A General Class of Models for Recurrent Events

Edsel A. Pena
University of South Carolina at Columbia
[E-Mail: pena@stat.sc.edu]

ENAR Talk, Arlington, VA, 3/18/02

Research support from NIH, NSF

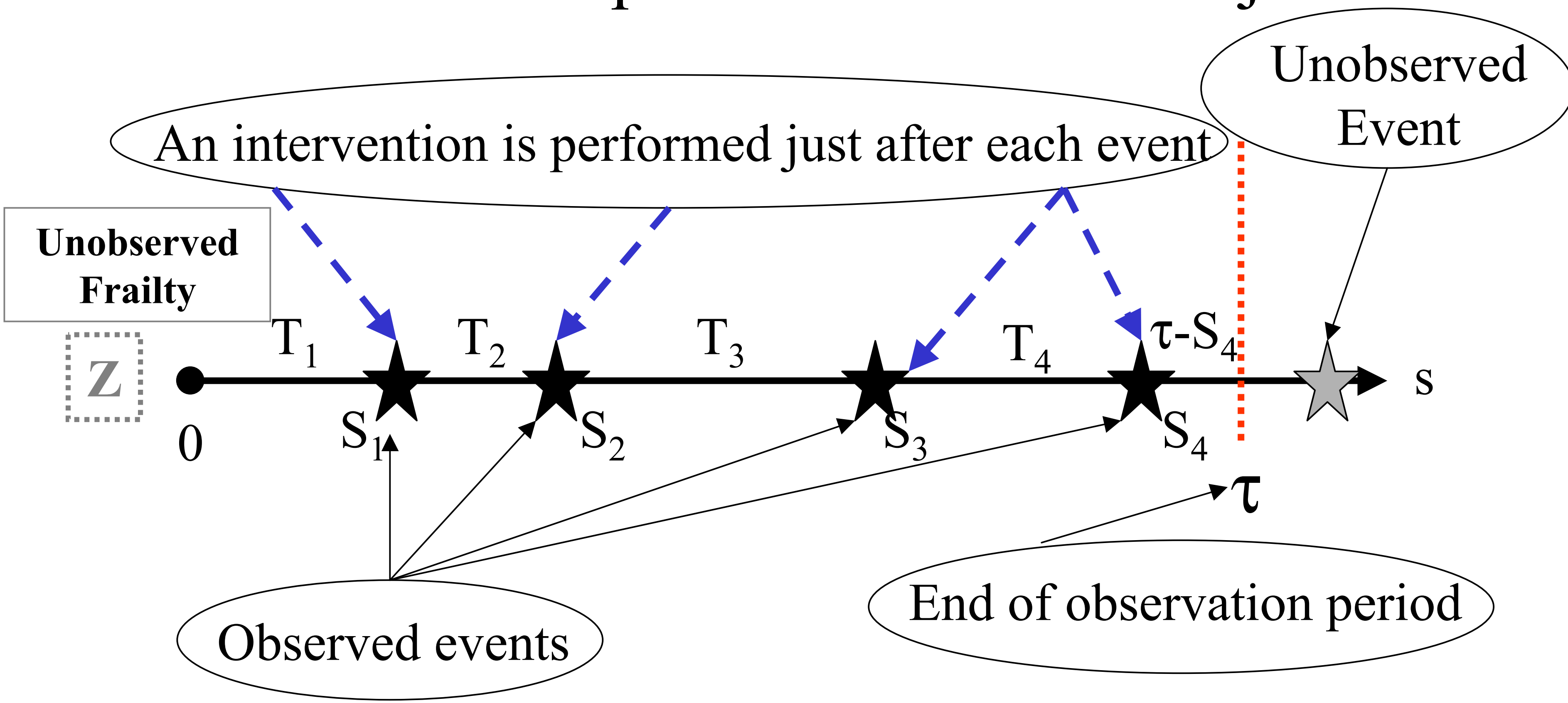
Recurrent Phenomena

Public Health and Medical Settings

- hospitalization of a subject with a chronic disease, e.g., end stage renal disease
- drug/alcohol abuse of a subject
- headaches
- tumor occurrence
- cyclic movements in the small bowel during fasting state
- depression
- episodes of epileptic seizures

Prevalent in other areas (reliability, economics, etc.) as well.

