

Proportional hazards and threshold regression: their theoretical and practical connections

Mei-Ling Ting Lee · G. A. Whitmore

Received: 11 February 2009 / Accepted: 24 October 2009 / Published online: 4 December 2009
© Springer Science+Business Media, LLC 2009

Abstract Proportional hazards (PH) regression is a standard methodology for analyzing survival and time-to-event data. The proportional hazards assumption of PH regression, however, is not always appropriate. In addition, PH regression focuses mainly on hazard ratios and thus does not offer many insights into underlying determinants of survival. These limitations have led statistical researchers to explore alternative methodologies. Threshold regression (TR) is one of these alternative methodologies (see Lee and Whitmore, *Stat Sci* 21:501–513, 2006, for a review). The connection between PH regression and TR has been examined in previous published work but the investigations have been limited in scope. In this article, we study the connections between these two regression methodologies in greater depth and show that PH regression is, for most purposes, a special case of TR. We show two methods of construction by which TR models can yield PH functions for survival times, one based on altering the TR time scale and the other based on varying the TR boundary. We discuss how to estimate the TR time scale and boundary, with or without the PH assumption. A case demonstration is used to highlight the greater understanding of scientific foundations that TR can offer in comparison to PH regression. Finally, we discuss the potential benefits of positioning PH regression within the first-hitting-time context of TR regression.

Keywords Boundary · Brownian motion · First hitting time · Gamma process · Poisson process · Process time · Renewal process · Stochastic process · Survival time · Time to event

M.-L. T. Lee (✉)
University of Maryland, College Park, MD, USA
e-mail: MLTLEE@UMD.EDU

G. A. Whitmore
McGill University, Montreal, QC, Canada

1 Introduction

Proportional hazards (PH) regression has been an established methodology for analyzing survival and time-to-event data for almost four decades and has proven its usefulness to practitioners in many different disciplines ranging from engineering to medicine (see [Lee and Whitmore 2006](#), for a review). The methodology is a standard module in major statistical software packages and a topic found in most statistics courses dealing with survival data. It is easy to use and understand, reasonably robust, and suited to diverse applications. The development of PH regression is usually attributed to Sir David Cox and, thus, often called Cox regression ([Cox 1972](#)).

The proportional hazards assumption of PH regression, however, is not always suitable and, thus, statistical researchers have explored many alternatives. Threshold regression (TR) is one of these alternative methodologies. It has a long history but only recently has begun to attract major interest (see [Lee and Whitmore 2006](#), for a review; also, [Aalen et al. 2008](#)). The connection between TR and PH regression has been examined in previous published work but the investigations have been limited in scope. For example, [Lee et al. \(2009b\)](#) use an observational case study to compare empirically the benefits of TR over PH regression in understanding the influence of smoking on lung cancer. Their analysis of the study data is cross-sectional and provides no theoretical connections between TR and PH regression. The same authors, in a second investigation of the same data set, look at the application of TR with Markov decomposition to handling the longitudinal features of the data. Their investigation shows that PH regression is consistent with their decomposition procedure ([Lee et al. 2009c](#)).

In this article, we study the theoretical connections between these two regression methodologies in greater depth and show that PH regression is, for most purposes, a special case of TR. We then set out to explore various ways in which TR models can be made to have the PH property. We discuss how to estimate the TR time scale and boundary, with or without the PH assumption. A case demonstration, based on a randomized clinical trial setting, is used to highlight the greater understanding of scientific foundations that TR can offer in comparison to PH regression. Finally, we discuss the potential benefits of positioning PH regression within the first-hitting-time context of TR regression.

In our exposition, we will choose terminology that comes from the fields of health and medicine. The reader will easily see, however, how the terms can be transformed to another area of application, such as engineering. For example, we speak of subjects rather than items, survival time rather than failure time, deteriorating health rather than physical degradation, and so on.

2 Threshold and proportional hazards regression models

We first describe the two kinds of models and comment on their suitability for understanding scientific phenomena.

2.1 Threshold regression model

The basic mathematical setting for threshold regression is the first hitting time of a boundary by a stochastic process. If S , B and $\{Y(t), t \geq 0\}$ denote the first hitting time, boundary and stochastic process, respectively, then their interconnection can be expressed mathematically as follows:

$$S = \inf\{t : Y(t) \in B\}, \quad (1)$$

where initial level $Y(0) \notin B$. In a medical context, the stochastic process $\{Y(t)\}$ describes the time trajectory of health or disease for a subject. The parameter t denotes time. The boundary B is a critical health state, disease state or other medical end point, such as death, a diagnosis of cancer, or hospital discharge. The first hitting time S is the time for the sample path of the stochastic process to first reach the boundary B . It is this first hitting time, or FHT for short, that is the time-to-event or survival time of interest. As we will show later, definition (1) can be extended mathematically in several important ways. For example, we will allow the boundary B to vary with time rather than remain fixed.

From the preceding model, it can be seen that a TR model has three building blocks: (1) a stochastic process that describes the evolution of a subject's underlying health state; (2) a boundary or threshold that defines a critical level or condition that triggers the event of interest when it is reached by the process for the first time; and (3) a time scale on which the process unfolds. Each of these building blocks may have parameters that depend on a covariate vector \mathbf{z} through regression link functions. After a specification of the threshold regression model, the regression functions can be estimated and various inferences can then be made using conventional statistical theory.

2.2 Proportional hazards regression model

The PH regression model assumes that the time to the event or endpoint of interest is a positive random variable with a hazard function of the following form:

$$h(t) = h_0(t) \exp(\mathbf{z}\boldsymbol{\beta}) \quad (2)$$

Here $h_0(t)$ is a fixed baseline hazard function, \mathbf{z} is a row vector of covariates, and $\boldsymbol{\beta}$ is a column vector of regression coefficients that are to be estimated. We limit our attention to fixed covariates in our investigation here but recognize that interesting analytical extensions are offered by considering time-varying covariates of the form $\mathbf{z}(t)$.

2.3 Suitability to real-world applications

The wide use of the proportional hazard model reflects more its mathematical convenience and ease of interpretation than its realism. The occurrence of proportional hazard functions in nature, however, is actually rare, although the model specification

may provide an adequate approximation in some applications. In contrast, first hitting time models are ubiquitous. By using special constructions, the PH feature can be embedded in TR models. Not surprisingly, because TR models have more building blocks (a stochastic process, absorbing boundary, and process time scale), one discovers that different TR models can produce the same family of proportional hazard functions, as we discuss in the next section. In this paper, we will show that a PH model might be viewed as a reduced mathematical form of a TR model, albeit a highly specialized form. The reduced form discards features of the TR model and, as a consequence, gives a more limited view of the underlying scientific phenomenon.

