Modeling and Non- and Semi-Parametric Inference with Recurrent Event Data

Edsel A. Peña

Department of Statistics
University of South Carolina
Columbia, SC 29208
Talk at Columbia University
November 19, 2007
Some (Recurrent) Events

- Submission of a manuscript for publication.
- Occurrence of tumor.
- Onset of depression.
- Patient hospitalization.
- Machine/system failure.
- Occurrence of a natural disaster.
- Non-life insurance claim.
- Change in job.
- Onset of economic recession.
- At least a 200 points decrease in the DJIA.
- Marital disagreement.
Event Times and Distributions

- $T$: the time to the occurrence of an event of interest.
- $F(t) = \Pr\{T \leq t\}$: the distribution function of $T$.
- $S(t) = \bar{F}(t) = 1 - F(t)$: survivor/reliability function.
- Hazard rate/ probability and Cumulative Hazards:

  **Cont:** \[ \lambda(t)dt \approx \Pr\{T \leq t + dt|T \geq t\} = \frac{f(t)}{S(t-)}dt \]

  **Disc:** \[ \lambda(t_j) = \Pr\{T = t_j|T \geq t_j\} = \frac{f(t_j)}{S(t_j-)} \]

  **Cumulative:** \[ \Lambda(t) = \int_0^t \lambda(w)dw \quad \text{or} \quad \Lambda(t) = \sum_{t_j \leq t} \lambda(t_j) \]
0 < t_1 < \ldots < t_M = t, M(t) = \max |t_i - t_{i-1}| = o(1),

\[ S(t) = \Pr\{T > t\} = \prod_{i=1}^{M} \Pr\{T > t_i | T \geq t_{i-1}\} \approx \prod_{i=1}^{M} \left[ 1 - \{\Lambda(t_i) - \Lambda(t_{i-1})\} \right] . \]

\( S \) as a product-integral of \( \Lambda \): When \( M(t) \to 0 \),

\[ S(t) = \prod_{w \leq t} \left[ 1 - \Lambda(dw) \right] \]

In general, \( \Lambda \) in terms of \( F \): \( \Lambda(t) = \int_0^t \frac{dF(w)}{1 - F(w-)} \).
Estimation of $F$ and Why?

Most Basic Problem: Given a sample $T_1, T_2, \ldots, T_n$ from an unknown distribution $F$, to obtain an estimator $\hat{F}$ of $F$.

Why is it important to know how to estimate $F$?
- Functionals/parameters $\theta(F)$ of $F$ (e.g., mean, median, variance) can be estimated via $\hat{\theta} = \theta(\hat{F})$.
- Prediction of time-to-event for new units.
- Knowledge of population of units or event times.
- For comparing groups, e.g., thru a statistic

$$Q = \int W(t) d \left[ \hat{F}_1(t) - \hat{F}_2(t) \right]$$

where $W(t)$ is some weight function.
Gastroenterology Data: Aalen and Husebye (’91)

Migratory Motor Complex (MMC) Times for 19 Subjects

Question: How to estimate the MMC period dist, $F$?
Parametric Approach

- Unknown df $F$ is assumed to belong to some parametric family (e.g., exponential, gamma, Weibull)

$$\mathcal{F} = \{ F(t; \theta) : \theta \in \Theta \subset \mathbb{R}^p \}$$

with functional form of $F(\cdot; \cdot)$ known; $\theta$ is unknown.

- Based on data $t_1, t_2, \ldots, t_n$, $\theta$ is estimated by $\hat{\theta}$, say, via maximum likelihood (ML). $\hat{\theta}$ maximizes likelihood

$$L(\theta) = \prod_{i=1}^{n} f(t_i; \theta) = \prod_{i=1}^{n} \lambda(t_i; \theta) \exp\{-\Lambda(t_i; \theta)\}.$$

- DF $F$ estimated by: $\hat{F}_{pa}(t) = F(t; \hat{\theta})$. 
Parametric Estimation: Asymptotics