A Pictorial Representation: One Subject



An observable covariate vector: $\mathbf{X}(s) = (X_1(s), X_2(s), \ldots, X_q(s))^t$

Features in Recurrent Event Modeling

- Intervention effects after each event occurrence.
- Effects of accumulating event occurrences on the subject. Could be a weakening or an strengthening effect.
- Effects of possibly time-dependent covariates.
- Possible associations of event occurrences for a subject.
- A possibly random observation period per subject.
- Number of observable events per subject is random and is informative on stochastic mechanism generating events.
- Informative right-censoring mechanism for the inter-event time that covers end of observation period.

Random Entities: One Subject

- $\mathbf{X}(s)$ = covariate vector, possibly time-dependent
- $T_1, T_2, T_3, \ldots$ = the inter-event or gap times
- $S_1, S_2, S_3, \ldots$ = the calendar times of occurrences
- $\mathbf{F}^+ = \{F_s^+ : 0 \leq s\}$ = filtration including info about interventions, covariate, etc. in [0, s]
- Z = unobserved frailty variable
- $N^+(s)$ = number of events observed on or before calendar time s
- $Y^+(s)$ = indicator of whether the subject is still at risk just before calendar time s

A General Class of Models

$\{A^+(s|Z): s \geq 0\}$ is a predictable non-decreasing process such that given Z and with respect to the filtration $\mathbf{F}^+$:

$$\left\{M^+(s \mid Z) = N^+(s) - A^+(s \mid Z):\ s \geq 0\right\}$$

is a square-integrable zero-mean (local) martingale. As in previous works (Aalen, Gill, Andersen and Gill, Nielsen, et al, others) we assume

$$A^+(s \mid Z) = \int_0^s Y^+(w)\lambda(w \mid Z)dw$$

Modeling the Intensity Process [Pena and Hollander, to appear]

Specify, possibly in a dynamic fashion, a predictable, observable process $\{E(s): 0 \leq s \leq \tau\}$, called the *effective age process*, satisfying

- $E(0) = e_0 \geq 0$;
- $E(s) \geq 0$ for every s;
- On $[S_{k-1}, S_k)$, $E(s)$ is monotone and differentiable with a nonnegative derivative.

Specification of the Intensity Process

$$\lambda(s \mid Z) = Z\lambda_0\left[E(s)\right]\rho\left[N^+(s-);\alpha\right]\psi\left[\beta^t X(s)\right]$$

Model Components

- $\lambda_0(.)$ = an unknown baseline hazard rate function, possibly parametrically specified.