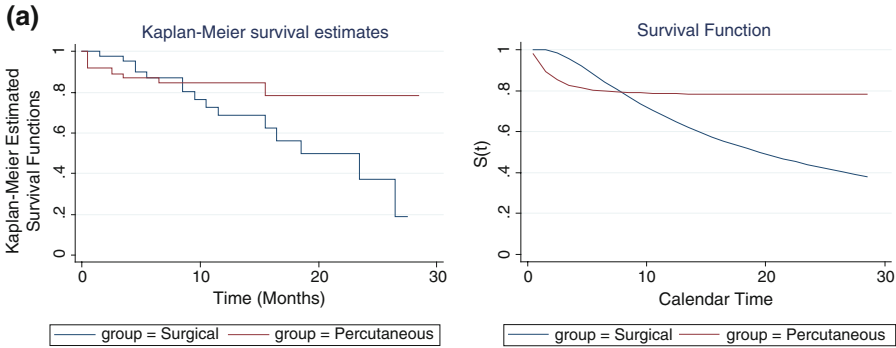
2.4 Time to infection in kidney dialysis: a TR demonstration

To demonstrate the TR approach in a typical medical setting, we present results for a study reported in Klein and Moeschberger (2003, pp. 6–7) based on original data reported in Nahman et al. (1992). The study looks at “time to first exit-site infection (in months) in patients with renal insufficiency, 43 patients utilized a surgically placed catheter (group 1) and 76 patients utilized a percutaneous placement of their catheter (group 2).” We fit a model to the data based on the FHT of a Wiener diffusion process. Time to infection is taken as the FHT of the zero level for the process starting at health level $y_0 > 0$. The model has two parameters: initial health level y_0 and process mean μ , which we link to an indicator variable *percutaneous* representing the catheter placement method (surgical 0, percutaneous 1). The link functions are the natural logarithm for y_0 (i.e., $\ln y_0$) and an identity function for μ . For expository convenience, we do not build an elaborate TR model. For example, we use the calendar time scale and only one covariate (*percutaneous*). To estimate our TR model, we employ a software package that is publicly available on a personal research website (Lee 2009).

Panel (a) of Fig. 1 shows the Kaplan-Meier (KM) plots for the two groups on the left and the corresponding fitted TR survival curves on the right. The fitted model appears plausible. The survival curves cross which suggests strongly that the proportional hazards assumption doesn't hold. The survival curve of the percutaneous group approaches a plateau at a survival probability level well above zero which indicates that a proportion of patients may be immune to infection. Both the crossing feature and plateau feature are accommodated naturally by the first hitting time model.

Panel (b) presents the TR regression output from the software package. The covariate *percutaneous* has a significant negative coefficient (-1.0731) with respect to $\ln y_0$ which suggests that the percutaneous patients have a lower health level at the start of the study. The precipitous fall in the survival function within the first month illustrates this effect. The covariate has a significant positive coefficient ($.6377$) for μ which shows that these same patients progress to infection less slowly. In fact, the estimated value of μ for percutaneous patients is positive ($-.0959 + .6377 = .542$), which indicates that the Wiener diffusion process is tending to drift away from the boundary at zero and, hence, the presence of the plateau in the survival curve that is seen in panel (a).

Finally, panel (c) shows the hazard functions for these fitted survival curves. The hazard for the percutaneous group is seen to drop sharply toward zero and that of



(b)

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
$\ln y_0$					
_group_2	-1.073	.189	-5.67	0.000	-1.444 - .702
_cons	1.411	.143	9.84	0.000	1.130 1.692
μ					
_group_2	.637	.128	4.98	0.000	.387 .889
_cons	-.096	.076	-1.25	0.210	-.246 .054

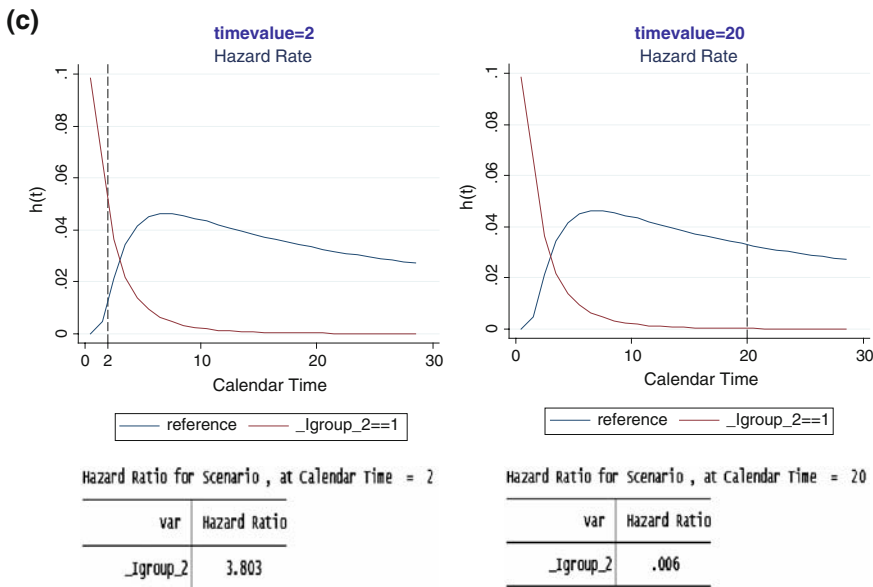


Fig. 1 A TR demonstration for time to infection in kidney dialysis using two methods of catheter placement. The TR model is the first hitting time of the zero level in a Wiener diffusion process. Process parameters are the natural logarithm of the initial health level $\ln y_0$ and the process mean μ . Panel (a) shows Kaplan-Meier plots and fitted survival curves for the TR model. Panel (b) shows TR regression output from a public software. Panel (c) shows the hazard functions for the two methods and hazard ratios at two time points

the surgical group first rises and then tends to level off at a high level. The hazard functions are clearly not proportional. The hazard ratios for the percutaneous group relative to the surgical group are calculated at 2 months and 20 months. The differing ratios show the non-proportionality in sharp contrast.

3 Generating PH functions using the TR model

We now present two methods of constructing TR models that possess the PH feature. As will be seen from the methods, every family of proportional hazard functions for fixed covariates can be generated by a suitable TR model under mild regularity conditions. The first construction method varies the time scale of a TR model. The second varies the boundary of a TR model. It might seem that a third method, namely, varying the stochastic process itself, is also available. As we discuss in Sect 3.3, this third option is not so accessible.