- When $\mathcal{F}$ holds, MLE of $\theta$ satisfies

$$\hat{\theta} \sim \text{AN} \left( \theta, \frac{1}{n} \mathcal{I}(\theta)^{-1} \right);$$

$$\mathcal{I}(\theta) = \text{Var} \{ \frac{\partial}{\partial \theta} \log f(T_1; \theta) \} = \text{Fisher information.}$$

- Therefore, when $\mathcal{F}$ holds, by $\delta$-method, with

$$\dot{F}(t; \theta) = \frac{\partial}{\partial \theta} F(t; \theta)$$

then

$$\hat{F}_{pa}(t) \sim \text{AN} \left( F(t; \theta), \frac{1}{n} \dot{F}(t; \theta)' \mathcal{I}(\theta)^{-1} \dot{F}(t; \theta) \right).$$
Nonparametric Approach

- **No assumptions** are made regarding the family of distributions to which the unknown df $F$ belongs.

- Empirical Distribution Function (EDF):

  \[ \hat{F}_{np}(t) = \frac{1}{n} \sum_{i=1}^{n} I\{T_i \leq t\} \]

- $\hat{F}_{np}(\cdot)$ is a **nonparametric** MLE of $F$.

- Since $I\{T_i \leq t\}, i = 1, 2, \ldots, n$, are IID $\text{Ber}(F(t))$, by Central Limit Theorem,

  \[ \hat{F}_{np}(t) \sim \text{AN} \left( F(t), \frac{1}{n} F(t)[1 - F(t)] \right). \]
An Efficiency Comparison

- Assume that family $\mathcal{F} = \{ F(t; \theta) : \theta \in \Theta \}$ holds. Both $\hat{F}_{pa}$ and $\hat{F}_{np}$ are asymptotically unbiased.

- To compare under $\mathcal{F}$, we take ratio of asymptotic variances to give the efficiency of parametric estimator over nonparametric estimator.

\[
\text{Eff}(\hat{F}_{pa}(t) : \hat{F}_{np}(t)) = \frac{F(t; \theta)[1 - F(t; \theta)]}{\hat{F}(t; \theta)'\mathcal{I}(\theta)^{-1} \hat{F}(t; \theta)}.
\]

- When $\mathcal{F} = \{ F(t; \theta) = 1 - \exp\{-\theta t\} : \theta > 0 \}$, then

\[
\text{Eff}(\hat{F}_{pa}(t) : \hat{F}_{np}(t)) = \frac{\exp\{\theta t\} - 1}{(\theta t)^2}.
\]
Efficiency: Parametric/Nonparametric

Asymptotic efficiency of parametric versus nonparametric estimators under a correct negative exponential family model.

Effi of Para Relative to NonPara in Expo Case

![Graph showing efficiency of parametric relative to nonparametric estimators in the exponential case.](image-url)
Whither Nonparametrics?

Consider however the case where the negative exponential family is fitted, but it is actually not the correct model. Let us suppose that the gamma family of distributions is the correct model.

Under wrong model, with \( \bar{T} = \frac{1}{n} \sum_{i=1}^{n} T_i \) the sample mean, the parametric estimator of \( F \) is

\[
\hat{F}_{pa}(t) = 1 - \exp\{-t/\bar{T}\}.
\]

Under gamma with shape \( \alpha \) and scale \( \theta \), and since \( \bar{T} \sim AN(\alpha/\theta, \alpha/(n\theta^2)) \), by \( \delta \)-method

\[
\hat{F}_{pa}(t) \sim AN \left( 1 - \exp \left\{ -\frac{\theta t}{\alpha} \right\}, \frac{1}{n} \frac{(\theta t)^2}{\alpha^3} \exp \left\{ -\frac{2(\theta t)}{\alpha} \right\} \right).
\]
Efficiency: Under a Mis-specified Model

Simulated Effi: $\text{MSE(Non-Parametric)}/\text{MSE(Parametric)}$ under a \textit{mis-specified} exponential family model.

True Family of Model: Gamma Family
**MMC Data: Censoring Aspect**

For each unit, red mark is the potential termination time.

