + 1} \cdot x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$$

with $\dfrac{g(t_m; q_j + 1)}{q_j + 1}$ a positive function of time t_m. The second term $x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ is nonnegative, but not identically 0, for example, when it equals a single indicator variable z. By definition, $S(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ is a decreasing function of t_m for each unique possible value of the entries of $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$, and so $\Lambda(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ must be an increasing function of t_m for functions of $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$ determined by the fractional polynomial model.

Let $J = \{j : 1 \le j \le p\}$, $J_1 = \{j : 1 \le j \le p, q_j \ne 0\}$, and $J_2 = J \setminus J_1$. Partition $\boldsymbol{x}(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ into $\boldsymbol{x}_1(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ with entries $x_j(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ for $j \in J_1$ and associated slope parameter vector $\boldsymbol{\beta}_1$ as well as $\boldsymbol{x}_2(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ with entries $x_j(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ for $j \in J_2$ depending only on $\boldsymbol{w}_{m-1}$ and $\boldsymbol{z}_m$ but not on t_m with associated slope parameter vector $\boldsymbol{\beta}_2$. The conditional hazard function $\lambda(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m)$ then satisfies

$$\begin{aligned} \lambda(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) &= \boldsymbol{x}^{\mathrm{T}}(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \boldsymbol{\beta} \\ &= \boldsymbol{x}_1^{\mathrm{T}}(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \boldsymbol{\beta}_1 + \boldsymbol{x}_2^{\mathrm{T}}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \boldsymbol{\beta}_2 \end{aligned}$$

so that

$$\lambda(t_m; \boldsymbol{w}_{m-1}, \boldsymbol{z}_m) = \sum_{j \in J_1} g(t_m; q_j) \cdot x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \beta_j + \sum_{j \in J_2} x_{j,2}(\boldsymbol{w}_{m-1}, \boldsymbol{z}_m) \cdot \beta_j.$$

Assume for now that J_2 is empty. For models based on a single transform of

t_m,

$$\lambda\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) = \lambda\left(t_m\right) = g\left(t_m; q_1\right) \cdot \beta_1$$

can be guaranteed to be positive by restricting the associated slope β_1 to be positive-valued. For models based on multiple transforms of t_m but not on $\boldsymbol{w}_{m-1}$ and z_m,

$$\lambda\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) = \lambda\left(t_m\right) = \sum_{j \in J_1} g\left(t_m; q_j\right) \cdot \beta_j$$

can generate complicated integrals $\Lambda\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) = \Lambda\left(t_m\right)$ that are not increasing in t_m in general. Restricting all the associated slopes β_j for $j \in J_1$ to be positive-valued is a straightforward way to guarantee that $\Lambda\left(t_m\right)$ is increasing in t_m. More general models with

$$\lambda\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) = \sum_{j \in J_1} g\left(t_m; q_j\right) \cdot x_{j,2}\left(\boldsymbol{w}_{m-1}, z_m\right) \cdot \beta_j$$

generate integrals $\Lambda\left(t_m; \boldsymbol{w}_{m-1}, z_m\right)$ that need to be increasing in each unique choice for the terms $x_{j,2}\left(\boldsymbol{w}_{m-1}, z_m\right)$. This can be guaranteed in a straightforward way once again by restricting all the associated slopes β_j for $j \in J_1$ to be positive-valued.

On the other hand, when J_2 is nonempty, the assumption of all positive-valued slope parameters would be too restrictive. For example, consider a model based on two primary predictors: t_m and a time-invariant indicator variable z for membership in one of two groups of study participants with additive effects to t_m and z. The term

$$\boldsymbol{x}_1^{\mathrm{T}}\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) \cdot \boldsymbol{\beta}_1 = \boldsymbol{x}_1^{\mathrm{T}}\left(t_m\right) \cdot \boldsymbol{\beta}_1$$

models the dependence of the conditional hazard function on t_m in terms of $p-1$ parameters while

$$\boldsymbol{x}_2^{\mathrm{T}}\left(\boldsymbol{w}_{m-1}, z_m\right) \cdot \boldsymbol{\beta}_2 = z \cdot \beta_p$$

provides for a shift in the conditional hazard function for the group coded as $z = 1$ compared to the conditional hazard function for the group coded as $z = 0$. In general, this shift can be in the negative or the positive direction, and so the slope of β_2 needs to be allowed to be negative. However, allowing for negative entries for $\boldsymbol{\beta}_2$ means that $\boldsymbol{x}^{\mathrm{T}}\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) \cdot \boldsymbol{\beta}$ and $\boldsymbol{h}^{\mathrm{T}}\left(t_m; \boldsymbol{w}_{m-1}, z_m\right) \cdot \boldsymbol{\beta}$ can sometimes take on negative or zero values. See Section 2.3 for how to handle such cases.

Hazard ratios are naturally generated for models based on proportional hazard rates. Hazard ratios are not needed to assess a fractional polynomial model but can be generated if desired. For example, consider the simple fractional polynomial model $\lambda\left(t_m; t_{m-1}, z\right)$ based on the current time value t_m, the previous time value t_{m-1}, and a single covariate z with values 1 and 2. The associated hazard ratio is given by $\lambda\left(t_m; t_{m-1}, 2\right) / \lambda\left(t_m; t_{m-1}, 1\right)$, that is, the conditional hazard rate for $z = 2$ divided by the conditional hazard rate for $z = 1$. Note that in general this is a function of the current time t_m and the previous time t_{m-1}.

2.3. Maximum Likelihood Estimation

Let $s \in S = \{s : 1 \leq s \leq n\}$ denote indexes for n study participants who provide multiple event time data consisting of observations

$$\boldsymbol{w}_{m,s} = \left(t_{1,s} \quad \cdots \quad t_{m,s} \quad c_{1,s} \quad \cdots \quad c_{m,s} \quad \boldsymbol{z}_{1,s}^{\mathrm{T}} \quad \cdots \quad \boldsymbol{z}_{m,s}^{\mathrm{T}} \right)^{\mathrm{T}}$$

for $1 \leq m \leq M(s) \leq M$. This allows for missing data, but just for large time indexes $M(s) < m \leq M$. This is the kind of missing data occurring in the example data set of Section 3. More complex missing data can be readily handled if desired (see [11]). The total number of measurements is thus

$$M(S) = \sum_{s \in S} M(s).$$

All times $t_{m,s}$ for all participants s can be uncensored, but not all of them should be censored.