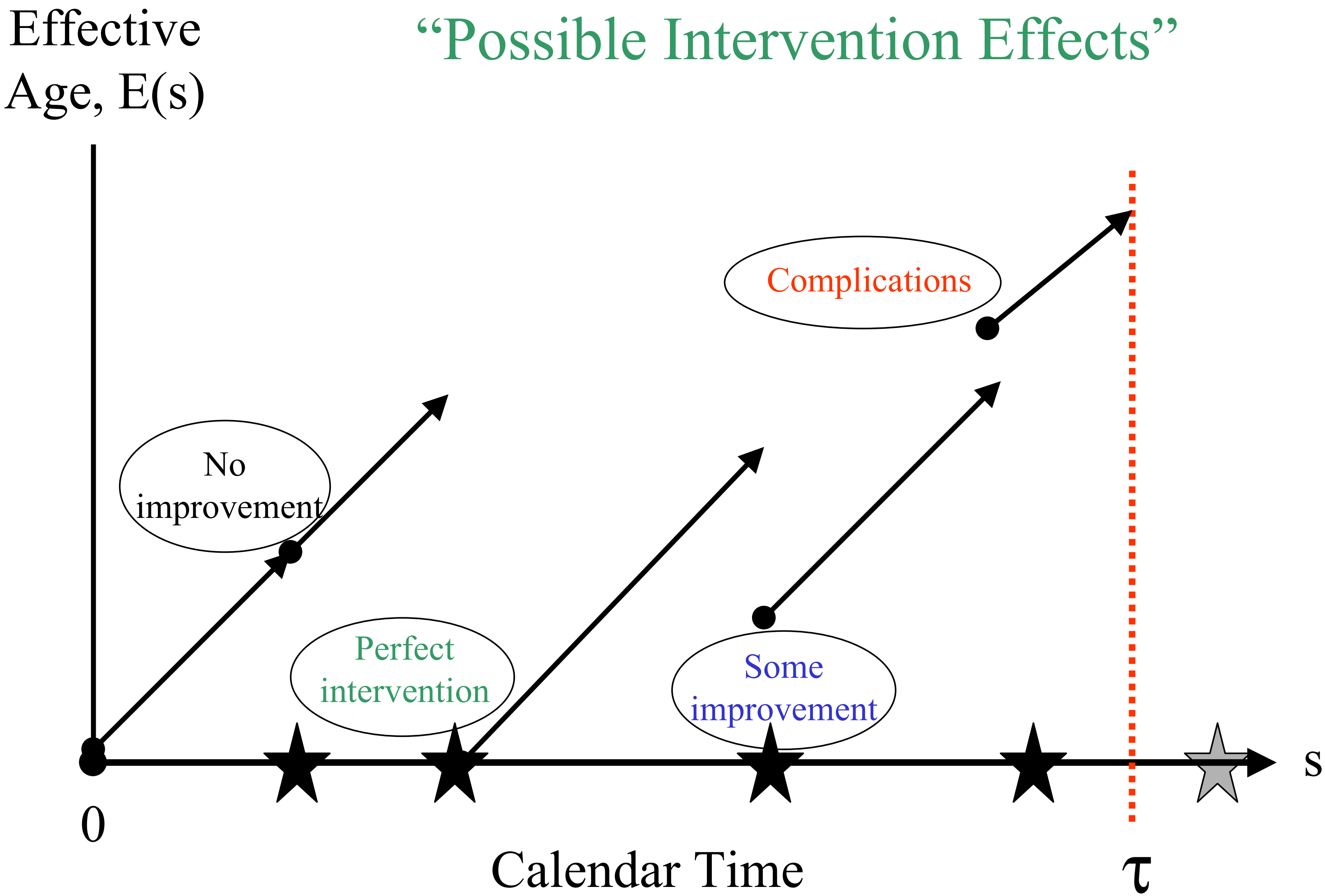
- E(s) = ***effective age*** of the subject at calendar time s. Idea is that a performed intervention changes the effective age of subject acting on the baseline hazard rate.

- $\rho(.;\alpha)$ = a +function on {0,1,2,…} of known form with $\rho(0;\alpha) = 1$ and with unknown parameter α. Encodes effect of accumulating event occurrences on the subject.

- $\psi(.)$ = positive link function containing the effect of subject covariates. β is unknown.

- Z = unobservable frailty variable, which when integrated out, induces associations among the inter-event times.

Illustration: Effective Age Process
"Possible Intervention Effects"



Special Cases of the Class of Models

- IID "Renewal" Model without frailties: Considered by Gill ('81 AS), Wang and Chang ('99, JASA), Pena, Strawderman and Hollander ('01, JASA).

$$\mathrm{E(s)} = \mathrm{s} - \mathrm{S}_{\mathrm{N}^{+}(s-)}; Z = 1; \rho(k;\alpha) = 1; \psi(\mathrm{w}) = 1.$$

- IID "Renewal" Model with frailties: Considered by Wang and Chang ('99), PSH ('01).

$$\mathrm{E(s)} = \mathrm{s} - \mathrm{S}_{\mathrm{N}^{+}(s-)}; Z \sim Ga(\gamma,\gamma); \rho(k;\alpha) = 1; \psi(\mathrm{w}) = 1.$$

Generality and Flexibility

• Extended Cox PH Model: Considered by Prentice, Williams, and Petersen (PWP) ('81); Lawless ('87), Aalen and Husebye ('91).

$$E(s) = s - S_{N^+(s-)}; Z = 1; \rho(k;\alpha) = 1; \psi(w) = exp(w).$$

• Also by PWP ('81), Brown and Proschan ('83) and Lawless ('87) called a "minimal repair model" in the reliability literature.