**Remark:** All 19 MMC times **completely** observed.
Estimation of $F$: With Censoring

- For $i$th unit, a right-censoring variable $C_i$ with $C_1, C_2, \ldots, C_n$ IID df $G$.
- Observables are $(Z_i, \delta_i), i = 1, 2, \ldots, n$ with $Z_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i \leq C_i\}$.
- **Problem:** For observed $(Z_i, \delta_i)$s, to estimate df $F$ or hazard function $\Lambda$ of the $T_i$s.

**Nonparametric Approaches:**
- Nonparametric MLE (Kaplan-Meier).
- Martingale and method-of-moments.

**Pioneers:** Kaplan & Meier; Efron; Nelson; Breslow; Breslow & Crowley; Aalen; Gill.
Product-Limit Estimator

Counting and At-Risk Processes:

\[ N(t) = \sum_{i=1}^{n} I\{Z_i \leq t; \delta_i = 1\}; \]
\[ Y(t) = \sum_{i=1}^{n} I\{Z_i \geq t\} \]

Hazard probability estimate at \( t \):

\[ \hat{\Lambda}(dt) = \frac{\Delta N(t)}{Y(t)} = \frac{\text{# of Observed Failures at } t}{\text{# at-risk at } t} \]

Product-Limit Estimator (PLE):

\[ 1 - \hat{F}(t) = \hat{S}(t) = \prod_{w \leq t} \left[ 1 - \frac{\Delta N(t)}{Y(t)} \right] \]
Some Properties of PLE

- Nonparametric MLE of $F$ (Kaplan-Meier, ’58).
- PLE is a step-function which jumps only at observed failure times.
- With censored data, unequal jumps.
- Biased for finite $n$.
- When no censoring and no tied values: $\Delta N(t_{(i)}) = 1$ and $Y(t_{(i)}) = n - i + 1$, so

$$\hat{S}(t_{(i)}) = \prod_{j=1}^{i} \left[ 1 - \frac{1}{n - j + 1} \right] = 1 - \frac{i}{n}.$$
Stochastic Process Approach

- A martingale $M$ is a zero-mean process which models a fair game. With $\mathcal{H}_t =$ history up to $t$:

$$E\{M(s + t)|\mathcal{H}_t\} = M(t).$$

- $M(t) = N(t) - \int_0^t Y(w)\Lambda(dw)$ is a martingale, so with $J(t) = I\{Y(t) > 0\}$ and stochastic integration,

$$E\left\{\int_0^t \frac{J(w)}{Y(w)}dN(w)\right\} = E\left\{\int_0^t J(w)\Lambda(dw)\right\}.$$

- Nelson-Aalen estimator of $\Lambda$, and PLE:

$$\hat{\Lambda}(t) = \int_0^t \frac{dN(w)}{Y(w)}, \quad \text{so} \quad \hat{S}(t) = \prod_{w \leq t} [1 - \hat{\Lambda}(dw)].$$
Likelihood Process: Hazard-Based

- J. Jacod’s likelihood:

\[
L_t(\Lambda(\cdot)) = \prod_{w \leq t} [Y(w)\Lambda(dw)]^N(dw) [1 - Y(w)\Lambda(dw)]^{1-N(dw)}.
\]

- When \( \Lambda(\cdot) \) is continuous,

\[
L_t(\Lambda(\cdot)) = \left\{ \prod_{w \leq t} [Y(w)\Lambda(dw)]^N(dw) \right\} e^{-\int_0^t Y(w)\Lambda(dw)}.
\]

- With \( T(t) = \int_0^t Y(w)dw = \text{TTOT}(t) \), for \( \lambda(t) = \theta \),

\[
L_t(\theta) = \theta^N(t) \exp\{-\theta T(t)\}.
\]
Asymptotic Properties

NAE: \( \sqrt{n} [\hat{\Lambda}(t) - \Lambda(t)] \Rightarrow Z_1(t) \) with \( \{Z_1(t) : t \geq 0\} \) a zero-mean Gaussian process with

\[
d_1(t) = \text{Var}(Z_1(t)) = \int_0^t \frac{\Lambda(dw)}{S(w)\bar{G}(w^-)}.
\]