3.1 Generation by varying process time

Consider an increasing function $r(t|z)$, with $r(0|z) = 0$. Here $r(t|z)$ denotes a *process time* r expressed as a function of *calendar time* t . The function is shown as being dependent on the covariate vector z . Process time r is sometimes referred to as operational time, running time or analysis time. In a health context, process time will be some time-like measure that describes progression of disease, cumulative exposure to a toxin, or the like, much as cumulative mileage might serve as process time for an automobile. The requirement that $r(t|z)$ be an increasing function in t is adopted for expository convenience. The requirement can and should be relaxed to ‘non-decreasing’ in some applications where a process is occasionally interrupted in calendar time. For example, a copy machine component may deteriorate only when the machine is copying and not when it is idle. As a health example, a subject’s knee joint may deteriorate only when it is in use and not when it is at rest. The condition of the joint remains unchanged during intervals of rest. We do not consider this extension.

We start by choosing an FHT survival function defined in terms of process time r . We denote this function by $\bar{F}_0(r)$ and refer to it as our *reference* FHT survival function. For technical convenience, we limit our attention to cases where $\bar{F}_0(r)$ is a continuous decreasing function of r , leaving the non-increasing case to later studies. We present illustrations of reference FHT survival functions after explaining the general construction technique. Given a process time $r(t|z)$, the survival function $\bar{F}(t|z)$ in calendar time is obtained by substituting $r(t|z)$ for r in $\bar{F}_0(r)$, i.e.,

$$\bar{F}(t|z) = \bar{F}_0[r(t|z)]. \tag{3}$$

If survival function $\bar{F}(t|z)$ is also to possess the PH feature then by definition

$$\bar{F}(t|z) \equiv \exp \left[- \exp(z\beta) \int_0^t h_0(u) du \right] = \exp \left[- \exp(z\beta) H_0(t) \right], \tag{4}$$

where $H_0(t)$ defines the cumulative baseline hazard function. Correspondence (3) follows from the fact that $r(t|z)$ is an increasing function. Correspondence (4) is the definition of the PH model, operating in calendar time t . Equating (3) and (4) and solving for $r(t|z)$, as shown next, gives a family of process time functions for the reference FHT survival function.

$$r(t|z) = \bar{F}_0^{-1} \left\{ \exp \left[-\exp(z\beta) H_0(t) \right] \right\} \quad (5)$$

Function $\bar{F}_0^{-1}(\cdot)$ denotes the inverse of the reference survival function. The process time functions in (5) generate proportional hazard functions for survival time that have the desired cumulative baseline hazard function.

A special case arises if the cumulative baseline hazard function $H_0(t)$ is chosen to be that of the reference survival function $\bar{F}_0(t)$ defined in calendar time t . In this case, setting $z = 0$ in Eq. 5 and using the fact that $\bar{F}_0(t) = \exp[-H_0(t)]$ gives the identity:

$$r(t|0) = \bar{F}_0^{-1}[\bar{F}_0(t)] = t. \quad (6)$$

In this case, the baseline process time is identical to calendar time. Equation 5 also yields another interesting relationship, namely,

$$\bar{F}_0[r(t|z)] = \bar{F}_0(t)^{\exp(z\beta)}. \quad (7)$$

Equation 7 shows that the family of PH survival functions form a power distribution family in this case where the power coefficient is hazard ratio $\exp(z\beta)$.

We now present two illustrations of this construction method:

Example: Poisson process. A Poisson process generates a simple family of FHT distributions. We now show that an arbitrary family of proportional hazard functions can be generated by considering the time to the first event in a non-homogeneous Poisson process. We choose a Poisson process operating at a unit rate to define the reference FHT survival function. The survival probability corresponding to process time r for this reference Poisson case is $P = \bar{F}_0(r) = \exp(-r)$. The inverse function is $r = \bar{F}_0^{-1}(P) = -\ln(P)$. Solving for $r(t|z)$ in (5) gives a process time scale defined by:

$$r(t|z) = \exp(z\beta) H_0(t). \quad (8)$$

Thus, the Poisson version of the PH family is one in which the event of interest occurs randomly in a non-homogeneous Poisson process. The occurrence rate function in this process has the form $r(t|z)$ in (8). The effect of covariate vector z is simply to shrink or expand the rate at which time runs. Specifically, if the baseline time increment is denoted by dH_0 , the corresponding increment for a subject with covariate vector z is $dr = \exp(z\beta)dH_0$.

Example: Brownian motion process. Consider the FHT to a fixed boundary for a Brownian motion process $\{Y(t)\}$. As the reference case in this family, we take the

FHT of Brownian motion with unit variance that starts at $Y(0) = y_0 > 0$ and has a boundary B at zero. The reference FHT survival function is:

$$\bar{F}_0(r) = 1 - 2\Phi(-y_0/\sqrt{r}) \tag{9}$$

and, thus,

$$\bar{F}(t|\mathbf{z}) = 1 - 2\Phi(-y_0/\sqrt{r(t|\mathbf{z})}). \tag{10}$$

Here Φ denotes the standard normal c.d.f.. Inverting the reference survival function and solving (5) for process time function $r(t|\mathbf{z})$ gives:

$$r(t|\mathbf{z}) = \left(\frac{y_0}{\Phi^{-1} \left\{ \frac{1}{2} \left(1 - \exp[-e^{\mathbf{z}\beta} H_0(t)] \right) \right\}} \right)^2. \tag{11}$$

The construction that we have just described is intimately linked to the canonical formulation of collapsible survival models where $r(t|\mathbf{z})$ in the notation here corresponds to *ideal time* in the collapsible model. The basic idea of an ideal time scale is that a patient’s survival prospects going forward are completely characterized by the patient’s current position r on the ideal time scale. In other words, ideal time r measures the patient’s state of cumulative wear and tear, in health terms. Refer to Cox and Oakes (1984), Oakes (1995), Kordonsky and Gertsbakh (1997), Duchesne and Lawless (2000), and Duchesne and Rosenthal (2003) for more details.