$$E(s) = s; Z = 1; \rho(k;\alpha) = 1; \psi(w) = 1.$$

• A generalized Gail, Santner and Brown ('80) tumor occurrence model and Jelinski and Moranda ('72) software reliability model:

$$\mathrm{E(s)} = \mathrm{s} - \mathrm{S}_{\mathrm{N}^{+}(s-)}; Z = 1; \rho(k;\alpha) = \alpha - k + 1; \psi(\mathrm{w}) = \exp(w).$$

• Let $I_1, I_2, I_3, \ldots$ be IND Ber[p(s)] rvs and $\eta(s) = \sum_{i=1}^{N^{+}(s)} I_i.$

Let $\Gamma_k = \min\{j > \Gamma_{k-1} : I_j = 1\}$. If

$$E(s) = s - S_{\Gamma_{\eta(s-)}}$$

we generalize the BP ('83) and Block, Borges and Savits ('85) minimal repair model. Also considered in Presnell, Hollander and Sethuraman ('94, '97).

Other Models In Class

• Dorado, Hollander and Sethuraman (‘97), Kijima (‘89), Baxter, Kijima and Tortorella (‘96), Stadje and Zuckerman (‘91), and Last and Szekli (‘98):

$$\{A_j : j = 0,1,2,\ldots\} \text{ and } \{\Theta_j : j = 0,1,2,\ldots\}$$
$$E(s) = A_{N^+(s-)} + \Theta_{N^+(s-)}\left[s - S_{N^+(s-)}\right]$$

• Two simple forms for the ρ function:

$$\rho(k;\alpha) = \alpha^k; \quad \rho(k;\alpha) = \max\{\alpha - g(k), 0\}$$

α = initial measure of “defectiveness” or event “proneness.”

Relevance

• Flexibility and generality of this class of models will allow better modeling of observed phenomena, and allow testing of specific/special models using this *larger* class.

• **Question: Is this relevant in biostatistical modeling??**

• **Answer:** The ***fact*** that it contains models currently being used indicates the model's importance.

• **However**, a ***"paradigm shift"*** is needed in the data gathering since the model requires the assessment of the effective age.

• **But**, this could be provided by the medical or public health experts after each intervention.

On the Model's *Immediate* Applicability

Most often it is the case of

"A Data in Search of a Model;"

but, sometimes* as in this case, it is

"A Model in Search of a Data!"

*A modern example of such a situation is that which led to the 1919 Eddington expedition.

Some Issues on Inference

- Need to take into account the sum-quota data accrual scheme which leads to an informative random number of events and informative right-censoring (cf., PSH (‘01) in renewal model).

Example: Variances of EDF, PLE, and GPLE

$$\text{EDF}: \quad v_1(t) = F(t)\overline{F}(t) = \overline{F}(t)^2 \int_0^t \frac{d\Lambda(w)}{\overline{F}(w)}$$

$$\text{PLE}: \quad v_2(t) = \overline{F}(t)^2 \int_0^t \frac{d\Lambda(w)}{\overline{F}(w)\overline{G}(w)}$$

For GPLE in “Renewal (IID) Model” [PSH ‘01, JASA]:

$$v_3(t) = \bar{F}(t)^2 \int_0^t \frac{\mathrm{d}\Lambda(w)}{\bar{F}(w)\bar{G}(w)\left\{1 + \frac{1}{\bar{G}(w)} \int_w^\infty R(u-w)\mathrm{d}G(u)\right\}}$$

Identifiability: Model without Frailty

If for each $(\lambda_0(.),\alpha,\beta)$, the support of $E(T_1)$ contains $[0, \tau]$, and if $\rho(.;.)$ satisfies

$$\forall k \in \{0,1,2,\ldots\}, [\rho(k;\alpha^{(1)}) = \rho(k;\alpha^{(2)})] \Rightarrow \{\alpha^{(1)} = \alpha^{(2)}\},$$

then the statistical model is identifiable.

Other Statistical Issues

- Parameter Estimation, especially when baseline hazard is non-parametrically specified. In progress!
- Testing and Group Comparisons.
- Model Validation and Diagnostics.