PLE: \( \sqrt{n} [\hat{F}(t) - F(t)] \Rightarrow Z_2(t) \overset{st}{=} S(t)Z_1(t) \) so

\[
d_2(t) = \text{Var}(Z_2(t)) = S(t)^2 \int_0^t \frac{\Lambda(dw)}{S(w)\bar{G}(w^-)}.
\]

If \( \bar{G}(w) \equiv 1 \) (no censoring), \( d_2(t) = F(t)S(t) \)!
Regression Models

- **Covariates:** temperature, degree of usage, stress level, age, blood pressure, race, etc.

- How to account of covariates to improve knowledge of time-to-event.

- Modelling approaches:
  - Log-linear models:
    \[ \log(T) = \beta'x + \sigma \epsilon. \]
    The **accelerated failure-time model**. Error distribution to use? Normal errors **not** appropriate.
  - **Hazard-based models:** Cox proportional hazards (PH) model; Aalen’s additive hazards model.
Cox ('72) PH Model: Single Event

- Conditional on $x$, hazard rate of $T$ is:

$$\lambda(t|x) = \lambda_0(t) \exp\{\beta'x\}.$$

- $\hat{\beta}$ maximizes partial likelihood function of $\beta$:

$$L_P(\beta) \equiv \prod_{i=1}^{n} \prod_{t<\infty} \left[ \frac{\exp(\beta'x_i)}{\sum_{j=1}^{n} Y_j(t) \exp(\beta'x_j)} \right] \Delta N_i(t).$$

- Aalen-Breslow semiparametric estimator of $\Lambda_0(\cdot)$:

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^{n} dN_i(w)}{\sum_{i=1}^{n} Y_i(w) \exp(\hat{\beta}'x_i)}.$$
 MMC Data: Recurrent Aspect

Aalen and Husebye (’91) Full Data

**Problem:** Estimate inter-event time distribution.
Recurrent Events: In Complex Systems

A Reliability Bridge Structure

\[ \phi(x_1, x_2, x_3, x_4, x_5) = x_1 x_3 x_5 \vee x_2 x_3 x_4 \vee x_1 x_4 \vee x_2 x_5 \]
System’s Dynamic Evolution

After Component 2 Has Failed: Series-Parallel

After Components 2 and 4 Have Failed: Series
Points to Ponder in Modeling

- System fails under certain component failure configurations (called cut sets).
- Recurrent event of interest are the successive component failure occurrences.
- Component failures **dynamically** change effective structure function. Originally a bridge system; then after #2 fails, it is a series-parallel system; then after #2 and #4 fail, it is a series system.
- Component failures change component loads (essence of load-sharing system).
- System failure time may right-censor some component failure times.
Representation: One Subject

Covariate vector: $X(s) = (X_1(s), \ldots, X_q(s))$
Observables: One Subject

- $\mathbf{X}(s) =$ covariate vector, possibly time-dependent
- $T_1, T_2, T_3, \ldots =$ inter-event or gap times
- $S_1, S_2, S_3, \ldots =$ calendar times of event occurrences
- $\tau =$ end of observation period: Assume $\tau \sim G$
- $K = \max\{k : S_k \leq \tau\} =$ number of events in $[0, \tau]$
- $Z =$ unobserved frailty variable
- $N^\dagger(s) =$ number of events in $[0, s]$
- $Y^\dagger(s) = I\{\tau \geq s\} =$ at-risk indicator at time $s$
- $\mathbf{F}^\dagger = \{\mathcal{F}_s^\dagger : s \geq 0\} =$ filtration: information that includes interventions, covariates, etc.
Aspect of Sum-Quota Accrual

Remark: A unique feature of recurrent event modeling is the sum-quota constraint that arises due to a fixed or random observation window. Failure to recognize this in the statistical analysis leads to erroneous conclusions.