Let $\boldsymbol{x}\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right)$ be the observed vectors determined by some fractional polynomial model. For each $s \in S$ and $1 \leq m \leq M(s)$, the conditional likelihood terms $L_{m,s}(\boldsymbol{\beta})$ satisfy

$$L_{m,s}(\boldsymbol{\beta}) = \begin{cases} f\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right), & c_{m,s} = 0 \\ S\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right), & c_{m,s} = 1 \end{cases}$$

so that the associated log-likelihood term $\ell_{m,s}(\boldsymbol{\beta}) = \log\left(L_{m,s}(\boldsymbol{\beta}) \right)$ satisfies

$$\ell_{m,s}(\boldsymbol{\beta}) = \left(1 - c_{m,s} \right) \cdot \log\left(\lambda\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right) \right) - \Lambda\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right).$$

The likelihood $L(S; \boldsymbol{\beta})$ equals the product over $s \in S$ and $1 \leq m \leq M(s)$ of the likelihood terms $L_{m,s}(\boldsymbol{\beta})$ with associated log-likelihood

$$\ell(S; \boldsymbol{\beta}) = \log\left(L(S; \boldsymbol{\beta}) \right) = \sum_{s \in S} \sum_{m=1}^{M(s)} \ell_{m,s}(\boldsymbol{\beta}).$$

The maximum likelihood estimate $\boldsymbol{\beta}(S)$ of $\boldsymbol{\beta}$ maximizes the log-likelihood $\ell(S; \boldsymbol{\beta})$ over $\boldsymbol{\beta}$, which is achieved by solving the estimating equations

$$\frac{\partial \ell(S; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0.$$

For simplicity of notation, parameter estimates $\boldsymbol{\beta}(S)$ are denoted as functions of the index set S for the observed data used in their computation without hat (^) symbols.

The gradient vector $\dfrac{\partial \ell(S; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ has entries satisfying

$$\frac{\partial \ell(S; \boldsymbol{\beta})}{\partial \beta_j} = \sum_{s \in S} \sum_{m=1}^{M(s)} \frac{\partial \ell_{m,s}(\boldsymbol{\beta})}{\partial \beta_j}$$

$$= \sum_{s \in S} \sum_{m=1}^{M(s)} \left(\left(1 - c_{sm,} \right) \cdot \frac{x_j\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right)}{\boldsymbol{x}^{\mathrm{T}}\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right) \cdot \boldsymbol{\beta}} - h_j\left(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s} \right) \right)$$

for $j \in J$. The Hessian matrix $\boldsymbol{H}(S; \boldsymbol{\beta})$ has entries satisfying

$$\frac{\partial^2 \ell(S;\boldsymbol{\beta})}{\partial \beta_{j'} \partial \beta_j} = \sum_{s \in S} \sum_{m=1}^{M(s)} \frac{\partial^2 \ell_{m,s}(\boldsymbol{\beta})}{\partial \beta_{j'} \partial \beta_j}$$

$$= -\sum_{s \in S} \sum_{m=1}^{M(s)} \left((1 - c_{m,s}) \cdot \frac{x_{j'}(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) \cdot x_j(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s})}{\left(\boldsymbol{x}^{\mathrm{T}}(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) \cdot \boldsymbol{\beta} \right)^2} \right)$$

for $j, j' \in J$. The gradient vector and Hessian matrix can be used in a Newton-Raphson algorithm to compute the maximum likelihood estimate $\boldsymbol{\beta}(S)$.

When $\boldsymbol{x}^{\mathrm{T}}(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) \cdot \boldsymbol{\beta} \leq 0$ or $\boldsymbol{h}^{\mathrm{T}}(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) \cdot \boldsymbol{\beta} \leq 0$ for some s, set

$$\lambda(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) = \delta$$

for some small positive value δ like 0.00001 and also set

$$\Lambda(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) = \delta \cdot t_{m,s}.$$

However, leave the gradient vector and Hessian matrix unchanged so that maximum likelihood estimation uses actual derivatives with respect to the parameter vector $\boldsymbol{\beta}$.

Models based on a single transform with $p = 1$ can be computed directly. In this case, $\boldsymbol{\beta} = (\beta_1)$ and the gradient is a scalar value satisfying

$$\frac{\partial \ell(S;\boldsymbol{\beta})}{\partial \beta_1} = \sum_{s \in S} \sum_{m=1}^{M(s)} \left(\frac{1 - c_{m,s}}{\beta_1} - h_1(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s}) \right)$$

so that

$$\beta_1(S) = \frac{n - \sum_{s \in S} \sum_{m=1}^{M(s)} c_{m,s}}{\sum_{s \in S} \sum_{m=1}^{M(s)} h_1(t_{m,s}; \boldsymbol{w}_{m-1,s}, \boldsymbol{z}_{m,s})}.$$

The covariance matrix for the parameter estimate vector $\boldsymbol{\beta}(S)$ can be estimated in two ways: the model-based estimate treats the assumed model as the true model while the robust empirical (or sandwich) estimate allows for the true model to be different from the assumed model. The model-based estimate of the covariance matrix for $\boldsymbol{\beta}(S)$ equals the $J \times J$ matrix satisfying

$$\boldsymbol{\Sigma}_{\mathrm{MB}}(\boldsymbol{\beta}(S)) = -\boldsymbol{H}^{-1}(S;\boldsymbol{\beta}(S)),$$

that is, the negative of the inverse of the Hessian matrix evaluated at the estimated parameter vector $\boldsymbol{\beta}(S)$. The robust empirical estimate equals the $J \times J$ matrix satisfying

$$\boldsymbol{\Sigma}_{\mathrm{RE}}(\boldsymbol{\beta}(SC)) = \boldsymbol{H}^{-1}(S;\boldsymbol{\beta}(S)) \cdot \boldsymbol{G}(\boldsymbol{\beta}(S)) \cdot \boldsymbol{G}^{\mathrm{T}}(\boldsymbol{\beta}(S)) \cdot \boldsymbol{H}^{-1}(S;\boldsymbol{\beta}(S))$$

where $\boldsymbol{G}(\boldsymbol{\beta}(S))$ is the $J \times m(S)$ matrix with entries $\boldsymbol{G}_{j,m,s}(\boldsymbol{\beta}(S))$ for $1 \leq j \leq J$, $1 \leq m \leq M(s)$, and $s \in S$ satisfying

$$\boldsymbol{G}_{j,m,s}(\boldsymbol{\beta}(S)) = \frac{\partial \ell_{m,s}(\boldsymbol{\beta})}{\partial \beta_j}.$$

Note that the gradient vector $\dfrac{\partial \ell(S;\boldsymbol{\beta}(S))}{\partial \boldsymbol{\beta}}$ equals the $J \times 1$ vector generated

by summing the rows of $G(\boldsymbol{\beta}(S))$. The diagonal entries of $\boldsymbol{\Sigma}_{\mathrm{MB}}(\boldsymbol{\beta}(S))$ and $\boldsymbol{\Sigma}_{\mathrm{RE}}(\boldsymbol{\beta}(S))$ are estimates of variances for the parameter estimates corresponding to the entries of $\boldsymbol{\beta}(S)$, and so associated standard errors equal the square roots of these diagonal entries. These standard errors can be used to compute tests for zero individual model parameters to use in assessing specific theoretically important models. However, tests for parameters of adaptively generated models are typically significant as a consequence of the adaptive modeling process (as summarized in Section 2.5). Consequently, results for such tests are not reported for adaptive models generated in example analyses, but are reported for theory-based models.