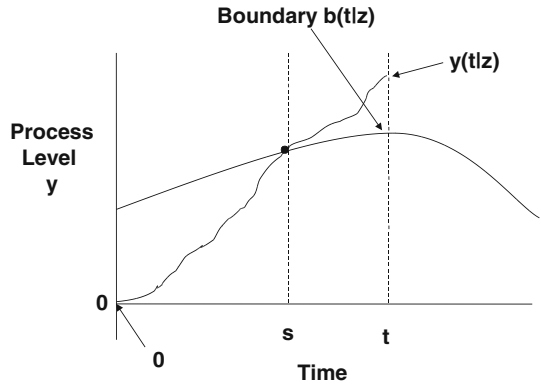
3.2 Generation by varying boundaries

Proportional hazard functions can also be constructed by varying boundaries in FHT contexts. Figure 2 shows an illustration of the basic framework for this construction method. The figure shows a typical sample path of a stochastic process $\{Y(t)\}$ and a boundary $b(t|\mathbf{z})$ that varies with time t and depends on the covariate vector \mathbf{z} . The sample path starts at the origin $Y(0) = 0$. To reach a level above the boundary $b(t|\mathbf{z})$ at any given time t , say level $y(t|\mathbf{z})$, the sample path must make a first transition to the boundary at some intermediate time s and then proceed in time interval $(s, t]$ from the boundary to level $y(t|\mathbf{z})$.

Let $Pr[b(t|\mathbf{z})]$ denote the probability that a sample path, starting at the origin, will lie above the boundary $b(t|\mathbf{z})$ at time t . Let $F(s|\mathbf{z})$ denote the c.d.f. for the FHT S of the boundary by the sample path. Note that the sample path will intersect the boundary at $b(s|\mathbf{z})$ where $S = s$. Finally, let $Pr[b(t|\mathbf{z}) - b(s|\mathbf{z})]$ denote the probability that the sample path will move from level $b(s|\mathbf{z})$ at time s to a level above $b(t|\mathbf{z})$ at time $t > s$.

The construction assumes the stochastic process has stationary independent increments. Translating the previous description into mathematical notation gives the following identity.

Fig. 2 Constructing boundaries in an FHT context that generate proportional hazard functions



$$Pr[(b(t|z))] = \int_0^t Pr[b(t|z) - b(s|z)]dF(s|z) \tag{12}$$

The framework in Fig. 2 and the identity (12) will now be applied in two examples. *Example : Brownian motion process.* In this illustration, the boundary of a standard Brownian motion is varied to produce a family of FHT distributions that have the PH property. Identity (12) has the form (13) in this case.

$$\Phi \left[-\frac{b(t|z)}{\sqrt{t}} \right] = \int_0^t \Phi \left[-\frac{b(t|z) - b(s|z)}{\sqrt{t - s}} \right] dF(s|z) \tag{13}$$

Here $b(t|z)$ is the boundary associated with covariate vector z and $F(s|z)$ is the c.d.f. of the FHT time (i.e., survival time). Recall that Brownian motion has independent stationary normal increments and, hence, the standard normal c.d.f. Φ is involved. The sample path of Brownian motion is continuous so the visual representation in Fig. 2 applies exactly to this case. This method of boundary construction and the identity (13) for Brownian motion was first proposed by Whitmore (1986).

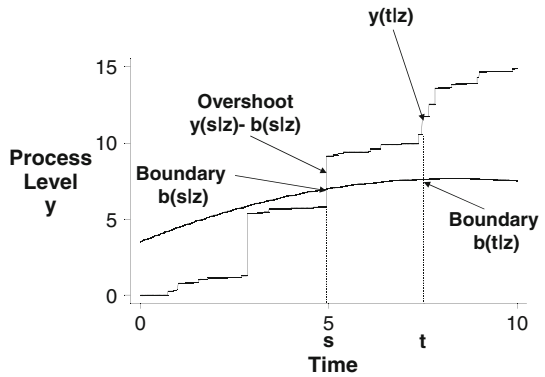
Now, to obtain boundaries that generate a family of PH functions, the following PH variant of the distribution function $F(s|z)$ is used in (13):

$$F(s|z) \equiv 1 - \exp[-\exp(z\beta)H_0(s)]. \tag{14}$$

The integral equation in (13) is numerically solved for $b(t|z)$ for each z of interest. A case demonstration presented later illustrates the method in this context.

Example: Gamma process. A gamma process offers a second example of how a boundary might be varied to produce a family of FHT distributions that have the PH property. We choose a gamma process with a unit scale parameter and shape parameter α . The gamma process has stationary independent gamma increments but does not possess continuous sample paths. The situation is illustrated in Fig. 3 which shows a simulated gamma sample path crossing a boundary.

Fig. 3 A boundary crossing for a gamma process which illustrates how the sample path overshoots the boundary at the first hitting time $S = s$



Identity (12) has the following form in this case:

$$\bar{G} [b(t|z)] = \int_0^t \bar{G}^+ [b(t|z) - b(s|z)] dF(s|z). \tag{15}$$

Here $\bar{G}(\cdot)$ denotes the complementary incomplete gamma function:

$$\bar{G}(w) = \int_w^\infty \frac{u^{\alpha-1}}{\Gamma(\alpha)} \exp(-u) du. \tag{16}$$

Observe that the shape parameter α is suppressed in the \bar{G} notation. The function $\bar{G}^+(\cdot)$ in (15) is a related function that we define in a moment. As in the preceding example with Brownian motion, identity (15) is used in this construction by replacing function $F(s|z)$ by the PH c.d.f. in (14). The identity is then solved numerically for $b(t|z)$ for any specified z .

We now return to the function $\bar{G}^+(\cdot)$. As mentioned earlier, the sample path of a gamma process is not continuous. It happens to be a series of random steps of random size as shown in Fig. 3. It follows therefore that the sample path will overshoot the boundary $b(s|z)$ at the FHT $S = s$. If we let u denote the overshoot distance, i.e., the difference $y(s|z) - b(s|z)$ as marked in the figure, v denote the difference $y(t|z) - b(s|z)$, and w denote the difference $b(t|z) - b(s|z)$ then we have:

$$\bar{G}^+(w) = \int_w^\infty \int_0^v q(u) g(v - u) du dv. \tag{17}$$

The notation in (17) suppresses the dependence on z . The function $g(\cdot)$ is the gamma p.d.f. The function $q(u)$ is the limiting p.d.f. of u , which is shown in the Appendix to have the following form:

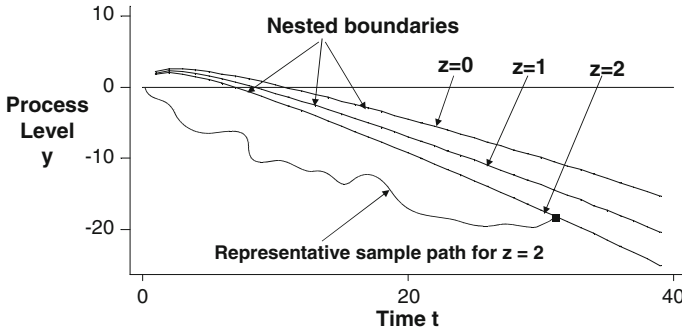


Fig. 4 Three nested boundaries in Brownian motion that generate proportional hazard functions for three levels of covariate z