\[
K = \max \left\{ k : \sum_{j=1}^{k} T_j \leq \tau \right\}
\]

\[
(T_1, T_2, \ldots, T_K) \text{ satisfies } \sum_{j=1}^{K} T_j \leq \tau < \sum_{j=1}^{K+1} T_j.
\]
Recurrent Event Models: IID Case

- **Parametric Models:**
  - **HPP:** $T_{i1}, T_{i2}, T_{i3}, \ldots$ IID $\text{EXP} (\lambda)$.
  - **IID Renewal Model:** $T_{i1}, T_{i2}, T_{i3}, \ldots$ IID $F$ where
    \[ F \in \mathcal{F} = \{ F(\cdot; \theta) : \theta \in \Theta \subset \mathbb{R}^p \}; \]
    e.g., Weibull family; gamma family; etc.

- **Non-Parametric Model:** $T_{i1}, T_{i2}, T_{i3}, \ldots$ IID $F$ which is some df.

- **With Frailty:** For each unit $i$, there is an *unobservable* $Z_i$ from some distribution $H(\cdot; \xi)$ and $(T_{i1}, T_{i2}, T_{i3}, \ldots)$, given $Z_i$, are IID with survivor function
  \[ [1 - F(t)]^{Z_i}. \]
A General Class of Full Models


\[ N^\dagger(s) = A^\dagger(s|Z) + M^\dagger(s|Z) \]

\[ M^\dagger(s|Z) \in \mathcal{M}_0^2 = \text{sq-int martingales} \]

\[ A^\dagger(s|Z) = \int_0^s Y^\dagger(w) \lambda(w|Z) \, dw \]

Intensity Process:

\[ \lambda(s|Z) = Z \lambda_0[\mathcal{E}(s)] \rho[N^\dagger(s-); \alpha] \psi[\beta^\top X(s)] \]
Effective Age Process: $\mathcal{E}(s)$
Effective Age Process, $\mathcal{E}(s)$

- **PERFECT** Intervention: $\mathcal{E}(s) = s - S_{N^+(s^-)}$.
- **IMPERFECT** Intervention: $\mathcal{E}(s) = s$.
- **MINIMAL** Intervention (BP ’83; BBS ’85):

  $$\mathcal{E}(s) = s - S_{\Gamma\eta(s^-)}$$

where, with \(I_1, I_2, \ldots\) IID BER(p),

$$\eta(s) = \sum_{i=1}^{N^+(s)} I_i$$ and $$\Gamma_k = \min\{j > \Gamma_{k-1} : I_j = 1\}.$$
Semi-Parametric Estimation: No Frailty

Observed Data for $n$ Subjects:

$$\{(X_i(s), N_{i}^{\dagger}(s), Y_{i}^{\dagger}(s), E_i(s)) : 0 \leq s \leq s^*\}, i = 1, \ldots, n$$

$N_{i}^{\dagger}(s) = \# \text{ of events in } [0, s] \text{ for } i\text{th unit}$

$Y_{i}^{\dagger}(s) = \text{at-risk indicator at } s \text{ for } i\text{th unit}$

with the model for the ‘signal’ being

$$A_{i}^{\dagger}(s) = \int_{0}^{s} Y_{i}^{\dagger}(v) \rho[N_{i}^{\dagger}(v-); \alpha] \psi[\beta^t X_i(v)] \lambda_0[E_i(v)] dv$$

where $\lambda_0(\cdot)$ is an unspecified baseline hazard rate function.
Processes and Notations

Calendar/Gap Time Processes:

\[ N_i(s, t) = \int_0^s I\{\mathcal{E}_i(v) \leq t\} N_i^\dagger (dv) \]

\[ A_i(s, t) = \int_0^s I\{\mathcal{E}_i(v) \leq t\} A_i^\dagger (dv) \]