2.4. Likelihood Cross-Validation

Partition the index set S into $k > 1$ disjoint sets $S(h)$, called folds, for $h \in H = \{h : 1 \le h \le k\}$. Note that all $M(s)$ measurements $\boldsymbol{w}_{m,s}$ for a given index s are allocated to the same fold. The LCV score for some model is defined as

$$\mathrm{LCV} = \prod_{h \in H} \left(L\left(S(h); \boldsymbol{\beta}\left(S \setminus S(h) \right) \right) \right)^{1/M(S)}.$$

In other words, evaluate the likelihood for each fold $S(h)$ using the parameter vector $\boldsymbol{\beta}(S \setminus S(h))$ computed by maximizing the likelihood for the complement $S \setminus S(h)$ of the fold, normalize it by the total number of measurements $M(S)$, and multiply these normalized deleted fold likelihoods together to get the LCV score.

When one or more slopes $\beta_j \le 0$ for some $j \in J_1$ are generated as part of the estimation computations, reset $\ell(S; \boldsymbol{\beta}) = -\Delta$ for some large value Δ such as 700 unless $\ell(S; \boldsymbol{\beta})$ is already smaller than $-\Delta$. For a model with parameter vector estimate $\boldsymbol{\beta}(S)$ having one or more slopes $\beta_j(S) \le 0$ for some $j \in J_1$, reset the LCV score to a very small value δ' such as 10^{-12}. These adjustments guarantee that the adaptive modeling process of Section 2.5 generates models with acceptable parameter vector estimates $\boldsymbol{\beta}(S)$.

A larger LCV score indicates a better model, but not necessarily a distinctly (substantially, significantly) better model. A χ^2-based LCV ratio test, analogous to a likelihood ratio test, can be used to decide if there is a distinct improvement or not in the LCV score. Following [7], these tests are expressed in terms of the percent decrease (PD) in the LCV score for the model with the smaller score compared to the model with the larger score and a cutoff for a distinct PD, changing with the sample size. A PD larger than the cutoff indicates that the model with the larger score distinctly (substantially, significantly) improves on the model with the smaller score. Otherwise, the model with the smaller score is a competitive alternative and, if also simpler, it is a parsimonious, competitive alternative and so a preferable choice. Examples of LCV ratio tests are provided in Section 3.

2.5. Adaptive Model Selection

An effective choice of a fractional polynomial conditional hazard rate model

based on a subset of the primary predictors $t_{m,s}$ and the coordinates of $w_{m-1,s}$ and $z_{m,s}$ is identified adaptively using a heuristic search process controlled by tolerance parameters indicating how much of a change in the LCV score is allowable at each stage of the process. Starting from a base conditional hazard rate model, which is usually the constant model, power transforms and also geometric combinations if requested are systematically added to the model. Then, the expanded model is contracted, removing extraneous transforms if any and adjusting the powers of the remaining transforms. LCV ratio tests are used to decide whether to stop the contraction or continue removing transforms. Only a brief overview of the adaptive modeling process is provided here; details are provided in [11]. Note that geometric combinations generalize standard interactions and so provide for a nonlinear assessment of the concept of moderation [12]. Adaptive modeling applies to both independent and correlated outcomes in a variety of other regression contexts such as linear regression [7], logistic regression [13], Poisson regression [14] [15], and discrete regression [16].

Power transforms of time values $t_{m,s}$ can cause computational problems for the adaptive modeling process when some of those time values are large. For this reason, fractional polynomial conditional hazard rate models are computed using the normalized event time values $t'_{m,s} = t_{m,s}/t^*$ and associated vectors $w'_{m,s}$ computed using these normalized times where

$$t^* \ge \max\left\{t_{m,s} : 1 \le m \le M(s), s \in S\right\}$$

is an upper bound on the observed time values. This generates estimates $\lambda'\left(t'_{m,s}; w'_{m-1,s}, z_{m,s}\right)$ and $S'\left(t'_{m,s}; w'_{m-1,s}, z_{m,s}\right)$ of the conditional hazard rate and the excess time probability function, respectively. Then, define the estimated conditional hazard rate and excess time probability function for the original time values $t_{m,s}$ as

$$\lambda\left(t_{m,s}; w_{m-1,s}, z_{m,s}\right) = \lambda'\left(t_{m,s}/t^*; w'_{m-1,s}, z_{m,s}\right)$$

and

$$S\left(t_{m,s}; w_{m-1,s}, z_{m,s}\right) = S'\left(t_{m,s}/t^*; w'_{m-1,s}, z_{m,s}\right)$$

for any $t_{m,s}$, $w_{m-1,s}$, and $z_{m,s}$. This adjustment guarantees that power transforms are computed using bounded time values within the interval (0, 1]. It is not needed when $t^* \le 1$ to start with, but that is unlikely to hold except in rare cases.

2.6. Modeling Constraints

A variety of constraints are needed so that a fractional polynomial provides an appropriate conditional hazard rate model. Specifically, covariate values $z_{i,m}$ for $1 \le i \le r$, $1 \le m \le M$ are restricted to be nonnegative (Section 2.2). Power transforms $g(t;q)$ have powers restricted to satisfy $q > -1$ (Section 2.2). Conditional hazard rates $\lambda\left(t_{m,s}; w_{m-1,s}, z_{m,s}\right)$ are adjusted to be positive-valued and their integrals $\Lambda\left(t_{m,s}; w_{m-1,s}, z_{m,s}\right)$ to be increasing in $t_{m,s}$ (as addressed in

Section 2.3). Slope parameters β_j for $j \in J_1$ corresponding to power transforms depending on time t are restricted to be positive-valued (as addressed in Section 2.4). Also, fractional polynomial conditional hazard rate models are computed using normalized event times (as addressed in Section 2.5).

3. Example Analyses

Example analyses are presented in this section demonstrating the use and applicability of conditional hazard regression modeling based on fractional polynomials. These analyses use data on recurrence of bladder cancer tumors as described in Section 3.1. Examples are provided of both theory-based modeling and adaptive modeling of these data. Analyses are conducted using SAS® Version 9.4. A SAS macro for conducting adaptive modeling in a variety of regression contexts including conditional hazard regression is available on request from the author.