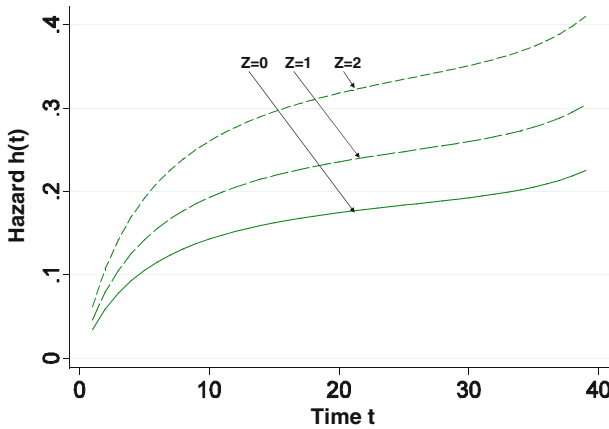


Fig. 5 Proportional hazard functions generated by Brownian motion reaching the three boundaries in Fig. 4

$$q(u) = \int_u^\infty \frac{1}{h} \exp(-h) dh \quad \text{for } u > 0. \tag{18}$$

The inner integral in (17) is a convolution of $y(s|z) - b(s|z)$ and $y(t|z) - y(s|z)$ which are probabilistically independent variates.

We observe that if α is large and overshoot is ignored then $\overline{G}^+(w)$ may be approximated by $\overline{G}(w)$, the complementary c.d.f. of a gamma distribution.

We now present one demonstration to illustrate the preceding methods of construction. We choose to vary the boundary in a Brownian motion. Figure 4 shows three boundaries in Brownian motion that generate proportional hazard functions for three levels of a covariate z . These proportional hazard functions appear in Fig. 5. The figures were produced by numerically solving (13) using the PH c.d.f. in (14) with a gamma cumulative baseline hazard function $H_0(s)$ corresponding to a shape parameter 2 and mean parameter 10. The PH regression parameter for z was set at $\beta = 0.3$ so that the hazard increases with z .

3.3 Generation by varying the stochastic process

One might anticipate that proportional hazard functions can be generated in a FHT context by holding the boundary and time scale fixed and letting the stochastic process vary over some parametric family. It appears difficult, however, to produce a family of PH functions in this manner because most stochastic processes do not have sufficient parametric flexibility for the task. Indeed, so far we have found only the Poisson process, among common stochastic process families, is capable of generating proportional hazard functions and, in this trivial case, they are constant functions. The outstanding question of whether any parametric family of stochastic processes exists that can produce an arbitrary family of PH functions while keeping the boundary and time scale fixed is left as an open question for future research.

4 Estimation methods for a boundary and process time

The preceding developments have demonstrated the theoretical connection between TR and PH models and shown that many kinds of TR models can produce a given family of PH functions. The theoretical connection takes on practical interest when we consider estimation of these models. The topic can be approached from two perspectives. First, if an investigator feels that the PH model is plausible (and perhaps tests do not reject this hypothesis) then the investigator may wish to explore the PH structure within a TR context. Second, the investigator may wish to estimate the boundary or process time function within the context of a TR model without imposing the PH requirement.

4.1 Estimation while retaining the PH property

In this estimation procedure the following estimated survival function from the PH regression model is substituted into the TR model in place of $\bar{F}(t|\mathbf{z})$:

$$\bar{F}_{PH}(t|\mathbf{z}) = \exp \left[- \exp(\mathbf{z}\hat{\boldsymbol{\beta}}) \hat{H}_0(t) \right]. \quad (19)$$

Here the subscript PH on the survival function designates the estimated function; $\hat{\boldsymbol{\beta}}$ denotes the estimated vector of regression coefficients and $\hat{H}_0(t)$ is the estimated cumulative baseline hazard function—both standard computer output in PH regression routines. After the substitution of $\bar{F}_{PH}(t|\mathbf{z})$ for the true survival function for various \mathbf{z} , the corresponding family of boundary functions or process time functions are estimated, as the case may be. For example, an investigator may feel that the PH time-to-event data in his or her study arise as FHTs of a zero-boundary in Brownian motion. Thus, the investigator may wish to estimate the family of process time functions that correspond to the family of estimated PH functions using (11).

4.2 Estimation without imposing the PH property

The second estimation procedure uses Kaplan-Meier (KM) estimates of the survival function, denoted by $\bar{F}_{KM}(t|\mathbf{z})$, for selected values of the covariate vector \mathbf{z} . These functions can provide estimates of process time functions $r(t|\mathbf{z})$ or boundary functions $b(t|\mathbf{z})$. The procedure simply replaces $\bar{F}(t|\mathbf{z})$ in earlier formulas by $\bar{F}_{KM}(t|\mathbf{z})$. For the preceding demonstrations involving Brownian motion, for example, the substitutions are made in expression (10) or expression (13), according to whether the process time function or boundary function is desired. To elaborate a little on the procedure for estimating the boundary, we observe that the KM estimator yields the following discrete approximation to the integral equation in (13):

$$\Phi \left[-\frac{b_j}{\sqrt{t_j}} \right] = \sum_{i=1}^j p_i \Phi \left[-\frac{b_j - b_i}{\sqrt{t_j - t_i}} \right], \quad (20)$$

where the pairs $(t_j, 1 - \sum_{i=1}^j p_i)$, $j = 1, 2, \dots$, define the KM estimator of the survival function $\bar{F}(t|\mathbf{z})$. The quantity $\Phi \left[-(b_j - b_i)/\sqrt{t_j - t_i} \right]$ on the right-hand side is taken to be 1/2 when $i = j$. Solving (20) by numerical iteration for the quantities b_j , $j = 1, 2, \dots$, gives the corresponding KM estimator of the boundary (see Whitmore 1984, 1986). For applications having continuous covariates, a generalized KM estimator of $\bar{F}_{KM}(t|\mathbf{z})$ for a fixed covariate \mathbf{z} might be used. This generalized estimator employs a kernel function modification of the standard KM estimator (Beran 1981). Its application in this setting remains to be developed.

Estimating a boundary function from the PH or KM estimator in the case of a gamma process is numerically more challenging. The procedure is straightforward in principle but requires the numerical evaluation of the double integral in (17).