Notational Reductions:

\[ \mathcal{E}_{ij-1}(v) \equiv \mathcal{E}_i(v) I_{(s_{ij-1}, s_{ij})} (v) I\{Y_i^\dagger (v) > 0\} \]

\[ \varphi_{ij-1}(w|\alpha, \beta) \equiv \frac{\rho(j - 1; \alpha) \psi\{\beta^t X_i [\mathcal{E}_{ij-1}^{-1}(w)]\}}{\mathcal{E}_{ij-1}' [\mathcal{E}_{ij-1}^{-1}(w)]} \]
For IID Renewal Model (PSH, 01) this simplifies to:

\[ Y_i(s, w|\alpha, \beta) = \sum_{j=1}^{N_i^+(s-)} I(T_{ij} \geq w) + I((s \wedge \tau_i) - S_{iN_i^+(s-)} \geq w) \]
Estimation of $\Lambda_0$

$$A_i(s, t|\alpha, \beta) = \int_0^t Y_i(s, w|\alpha, \beta)\Lambda_0(dw)$$

$$S_0(s, t|\alpha, \beta) = \sum_{i=1}^n Y_i(s, t|\alpha, \beta)$$

$$J(s, t|\alpha, \beta) = I\{S_0(s, t|\alpha, \beta) > 0\}$$

Generalized Nelson-Aalen ‘Estimator’:

$$\hat{\Lambda}_0(s, t|\alpha, \beta) = \int_0^t \left\{ \frac{J(s, w|\alpha, \beta)}{S_0(s, w|\alpha, \beta)} \right\} \left\{ \sum_{i=1}^n N_i(s, dw) \right\}$$
Estimation of $\alpha$ and $\beta$

- **Partial Likelihood (PL) Process:**

$$L_P(s^* | \alpha, \beta) = \prod_{i=1}^{n} N_i^\dagger(s^*) \prod_{j=1}^{N_i^\dagger(s^*)} \left[ \frac{\rho(j - 1; \alpha) \psi[\beta^t X_i(S_{ij})]}{S_0[s^*, E_i(S_{ij}) | \alpha, \beta]} \right] \Delta N_i^\dagger(S_{ij})$$

- **PL-MLE:** $\hat{\alpha}$ and $\hat{\beta}$ are maximizers of the mapping

$$(\alpha, \beta) \mapsto L_P(s^* | \alpha, \beta)$$

- **Iterative procedures. Implemented in an \texttt{R} package called \texttt{gcmrec} (Gonzaléz, Slate, Peña ’04).
Estimation of $\tilde{F}_0$

G-NAE of $\Lambda_0(\cdot)$: $\hat{\Lambda}_0(s^*, t) \equiv \hat{\Lambda}_0(s^*, t|\hat{\alpha}, \hat{\beta})$

G-PLE of $\tilde{F}_0(t)$:

\[
\hat{F}_0(s^*, t) = \prod_{w=0}^{t} \left[ 1 - \frac{\sum_{i=1}^{n} N_i(s^*, dw)}{S_0(s^*, w|\hat{\alpha}, \hat{\beta})} \right]
\]

For IID renewal model with $\mathcal{E}_i(s) = s - S_{iN_i^+(s_{-})}$, $\rho(k; \alpha) = 1$, and $\psi(w) = 1$, the estimator in PSH (2001) obtains.
Sum-Quota Effect: IID Renewal


\[
\sqrt{n}(\hat{F}(\cdot) - F(\cdot)) \rightarrow \text{GP}(0, \sigma^2(\cdot))
\]

\[
\sigma^2(t) = \bar{F}(t)^2 \int_0^t \frac{d\Lambda(w)}{\bar{F}(w)\bar{G}(w-)} [1 + \nu(w)]
\]

\[
\nu(w) = \frac{1}{\bar{G}(w-)} \int_w^\infty \rho^*(v - w) dG(v)
\]

\[
\rho^*(\cdot) = \sum_{j=1}^{\infty} F^*j(\cdot) = \text{renewal function}
\]
Semi-Parametric Estimation: With Frailty

- Recall the intensity rate:

\[ \lambda_i(s|Z_i, X_i) = Z_i \lambda_0[\mathcal{E}_i(s)] \rho[N_i^\dagger(s-); \alpha] \psi(\beta^t X_i(s)) \]

- Frailties \( Z_1, Z_2, \ldots, Z_n \) are unobserved and assumed to be IID Gamma(\( \xi, \xi \))

- Unknown parameters: \( (\xi, \alpha, \beta, \lambda_0(\cdot)) \)

- Use of the EM algorithm (Dempster, et al; Nielsen, et al), with frailties as missing observations.