3.1. Multiple Tumor Event Times for Bladder Cancer Patients

Data are provided in [5] on the recurrence of up to $M = 4$ tumors for $n = 86$ patients initially diagnosed with bladder cancer collected by [17]. Patients have from $M(s) = 1$ to $M(s) = 4$ possibly censored recurrence times for a total of $M(S) = 179$ time measurements. LCV scores are computed using $k = 5$ folds with fold sizes ranging from 15 (17.4%) to 20 (23.3%) patients and from 31 (17.3%) to 42 (23.5%) time measurements so that both fold sizes and fold complement sizes are not proportionately sparse. The cutoff for a distinct percent decrease (PD) in the LCV score for data with $m(S) = 179$ observations is 1.07%.

Patients have follow-up times $t_s^{\max}$ ranging from 0 to 64 months for $s \in S$. Observed tumor recurrence event times $t_{s,m}^{\text{cum}} \leq t_s^{\max}$ have values increasing in m. When a patient has a total of $0 \leq m' \leq 3$ tumor recurrences with $t_{m'}^{\text{cum}} < t_s^{\max}$, a censored time $t_{s,m'+1}^{\text{cum}} = t_s^{\max}$ is added to the measurements for this patient, and so then $m(s) = m' + 1 \leq 4$ for that patient and the only possible censored event time for all patients is the last one $t_{s,M(s)}^{\text{cum}}$. In cases where patients have $m' = 4$ total tumor recurrences and $t_{s,m'}^{\text{cum}} < t_s^{\max}$, a fifth censored time could be added but has not been to restrict to at most 4 event times (as in [5]). One patient has follow-up time $t_s^{\max} = 0$ with no observed tumor recurrences. Technically, this introduces a single censored event time $t_{s,1}^{\text{cum}} = 0$, which is trivial (*i.e.*, the associated likelihood has value 1), and so the follow-up time for this patient has been changed to $t_s^{\max} = 1$ so that $t_{s,1}^{\text{cum}} = 1$ is non-trivial. Example analyses use the consecutive times between tumors given by the differences $t_{s,m} = t_{s,m}^{\text{cum}} - t_{s,m-1}^{\text{cum}}$ with $t_{s,0}^{\text{cum}} = 0$. Censoring is unchanged.

Available covariates include treatment group, the initial number of tumors, and the initial tumor size. Patients are categorized into two treatment groups: 48 (55.8%) receiving a placebo treatment (group = 1) and 38 (44.2%) receiving thiotepa treatment (group = 2). Initial numbers of tumors range from 1 - 8 with 51

(59.3%), 11 (12.8%), 10 (11.6%), and 14 (16.3%) patients having 1, 2, 3, and 4-8, respectively. Initial tumor sizes range from 1 - 7 with 49 (57.0%), 10 (11.6%), 16 (18.6%), and 11 (12.8%) patients having 1, 2, 3, and 4-7, respectively. These covariates are time-invariant. They are denoted as group, number, and size in example analyses for brevity. The primary theory-based model for the data is the one based on these three covariates considered jointly.

Time-varying covariates are also considered in example analyses. These address possible dependence of the conditional hazard rate on event order and time. The dependence predictors include tumor order m with possible values 1 - 4, the current observed time t_m, the prior observed time t_{m-1} with $t_0 = 0$, and the prior cumulative time t_{m-1}^{cum} (or, equivalently, the sum of the prior time values $t_0 + \cdots + t_{m-1}$) with $t_0^{\text{cum}} = 0$. Note that the last two of these predictors induce dependence of an order 1 autoregressive type. Dependence predictors of higher order autoregressive types may have value especially for data with relatively large maximum numbers of event times M, but these are not considered in example analyses for brevity.

Of the 179 tumor recurrence times, 86 (48.0%) occur at tumor order 1, 46 (25.7%) at tumor order 2, 27 (15.1%) at tumor order 3, and 20 (11.2%) at tumor order 4. Also, 67 (37.4%) of these recurrence times are censored.

The example analyses are actually conducted using times t_m^{cum}/t^* normalized by $t^* = 60$ and their differences. However, the unadjusted symbols t_m^{cum} and t_m are used to denote times in the example analyses to reduce the complexity of the notation. The patient index s is dropped for the same reason.

3.2. Modeling the Conditional Hazard Rate

Existing methods [2] [3] [4] [5] for modeling multiple event times are extensions of Cox regression methods, and so ignore the general dependence on current and prior times except for how those times affect covariates. Thus, they restrict to estimating only such covariate effects. However, fractional polynomial conditional hazard rate models need to estimate dependence effects prior to conducting theory-based or adaptive modeling based on other covariates. This section provides example analyses addressing dependence for the bladder cancer event times using fractional polynomials.

To address the dependence of conditional hazard rates over time, consider models based on the following three time-varying predictors: tumor order m, the current time value t_m, and the prior time value t_{m-1}. The adaptive additive model in these three dependence predictors is based on untransformed tumor order m (*i.e.*, transformed by the power 1) without an intercept and with LCV score 1.07006. The associated moderation model in these three dependence predictors (*i.e.*, allowing for geometric combinations) is based on the single power transform $m^{0.99}$ without an intercept, almost the same model as the additive model.

An alternate approach for accounting for dependence would be the adaptive

model based on the following three time-varying predictors: tumor order m, the current time value t_m, and the prior cumulative time value t_{m-1}^{cum}, that is, with the prior time value replaced by the prior cumulative time value. The adaptive additive model in these three dependence predictors is based on untransformed event order m without an intercept, the same model as generated using the prior time value. The associated moderation model in these three dependence predictors is based on the single power transformed geometric combination $\left(m^{1.5} \cdot \left(t_{m-1}^{\text{cum}}\right)^{0.2} \cdot t_m^{0.08}\right)^{1.2}$ with an intercept and LCV score 1.07036. The associated additive model with LCV score 1.07006 has non-distinct PD 0.03% (*i.e.*, less than the cutoff of 1.07% for a distinct PD), and so is preferable as a parsimonious, competitive alternative.

This preferable dependence model generates increasing conditional hazard rate estimates of 1.73, 3.47, 5.20, and 6.93 as tumor order m varies from 1 - 4. Estimated excess time probability curves are plotted in **Figure 1**. The excess time probability decreases with increased times between tumors and more quickly for larger tumor orders m. Also, observed ranges for times between tumors tend to be smaller for larger tumor orders, contributing to higher estimated conditional hazard rates and more quickly decreasing estimated excess time probability curves. This dependence model is used next to address theory-based effects to non-dependence covariates using fractional polynomial conditional hazard rate models.