5 Benefits of having a TR interpretation of PH functions

We have shown that any family of proportional hazard functions can be generated by varying the time scales or boundaries of a TR model, subject to only mild regulatory conditions. We have also shown that different TR models can produce the same family of PH functions. This fact has several implications. First, knowledge of the applicable TR model represents more fundamental knowledge about the real phenomenon under study because it is a more refined or detailed statistical model than the PH model. Second, as a rule, the true TR model will not be identifiable from time-to-event and covariate data alone. A full identification usually requires subject matter knowledge that can help choose the correct TR model in terms of its component boundary, stochastic process or time scale. Third, many (indeed, most) TR models encountered in the real world will not correspond to PH families. The PH regression model is not suitable for these cases. The latter fact is one of the strongest arguments for having threshold regression in the statistical toolbox.

Given the pervasive use of PH regression, it is interesting to consider the potential benefits of interpreting PH regression results using a TR context. Our basic scenario for

this discussion is the following. Assume that an investigator has chosen to use the PH regression approach, has exercised due diligence and is satisfied that the proportional hazards assumption is plausible. We now pick up the story at this point and ask the question, What additional insights are offered by embedding the PH regression results in a TR context? The answer begins by considering the shape of the proportional hazard functions and studying the possible TR models that might have generated the family. Probing the causal forces behind the hazard function is a worthwhile endeavor in its own right.

Aalen and Gjessing (2001, p. 1), make the telling point that the “hazard rate is really an elusive concept, especially when one tries to interpret its shape considered as a function of time.” These authors are highlighting the important point that the hazard function is only a derivative feature that may lie on the pathway to understanding but is not the end of the journey itself. A deeper understanding will be obtained if the risk mechanism behind the hazard pattern is probed by the investigator. The TR model provides an investigator with a general conceptual framework for this kind of probe.

To illustrate the issue, suppose a disease is characterized by recurrent infections that require medical intervention (such as treatment with an antibiotic). The factors affecting the recurrence time for an infection can be studied by PH regression and results may suggest that the PH assumption is acceptable. There are (at least) two competing models for the occurrence of the infections. The first model assumes that infections for a patient arise from random exposures to the infective organism (a Poisson-type process). The second model assumes that the infection results when the preceding infection, that is already established in the patient, begins to rebound and gradually reaches a threshold where intervention is called for (a deterioration process, such as Brownian motion with drift). These two models can produce similar looking families of hazard functions for the recurrence intervals and these may appear almost proportional. Yet, the causal dynamics are clearly not the same. To distinguish the two models, clinicians would need to monitor infection levels over time for a patient. It may happen, in fact, that both models are correct, but in different circumstances or on different occasions - sometimes a new infection, sometimes a worsening of an existing infection. Studying PH regression results alone, however, does not encourage a search for greater insight and, in any case, does not provide the requisite insight.

We now present a case illustration that demonstrates the importance of looking beyond PH in investigating event time data and, in particular, the kinds of insights that can be obtained from an application of TR.

6 Case illustration

This illustration compares TR and PH regression results for two scenarios based on a simulated randomized clinical trial with two study arms and $n = 200$ subjects on each arm. The applicable TR model in this simulation is a Wiener diffusion model (see, for example, Lee and Whitmore 2006). The response variable on each arm is an inverse Gaussian survival time S with two parameters, namely, an initial health level y_0 and mean health change μ . Survival time is assumed to be measured in, say, years. The scenarios have no censoring. The regression structure for the TR model adopts a

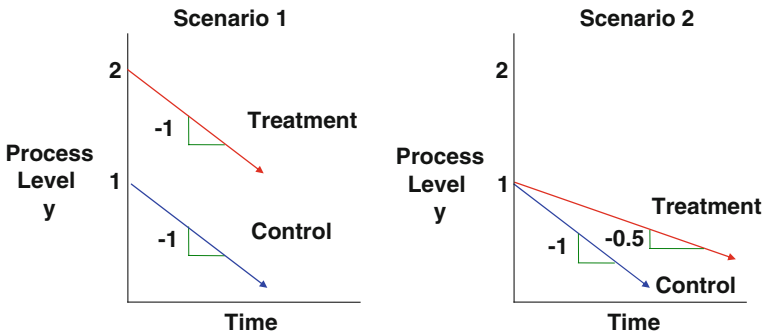


Fig. 6 Stylized description of two scenarios for a simulated clinical trial with two arms. Scenario 1 has y_0 twice as large on the treatment arm as on the control arm (2 versus 1), with a common value of μ (equal to -1). Scenario 2 has a value for μ that is half as large on the treatment arm as on the control arm (-0.5 versus -1) with a common value of y_0 (equal to 1)

logarithmic link function $\ln(y_0) = \alpha_0 + \alpha_1 \text{arm}$ for y_0 and an identity link function $\mu = \beta_0 + \beta_1 \text{arm}$ for μ , where arm is an indicator variable with values 0 and 1 for the control and treatment arms, respectively.

Scenario 1 in our simulation has y_0 twice as large on the treatment arm as on the control arm, with a common value of μ . Thus, the treatment effects for this scenario are $\alpha_1 = \ln(2) = .6931$ and $\beta_1 = 0$. Scenario 2 has a value for μ that is half as large on the treatment arm as on the control arm with a common value of y_0 . Thus, the treatment effects for this scenario are $\alpha_1 = \ln(1) = 0$ and $\beta_1 = 0.5$. Figure 6 illustrates the two scenarios graphically. The mean survival time on an arm is given by $E(S) = y_0/|\mu|$. Thus, both scenarios have the same mean survival times on corresponding arms, namely, a mean of 1 year on the control arm and 2 years on the treatment arm. The difference in the two scenarios is that the first scenario describes a medical context in which the treatment doubles initial health level y_0 of a patient at the outset of the study (relative to control) but leaves the rate of decline of health μ unchanged on both study arms. Such a treatment effect might correspond, for example, to a favorable surgical intervention carried out at the outset of the study, such as a successful heart transplant. The second scenario describes a context in which the treatment slows the rate of decline in health μ to half its previous value from the start of the study but leaves the initial health level y_0 unchanged on both study arms. Such a treatment effect might be found, for example, with ongoing drug therapy for a debilitating chronic disease, such as heart disease.

Table 1 presents TR and PH regression results for the two scenarios. The TR results show the true treatment effects α_1 and β_1 for each of the TR model parameters $\ln(y_0)$ and μ and the corresponding estimates, together with P -values. The PH results show the hazard ratio, its P -value, as well as the P -value for a global test of the PH property. Of course, by construction, the TR model is the true model.