- Estimator of baseline hazard function under no-frailty model plays an important role.

- Details in Peña, Slate & Gonzalez (JSPI, 2007).
First Application: MMC Data Set

Aalen and Husebye (1991) Data
Estimates of distribution of MMC period

![Graph showing survivor probability estimates for MMC period](image-url)
Second Application: Bladder Data Set

Bladder cancer data pertaining to times to recurrence for \( n = 85 \) subjects studied in Wei, Lin and Weissfeld ('89).
# Results and Comparisons

## Estimates from Different Methods for Bladder Data

<table>
<thead>
<tr>
<th>Cova</th>
<th>Para</th>
<th>AG</th>
<th>WLW Marginal</th>
<th>PWP Cond*nal</th>
<th>General Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $N(t-)$</td>
<td>$\alpha$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.98 (.07)</td>
</tr>
<tr>
<td>Frailty</td>
<td>$\xi$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>$\infty$</td>
</tr>
<tr>
<td>rx</td>
<td>$\beta_1$</td>
<td>-.47 (.20)</td>
<td>-.58 (.20)</td>
<td>-.33 (.21)</td>
<td>-.32 (.21)</td>
</tr>
<tr>
<td>Size</td>
<td>$\beta_2$</td>
<td>-.04 (.07)</td>
<td>-.05 (.07)</td>
<td>-.01 (.07)</td>
<td>-.02 (.07)</td>
</tr>
<tr>
<td>Number</td>
<td>$\beta_3$</td>
<td>.18 (.05)</td>
<td>.21 (.05)</td>
<td>.12 (.05)</td>
<td>.14 (.05)</td>
</tr>
</tbody>
</table>

*a* Effective Age is backward recurrence time ($\mathcal{E}(s) = s - S_{N^+(s)}$).

*b* Effective Age is calendar time ($\mathcal{E}(s) = s$).
On Asymptotic Properties

- Asymptotics under the no-frailty models.
- **Difficulty**: $\Lambda_0(\cdot)$ has $\mathcal{E}(s)$ as argument in the model; whereas, interest is usually on $\Lambda_0(t)$.
- **No** martingale structure in gap-time axis. MCLT not **directly** applicable.
- Under regularity conditions: **consistency** and **joint weak convergence** to Gaussian processes of standardized $(\hat{\alpha}, \hat{\beta})$ and $\hat{\Lambda}_0(s^*, \cdot)$.
- Results **extend** those in Andersen and Gill (AoS 82) regarding Cox PHM, though proofs different.
- Research on the asymptotics for the model **with frailty** in progress.
Asymptotics: Master Theorem

- \( \{H_i\} \) a sequence defined on \([0, s^*] \times [0, t^*]\).
- \( M_i(s, t) = \int_0^s I\{E_i(v) \leq t\} M_i^\dagger(dv) \).
- \( Y_i(s, t) \) - generalized at-risk process.
- Under some regularity conditions, and if

\[
\frac{1}{n} \sum_{i=1}^n H_i \otimes 2(s^*, \cdot) Y_i(s^*, \cdot) \xrightarrow{upr} v(s^*, \cdot),
\]

then, with \( \Sigma(s^*, t) = \int_0^t v(s^*, w) \Lambda_0(dw) \),

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\cdot H_i(s^*, w) M_i(s^*, dw) \Rightarrow GP(0, \Sigma(s^*, \cdot)).
\]
Relevant Empirical Measures

- **Simplified model (one unit):**

\[
\Pr\{dN_i^+(v) = 1 | \mathcal{F}_{s-}\} = Y_i^+(v)\lambda_0[\mathcal{E}_i(v)]\Xi_i(v; \eta)dv.
\]

- **Conditional PM** \(Q(s^*, w; \eta)\) on \(\{1, 2, \ldots, N^+(s^*-) + 1\} :\)

\[
Q(\{j\}; s^*, w; \eta) = \frac{\varphi_{j-1}(w; \eta)I\{\mathcal{E}(S_{j-1}) < w \leq \mathcal{E}(S_j)\}}{Y(s^*, w)}
\]

with \(S_{N^+(s^*-) + 1} = \min(s^*, \tau)\).