3.3. Modeling Theory-Based Effects

Analyses of multiple event times using the methods of [2] [3] [4] are supported by SAS PROC PHREG (for proportional hazards regression). Results generated using SAS for these methods applied to the bladder cancer event times are reported in [18] and described in what follows. Results for the method of [5] applied
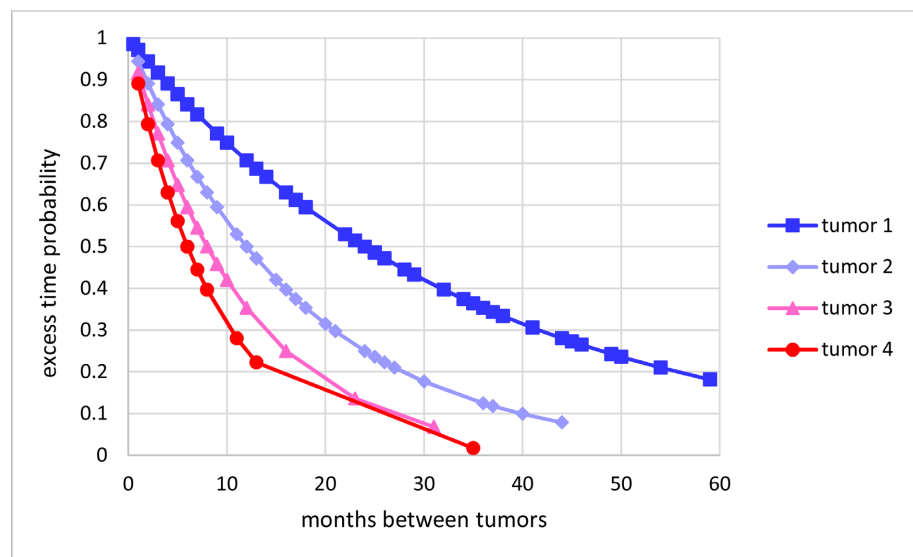

Figure 1. The estimated excess time probability as a function of time between bladder cancer tumors by tumor order.

to the bladder cancer event times are reported in that article.

The model of [2] is called the intensity model since it models the hazard rate. The joint intensity model in group, number, and size has associated p-values 0.022, <0.001, and 0.538, respectively, generated using model-based standard errors. These results support the conclusion of significant joint effects on the hazard rate due to group and number but not due to size. The estimate for group is negative indicating that thiotepa treatment reduces the hazard rate for times between bladder cancer tumors. The estimate for number is positive indicating that a larger number of initial tumors increases the hazard rate for times between bladder cancer tumors. The model of [3] is called the proportional means model. It has the same slope estimates as the intensity model but uses the robust empirical standard errors to test these slope estimates. The joint proportional means model in group, number, and size has associated p-values 0.075, 0.005, and 0.572, respectively, supporting a significant joint effect due to number (and also increasing in number as before), but not due to group and size.

The model of [4] is called a marginal Cox model since it treats the effects of group, number, and size as changing with tumor order m. It can be applied to times between tumors, and is then called a gap model, or alternately to cumulative times. It is more complex having different slopes for group, number, and size for each of the 4 values for tumor order m. Of these 12 slopes, only two estimates for the gap model are significant: the estimates for number for tumor orders 1 and 4 ($p = 0.002$ and $p = 0.034$, both with positive slope estimates). Results reported in [5] for their alternate marginal Cox model applied to the bladder cancer data are described as "providing some evidence, though not very strong, of thiotepa slowing down tumor recurrence" (they focus on only the effects to thiotepa over the 4 tumor orders controlling for number and size rather than on the effects for all three covariates in combination).

Results for the theory-based conditional hazard rate model based on fractional polynomials are reported in Table 1 including model-based and robust empirical tests for group, number, and size. Robust empirical test results are more conservative in the sense that standard errors are larger and so p-values are larger. However, for both of these two types of tests, estimates for group and number

Table 1. Assessing the theory-based model for times between bladder cancer tumors using conditional hazard rate modeling based on fractional polynomials.

Covariate[1]	Estimate	Model-based tests[2]		Robust tests[1,2]	
		SE	p	SE	p
group	−0.70	0.25	0.006	0.28	0.013
number	0.59	0.19	0.002	0.21	0.006
size	0.01	0.12	0.907	0.12	0.911

SE—standard error; [1]Group has values 1 (placebo) and 2 (thiotepa), number has values 1 - 8, and size has values 1 - 7. [2]Tests based on the model controlling for all three covariates as well as for tumor order to address dependence over time.

are significant at $p < 0.05$ while the estimate for size is not significant. This holds while controlling jointly for all three of these covariates as well as for the dependence predictor untransformed tumor order m as identified in Section 3.1. As for the intensity model of [2], the estimate for group is negative indicating that thiotepa treatment reduces the hazard rate for times between bladder cancer tumors and the estimate for number is positive indicating that a larger number of initial tumors increases the hazard rate for times between bladder cancer tumors.

The impact of these dependence and non-dependence covariates can be further assessed adaptively by contracting the full model in all 4 covariates to a parsimonious competitive alternative while restricting the contraction to leave the covariates untransformed. The generated model is based on untransformed effects to tumor order, group, and number with the estimate for group negative as before and the estimate for number positive as before. The inclusion of group and number in this adaptively reduced model means that the removal of each of these from the model generates a distinct PD in the LCV score, providing further support for conclusions about the significance of their joint effects on time between bladder cancer tumors.

The above result indicates that there is a distinct effect to group controlling for number. This raises the question of what effect group has by itself, not controlling for number and size, but still allowing for dependence on untransformed tumor order. The estimated slope for group in this model is negative as before, but it is non-significant for both the model-based ($p = 0.349$) and the robust empirical ($p = 0.436$) tests. This indicates that the conclusion of significance of the effect of group based on theory-based models requires controlling for number.

3.4. Adaptive Modeling Results

This section addresses adaptive modeling of times between bladder cancer tumors. These models assume that dependence is a function of only tumor order m as supported by the results of Section 3.1, but allows for its possible transformation rather than keeping it untransformed as in Sections 3.1-3.2. Section 3.3.1 addresses the impact of group by itself while Section 3.3.2 addresses the combined impact of group and number. Effects to size are not addressed for brevity. This seems reasonable given that size has weak effects for the theory-based models of Section 3.2.

3.4.1. Adaptive Modeling in Terms of Group

The adaptive additive model in tumor order m (to account for dependence) and group is the model based on untransformed tumor order by itself and does not depend on group. This is the same model reported in Section 3.1 with LCV score 1.07006. The associated adaptive moderation model in these two predictors is the model based on the two transformed geometric combinations $\left(\text{group} \cdot m^{-1.22}\right)^{-0.4}$ and $\left(\text{group} \cdot m^{-7}\right)^{-5.31}$ without an intercept and with LCV score 1.08477. The LCV score for the additive model generates a distinct PD of 1.36%, indicating that group distinctly moderates the effect of tumor order on time between blad-

der cancer tumors.