In scenario 1, an analyst using PH regression would discover a significant hazard ratio of .4579 with P -value $<.0001$, indicating that the treatment risk is 46% of the control risk. The global test of the PH property, however, shows that the PH assumption is untenable (P -value $<.0001$). The analyst would then need to proceed to some other model. If the analyst happened upon the true TR model (possibly following

Table 1 A comparison of TR and PH regression results for two scenarios based on a simulated randomized clinical trial with two arms and $n = 200$ cases on each arm

Proportional hazards			Threshold regression			
Haz. ratio	<i>P</i> -value	PH test	Reg. Coeff.	True	Est.	<i>P</i> -value
<i>Scenario 1^a</i>						
.4579	<.0001	<.0001	α_1 in $\ln(y_0)$.6931	.7752	<.0001
			β_1 in μ	0	-.1247	.252
<i>Scenario 2^b</i>						
.5581	<.0001	.458	α_1 in $\ln(y_0)$	0	.0844	.232
			β_1 in μ	.5	.4282	<.0001

The response variable is an inverse Gaussian survival time with two parameters: initial health level y_0 and mean health change μ . The scenarios have no censoring. Scenario 1 has y_0 twice as large on the treatment arm as on the control arm, with a common value of μ . Scenario 2 has μ half as large on the treatment arm as on the control arm with a common value of y_0 . Both scenarios have the same mean survival time

^a Control arm: $y_0 = 1$, $\mu = -1.0$, Treatment arm: $y_0 = 2$, $\mu = -1.0$

^b Control arm: $y_0 = 1$, $\mu = -1.0$, Treatment arm: $y_0 = 1$, $\mu = -0.5$

discussions with the principal investigator) then the TR regression results shown in the top righthand panel of Table 1 would be obtained. The table shows a significant positive treatment effect for $\ln(y_0)$ but an insignificant effect for μ (as expected). Thus, importantly, the analyst would see that the treatment elevates the initial health of a subject but leaves the rate of subsequent health decline unchanged. The significant coefficient for α_1 translates into an estimated $\exp(.7752) = 2.2$ multiplier for initial health level y_0 (the true value is $\exp(.6931) = 2$).

In scenario 2, an analyst using PH regression would discover a significant hazard ratio of .5581, indicating that the treatment risk is 56% of the control risk. The global test of the PH property shows that the PH assumption is tenable (*P*-value .458). The analyst might then feel confident in reporting the finding. The principal investigator, however, is not much enlightened by this finding. A significant treatment effect has been found but the question of how the treatment is acting on the subject is not known from the PH regression finding. Alternatively, if the analyst happened to employ the true TR model then the regression results in the lower righthand panel of Table 1 show a significant positive treatment effect for μ but an insignificant effect for $\ln(y_0)$. Thus, the analyst would see that the treatment slows the decline in health but has an insignificant effect on the initial health level $\ln(y_0)$. The significant coefficient for β_1 translates into an estimated .4282 rise in the value of μ . The investigator would thereby have these extra insights into the treatment effect. These insights, when combined with medical knowledge, would presumably deepen the investigator's understanding of the treatment mechanism.

The differing results for the two scenarios can be anticipated from the plots in Fig. 7. The plot in the lefthand panel for each scenario shows the true inverse Gaussian hazard functions for the two study arms. The righthand plot in each scenario shows the natural logarithm of the ratio of the hazard functions for the two arms. For scenario 1, the two hazard functions converge to the same hazard level and the log-hazard ratio declines sharply to zero. The proportional hazard assumption is clearly untenable, as the PH test

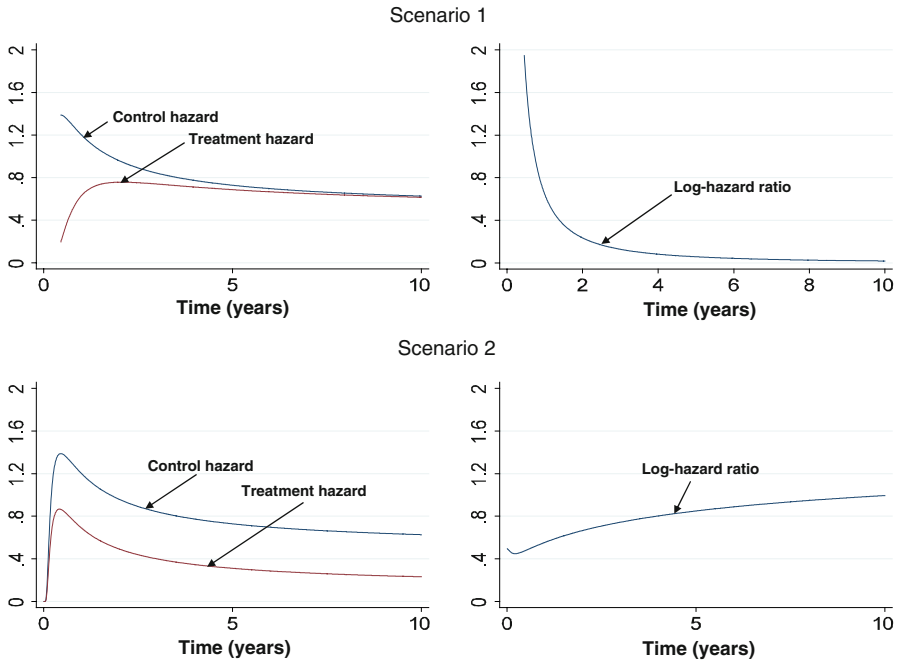


Fig. 7 Plots of the true inverse Gaussian hazard functions for the two study arms (the lefthand panel for each scenario) as well as plots of the log-hazard ratio (the righthand panel for each scenario)

result in Table 1 shows. The rapid convergence of the hazard functions indicates that the gains in risk from treatment come early in this scenario and gradually disappear. At about five years, the treatment and control patients face indistinguishable risks. For scenario 2, the two hazard functions are much closer to being proportional, although their log-hazard ratio more than doubles (on the log-scale) within the time span of the graph. The global test for the PH property does not detect this non-proportionality with the sample sizes involved in the simulation.

7 Closing remark

We have studied the theoretical connections between TR and PH regression methodologies to show that TR regression can encompass the PH property. Our wish is to encourage researchers to consider TR as an alternative regression approach when there is evidence or suspicions that the PH property does not hold. Even where the PH feature does hold reasonably well, we recommend that investigators use a TR modeling framework to delve into the hazard structure to understand better the patterns of risk being exhibited.