Estimated conditional hazard rates for this moderation model are depicted in **Figure 2**. For all 4 tumor orders, the conditional hazard rate for patients given thiotepa is lower than for patients given placebo treatment with the effect strongest for tumor order 2.

3.4.2. Adaptive Modeling in Terms of Group and Number

In order to generate models in number that are more readily interpretable, analyses in this section use $\min(\text{number}, 4)$, that is, number bounded to be no more than 4, combining the smaller categories 4 - 8 into one. The additive model in tumor order m, group, and $\min(\text{number}, 4)$ is the model based on $m^{0.6}$, $\left(\min(\text{number}, 4)\right)^{12}$, and group with LCV score 1.11380. The associated moderation model in these three predictors is the model based on the two transformed geometric combinations

$$\left(\text{group} \cdot m^{-2} \cdot \left(\min(\text{number}, 4)\right)^{-0.6}\right)^{-0.5}$$

and

$$\left(\text{group} \cdot \left(\min(\text{number}, 4)\right)^{19} \cdot m^{-7}\right)^{1}$$

without an intercept and with LCV score 1.12023. The LCV score for the additive model generates a non-distinct PD of 0.57%, indicating that the additive model is preferable as a competitive alternative. It is also simpler since the effect of group is the same for all values of tumor order and of $\min(\text{number}, 4)$.

Estimated conditional hazard rates for patient on thiotepa treatment based on the above additive model are depicted in **Figure 3**. Estimated conditional hazard rates are increasing in tumor order with, for each tumor order, essentially the
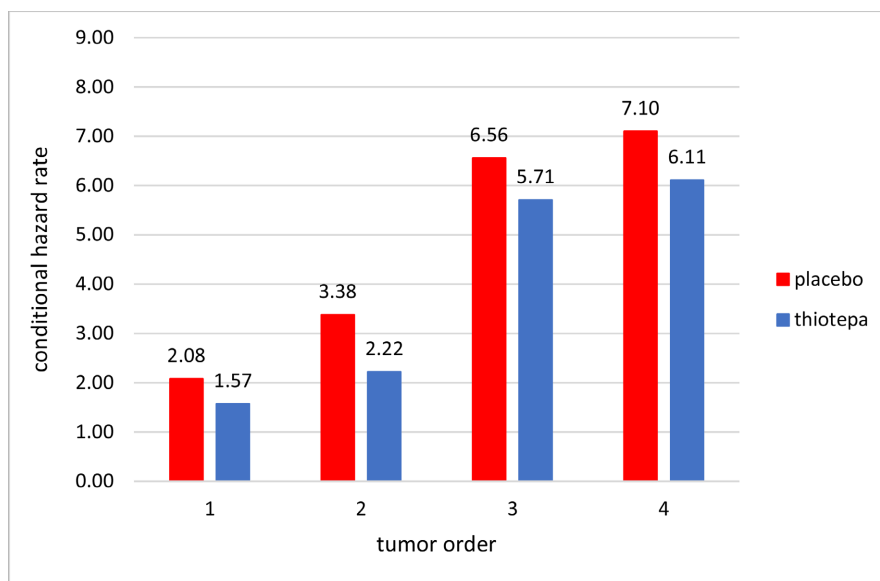

Figure 2. The estimated conditional hazard rate as a function of tumor order and treatment group (placebo or thiotepa) for times between bladder cancer tumors.

same values for numbers 1 - 3 and much larger values for numbers 4 - 8 combined. In all cases, estimated conditional hazard rates are larger by 1.07 units for patients on placebo treatment (not plotted in **Figure 3**).

Figure 4 provides plots for excess time probability curves for number 1 compared to numbers 4 - 8, each by group for tumor order 1. For each group, the curves for numbers 4 - 8 decrease more quickly than the curves for number 1 (which is about the same as for number 2 and for number 3, but these are not plotted). For each of the number 1 and 4 - 8 cases, the curves for the placebo
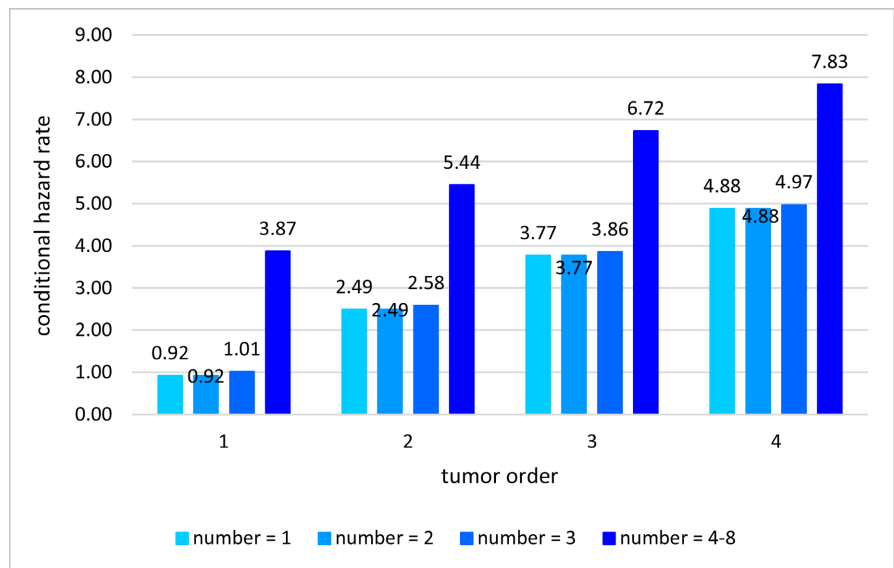
Figure 3. The estimated conditional hazard rate as a function of tumor order and initial number for times between bladder cancer tumors for patients on thiotepa treatment; estimates for placebo treatment are all 1.07 units larger.
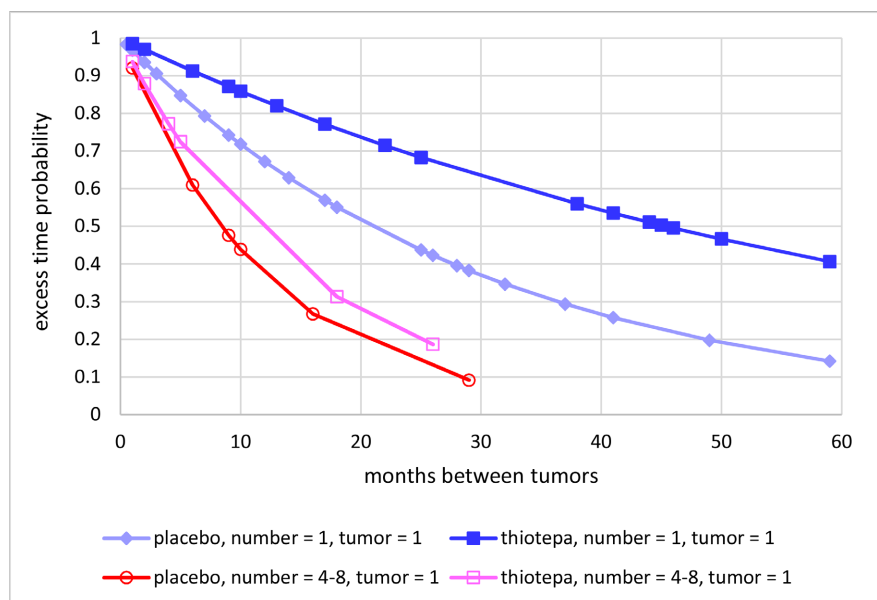
Figure 4. The estimated excess time probability as a function of time between bladder cancer tumors by initial number 1 and numbers 4-8, each by group for tumor order 1.