TR is a relatively new approach to analyzing event-time data but is gradually being applied in a wide assortment of medical and health studies. In support of this expansion, researchers are taking up theoretical and methodological investigations of TR in order to understand the model better and to offer a richer assortment of tools that are needed in practical statistical work. Recent publications include [Whitmore and](#)

Su (2007) who use TR to model low birth weights, Tong et al. (2008) who consider a bivariate TR model for joint analysis of current status and marker data, and Lee et al. (2008) who use a mixture version of TR to model data from a multiple myeloma clinical trial. More recently, Balka et al. (2009) implement cure models based on first hitting times for Wiener processes, Lee et al. (2009a) analyze occupational exposures to diesel exhaust using a TR model, Yu et al. (2009) investigate semi-parametric variants of TR, and Pennell et al. (2009) incorporate random effects in TR.

Acknowledgments This research is supported in part by CDC/NIOSH grant OH008649 (Lee) and by a research grant from the Natural Sciences and Engineering Research Council of Canada (Whitmore). The authors acknowledge the helpful comments of an associate editor and two reviewers that have greatly improved the manuscript. The authors would also like to thank Tao Xiao for preparing Figure 1.

Appendix

We derive the overshoot density $q(u)$ in (18) for a gamma process using limiting arguments in the theory of renewal processes. In brief, consider any partition of the time interval $[0, t]$ into n small increments $\Delta t > 0$ so $t = n\Delta t$. The corresponding gamma increments $Y_j(\Delta t)$ for time increments $j = 1, \dots, n$, form a renewal process with independent and identically distributed renewal intervals. We assume that the gamma process $\{Y(t)\}$ starts in equilibrium at $t = 0$. The excess $U > 0$ of the renewal interval that contains the first hitting time of boundary $b(s|z)$ has the following c.d.f. (see, for example, Ross 1996, p. 116):

$$Q_{\Delta t}(u) = \frac{1}{\mu} \int_0^u \bar{G}_{\Delta t}(h) dh \quad \text{for } u > 0, \tag{21}$$

where $\bar{G}_{\Delta t}(h)$ is the complementary c.d.f. of the gamma renewal interval for a time increment Δt . The notation $\mu = \alpha \Delta t$ is the mean length of the renewal interval. Taking the derivative of (21), the corresponding p.d.f. $q_{\Delta t}(u)$ is found to be:

$$q_{\Delta t}(u) = \frac{1}{\mu} \bar{G}_{\Delta t}(u) = \frac{1}{\alpha \Delta t} \int_u^\infty \frac{h^{\alpha \Delta t - 1}}{\Gamma(\alpha \Delta t)} \exp(-h) dh. \tag{22}$$

Taking the limit as Δt goes to zero and noting that $\lim_{a \rightarrow 0+} a\Gamma(a) = 1$, we obtain the limiting p.d.f.:

$$q(u) = \int_u^\infty \frac{1}{h} \exp(-h) dh \quad \text{for } u > 0. \tag{23}$$

Observe that this function is not defined at zero but it does integrate to 1 in the following limiting sense:

$$\lim_{a \rightarrow 0^+} \int_a^{\infty} q(u) du = 1. \quad (24)$$

References

- Aalen OO, Gjessing HK (2001) Understanding the shape of the hazard rate: a process point of view. *Stat Sci* 16:1–22
- Aalen OO, Borgan O, Gjessing HK (2008) *Survival and event history analysis: a process point of view (statistics for biology and health)*. Springer, New York
- Balka J, Desmond AF, McNicholas PD (2009) Review and implementation of cure models based on first hitting times for Wiener processes. *Lifetime Data Anal* 15:147–176
- Beran R (1981) *Nonparametric regression with randomly censored survival data*. Technical Report. University of California, Berkeley, CA
- Cox DR (1972) Regression models and life tables (with discussion). *J R Stat Soc Ser B* 34:187–230
- Cox DR, Oakes D (1984) *Analysis of survival data*. Chapman and Hall, London
- Duchesne T, Lawless J (2000) Alternative time scales and failure time models. *Lifetime Data Anal* 6:157–179
- Duchesne T, Rosenthal JS (2003) On the collapsibility of lifetime regression models. *Adv Appl Prob* 35:755–772
- Klein JP, Moeschberger ML (2003) *Survival analysis: techniques for censored and truncated data*, 2nd edn. Springer-Verlag, New York
- Kordonsky KB, Gertsbakh I (1997) Multiple time scales and the lifetime coefficient of variation: engineering applications. *Lifetime Data Anal* 3:139–156
- Lee M-LT (2009) Personal research webpage for TR and other software. <http://sph.umd.edu/epib/faculty/mltlee/index.html>
- Lee M-LT, Whitmore GA (2006) Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Stat Sci* 21:501–513
- Lee M-LT, Whitmore GA, Chang M (2008) A threshold regression mixture model for assessing treatment efficacy in a multiple myeloma clinical trial. *J Biopharm Stat* 18:1136–1149
- Lee M-LT, Whitmore GA, Laden F, Hart JE, Garshick E (2009a) A case-control study relating railroad worker mortality to diesel exhaust exposure using a threshold regression model. *J Stat Plan Infer* 139:1633–1642
- Lee M-LT, Whitmore GA, Rosner B (2009b) Benefits of threshold regression: a case-study comparison with Cox proportional hazards regression (submitted, under revision)
- Lee M-LT, Whitmore GA, Rosner B (2009c) Threshold regression for survival data with time-varying covariates. *Stat Med* (accepted, in press)
- Nahman NSJr, Middendorf DF, Bay WH, McElligott R, Powell S, Anderson J (1992) Modification of the percutaneous approach to peritoneal dialysis catheter placement under peritoneoscopic visualization: Clinical results in 78 patients. *J Am J Nephrol* 3:103–107
- Oakes D (1995) Multiple time scales in survival analysis. *Lifetime Data Anal* 1:7–18
- Pennell ML, Whitmore GA, Lee M-LT (2009) Bayesian random effects threshold regression with application to survival data with nonproportional hazards. *Biostatistics* (accepted, in press)
- Ross SM (1996) *Stochastic processes*, 2nd edn. Wiley, New York
- Tong X, He X, Sun J, Lee M-LT (2008) Joint analysis of current status and marker data: an extension of a bivariate threshold model. *Int J Biostat* 4(1):Article 21
- Whitmore GA (1984) Barrier estimation using first passage time data from Brownian motion. Working paper. McGill University, Montreal
- Whitmore GA (1986) First passage time models for duration data—regression structures and competing risks. *Statistician* 35:207–219
- Whitmore GA, Su Y (2007) Modeling low birth weights using threshold regression: results for U.S. birth data. *Lifetime Data Anal* 13:161–190
- Yu Z, Tu W, Lee M-LT (2009) A semiparametric threshold regression analysis of sexually transmitted infections in adolescent women. *Stat Med* (accepted, in press)