group decrease more quickly than the curves for the thiotepa group with a stronger thiotepa benefit for number 1 than for numbers 4 - 8.

Figure 5 provides plots for excess time probability curves for the same cases as in Figure 4 but for tumor order 2. The patterns are similar to those of Figure 4, but with curves decreasing more quickly than associated curves of Figure 4. Plots are not provided for tumor order 3 - 4, but in general curves tend to decrease more quickly for larger tumor orders, but observed data are also relatively sparse for tumor order 3 - 4.

4. Discussion

An adaptive approach is formulated for analyzing multiple event time data. Conditional hazard rates are modeled in terms of dependence on both time and covariates using fractional polynomials, that is, weighted sums of products of powers of time and other available predictors. Models are restricted so that conditional hazard rates have positive values and so that estimated excess time probability functions decrease with time for functions of the other predictors determined by the model. Maximum likelihood is used to estimate parameters adjusting for right censored event times.

Likelihood cross-validation (LCV) scores are used to compare models. LCV scores are normalized products of deleted likelihoods for $k = 5$ subsets, called folds, of the data computed using parameters estimated using the complements of those folds. A model with a larger LCV score is a better model, but may not be a distinctly better model. LCV ratio tests generalizing standard likelihood ratio tests are used to identify distinct differences in LCV scores. These tests are based on a cutoff for a distinct percent decrease in the LCV score.
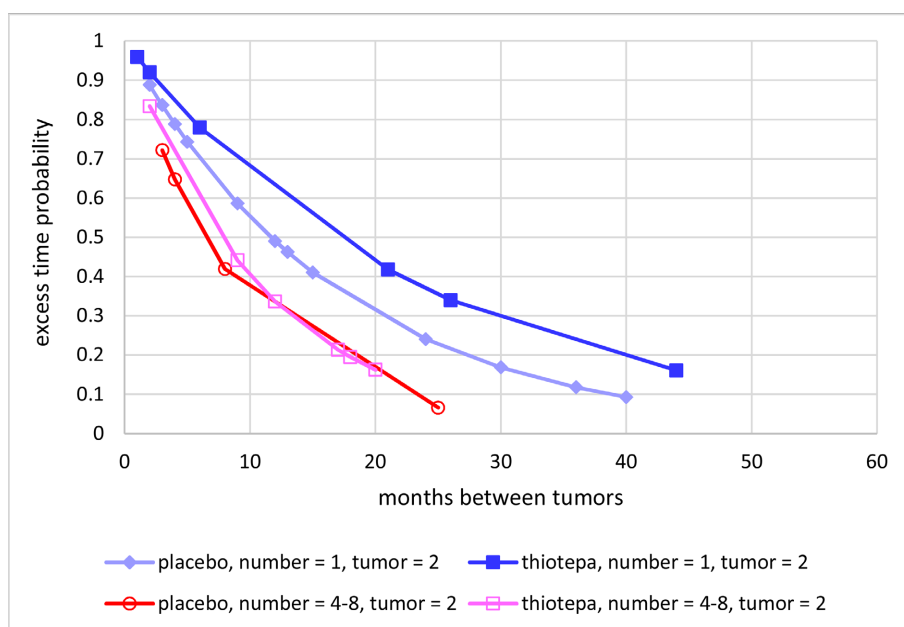

Figure 5. The estimated excess time probability as a function of time between bladder cancer tumors by initial number 1 and numbers 4-8, each by group for tumor order 2.

Adaptive searches through alternative models are used to identify effective models for a given data set in the sense that the removal of any of the model's terms generates a distinct decrease in the LCV score. This is accomplished in two stages: an expansion to grow the model followed by a contraction to prune ineffective terms if any from the expanded model. Geometric combinations equal to products of powers of multiple predictors can optionally be generated. These generalize standard interactions and so provide for a nonlinear assessment of moderation.

Conditional hazards regression is demonstrated using data consisting of up to four possibly censored tumor recurrence times for patients diagnosed with bladder cancer. Dependence is modeled for times between tumor recurrence in terms of tumor order, the current time to recurrence, the prior time to recurrence, and the prior cumulative time. Adaptive analyses indicate that dependence for these times between tumor recurrence is reasonably modeled in terms of tumor order by itself, indicating that adaptive modeling of the tumor recurrence data can reasonably base dependence solely on tumor order. Figure 1 provides plots for associated estimated excess time probability curves by tumor order. This dependence model is used to conduct theory-based modeling of the times between tumor recurrence.

The example analyses start with an assessment of theory-based modeling. The primary theory-based model for times between tumor recurrence is based on the joint covariate effects of treatment group (thiotepa versus placebo), initial number of tumor (1 - 8), and initial size of the tumors (1 - 7). Existing methods for analyzing these data provide conflicting conclusions for this theory-based model. The intensity model of [2] identifies significant effects to group and initial number using model-based tests for zero slopes. The proportional means model of [3] generates the same slope estimates as the intensity model but uses robust empirical tests which identify only a significant effect to initial number. The marginal Cox model of [4] assumes that the slopes for the three covariates change with the four tumor orders, thereby increasing the number of parameters four-fold and so possibly overfitting the data. This model only identifies significant effects to initial number for two of the four tumor orders. The alternate marginal Cox model of [5] focuses on the group effect controlling for the other effects and concludes that there is "some evidence, though not very strong" of a group effect. It is not clear which of these models should be used to reach an appropriate theory-based conclusion.

On the other hand, fractional polynomial modeling indicates that there are significant effects to group and initial number using both model-based and robust empirical tests (Table 1). This approach differs from the approaches of [2] and [3] by estimating the dependence as well as the covariate effects, which results in consistent conclusions about the group effect for both model-based and robust empirical tests. It differs from the marginal Cox approaches of [4] and [5] in using tumor order to model dependence rather than having all covariate effects change with tumor order, which apparently is more effective due to gene-

rating a more parsimonious model. All five approaches indicate that initial tumor size has no effect.

Further assessment of the theory-based model using fractional polynomial modeling indicates that there is a distinct effect to group, but only after controlling for the effect to initial number and not when considered by itself. Further adaptive modeling indicates that group considered by itself distinctly moderates the effect of tumor order on times between tumor recurrence. Figure 2 provides a comparison of estimated conditional hazard rates by tumor order and group for this moderation model.

Adaptive modeling also identifies combined effects to tumor order, group, and initial number of tumors, but with the latter covariate adjusted to combine the small categories of 4 - 8 numbers in order to provide for a more readily interpretable effect to initial number of tumors. The additive model in these predictors is preferable over the associated moderation model. Figure 3 displays the estimated conditional hazard rate over tumor order and initial number for the thiotepa group. The conditional hazard rate for each tumor order is almost the same for each of initial numbers 1 - 3 and then increases for initial numbers 4 - 8 combined and this pattern increases with increasing tumor order. In all cases, the conditional hazard rate for patients on placebo treatment is 1.07 units higher. Figure 4 & Figure 5 provide plots of estimated excess time probability curves generated by this model for tumor order 1 and 2, respectively. Plots are not provided for tumor order 3 and 4, but in general these curves tend to decrease more quickly for larger tumor order. However, observed data are relatively sparse for tumor order 3 and 4.

The example analyses are limited in only addressing recurrent event time data and not more general multiple event time data, but the methods have been formulated to handle more general cases. Censoring for these analyses can only occur for the last event for each participant. However, conditional hazards regression has been formulated in terms of previous time values whether they are censored or not and so generalizes to more complex censoring situations. Conditional hazards regression can also be applied to model clustered multiple event time data (as in [19]) as long as the clusters can be assigned appropriate event orders indexes. Future work is needed to investigate these more complex multiple event time data issues.

In summary, conditional hazard regression modeling has been formulated to model dependence across multiple event times as well as the effects of covariates. It has also been demonstrated using both theory-based and adaptive analyses of times between tumor recurrence for bladder cancer patients. Results of example analyses indicate that adaptive conditional hazard rate modeling can generate useful insights into multiple event time data, but future work is needed to apply this method to more general sets of data.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Kalbfleisch, J.D. and Schaubel, D.E. (2023) Fifty Years of the Cox Model. *Annual Review of Statistics and Its Application*, **10**, 1-23. https://doi.org/10.1146/annurev-statistics-033021-014043

[2] Andersen, P.K. and Gill, R.D. (1982) Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, **10**, 1100-1120. https://doi.org/10.1214/aos/1176345976

[3] Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000) Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. *Journal of the Royal Statistical Society Series B*, **62**, 711-730. https://doi.org/10.1111/1467-9868.00259

[4] Prentice, R.L., Williams, B.J. and Peterson, A.V. (1981) On the Regression Analysis of Multivariate Failure Time Data. *Biometrika*, **68**, 373-379. https://doi.org/10.1093/biomet/68.2.373

[5] Wei, L.J., Lin, D.Y. and Weissfeld, L. (1989) Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, **84**, 1065-1073. https://doi.org/10.1080/01621459.1989.10478873

[6] Royston, P. and Altman, D.G. (1994) Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modeling. *Applied Statistics*, **43**, 429-467. https://doi.org/10.2307/2986270

[7] Knafl, G.J. (2018) Adaptive Fractional Polynomial Modeling. *Open Journal of Statistics*, **8**, 159-186. https://doi.org/10.4236/ojs.2018.81011

[8] Knafl, G.J. (2023) An Adaptive Approach for Hazard Regression Modeling. *Open Journal of Statistics*, **13**, 300-315. https://doi.org/10.4236/ojs.2023.133016

[9] Amorim, L. and Cai, J. (2014) Modelling Recurrent Events: A Tutorial for Analysis in Epidemiology. *International Journal of Epidemiology*, **44**, 324-333. https://doi.org/10.1093/ije/dyu222

[10] Royston, P. and Sauerbrei, W. (2008) Multivariable Model-Building: A Practical Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables. John Wiley & Sons, Hoboken. https://doi.org/10.1002/9780470770771

[11] Knafl, G.J. and Ding, K. (2016) Adaptive Regression for Modeling Nonlinear Relationships. Springer International Publishing, Berlin.

[12] Baron, R.M. and Kenny, D.A. (1986) The Moderator-Mediator Variable Distinction in Social Psychology Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality & Social Psychology*, **51**, 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

[13] Knafl, G.J. (2018) A Reassessment of Birth Defects for Children of Participants of

the Air Force Health Study. *Open Journal of Epidemiology*, **8**, 187-200. https://doi.org/10.4236/ojepi.2018.84015

[14] Knafl, G.J. and Meghani, S.H. (2021) Modeling Individual Patient Count/Rate Data over Time with Applications to Cancer Pain Flares and Cancer Pain Medication Usage. *Open Journal of Statistics*, **11**, 633-654. https://doi.org/10.4236/ojs.2021.115038

[15] Knafl, G.J. (2022) Adaptive Regression for Nonlinear Interrupted Time Series Analyses with Application to Birth Defects in Children of Vietnam War Veterans. *Open Journal of Statistics*, **12**, 789-809. https://doi.org/10.4236/ojs.2022.126045

[16] Knafl, G.J. and Meghani, S.H. (2021) Regression Modeling of Individual-Patient Correlated Discrete Outcomes with Applications to Cancer Pain Ratings. *Open Journal of Statistics*, **12**, 456-485. https://doi.org/10.4236/ojs.2022.124029

[17] Byar, D.P. (1980) The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumors: Comparisons of Placebo, Pyridoxine, and Topical Thiotepa. In: Pavone-Macaluso, M., Smith, P.H. and Edsmyn, F., Eds., *Bladder Tumors and Other Topics in Urological Oncology*, Plenum, New York, 363-370. https://doi.org/10.1007/978-1-4613-3030-1_74

[18] SAS Institute Inc. (2004) SAS/STAT 9.1 User's Guide: Volume 5. SAS Institute Inc., Cary.

[19] Lee, E., Wei, L. and Amato, D. (1992) Cox-Type Regression Analyses for Large Numbers of Small Groups of Correlated Failure Time Observations. In: Klein, J.P. and Goel, P.K., Eds., *Survival Analysis: State of the Art*, *Nato Science*, Vol. 211, Springer, Dordrecht, 237-247. https://doi.org/10.1007/978-94-015-7983-4_14