- **Conditional PM** \(P(s^*, w; \eta)\) on \(\{1, 2, \ldots, n\} :\)

\[
P(\{i\}; s^*, w; \eta) = \frac{Y_i(s^*, w; \eta)}{n\mathbb{P}Y(s^*, w; \eta)}.
\]
Empirical Means & Variances

\[ \mathbb{P} f(D) = \frac{1}{n} \sum_{i=1}^{n} f(D_i) \]

\[ \mathbb{E}_Q(s^*, w; \eta) g(J) = \sum_{j=1}^{N^\dagger(s^*-)+1} g(j) Q(\{j\}; s^*, w; \eta) \]

\[ \nabla Q(s^*, w; \eta) g(J) = \mathbb{E}_Q(s^*, w; \eta) [g^2(J)] - (\mathbb{E}_Q(s^*, w; \eta) g(J))^2 \]

\[ \mathbb{E}_P(s^*, w; \eta) g(I) = \sum_{i=1}^{n} g(i) P(\{i\}; s^*, w; \eta) \]

\[ \nabla P(s^*, w; \eta) g(I) = \mathbb{E}_Q(s^*, w; \eta) [g^2(I)] - (\mathbb{E}_Q(s^*, w; \eta) g(I))^2 \]
Relevant Limit Functions

- \( s_0(s^*, w; \eta, \Lambda_0) = \text{plim} \, \mathbb{P}Y(s^*, w; \eta). \)

- Partial Likelihood Information Limit:

\[
\mathcal{I}_p(s^*, t; \eta, \Lambda_0) = \text{plim} \int_0^t \left\{ \left[ \mathbb{E}_{P(s^*, w; \eta)} \nabla Q(s^*, w; \eta) \left( \nabla_\eta \log \Xi_I(\mathcal{E}_{IJ-1}(w); \eta) \right) \right] + \nabla P(s^*, w; \eta) \mathbb{E}_{Q(s^*, w; \eta)} \left( \nabla_\eta \log \Xi_I(\mathcal{E}_{IJ-1}(w); \eta) \right) \right\} \times s_0(s^*, w; \eta, \Lambda_0) \Lambda_0(dw).
\]

- With \( e(s^*, w; \eta, \Lambda_0) = \text{plim} \frac{\mathbb{P}\nabla_\eta Y(s^*, w; \eta)}{\mathbb{P}Y(s^*, w; \eta)}, \) let

\[
A(s^*, t; \eta, \Lambda_0) = \int_0^t e(s^*, w; \eta, \Lambda_0) \Lambda_0(dw).
\]
Weak Convergence Results

As \( n \to \infty \) and under certain regularity conditions:

\[
\sqrt{n}(\hat{\eta}(s^*, t^*) - \eta) \Rightarrow N(0, \mathcal{I}_p(s^*, t^*; \eta, \Lambda_0)^{-1})
\]

\[
\sqrt{n}(\hat{\Lambda}_0(s^*, \cdot) - \Lambda_0(\cdot)) \Rightarrow GP(0, \Gamma(s^*, \cdot; \eta, \Lambda_0))
\]

where the limiting variance function is given by

\[
\Gamma(s^*, t; \eta, \Lambda_0) = \int_0^t \frac{\Lambda_0(dw)}{s_0(s^*, w; \eta)} + A(s^*, t; \eta, \Lambda_0) \mathcal{I}_p(s^*, t^*; \eta, \Lambda_0)^{-1} A(s^*, t; \eta, \Lambda_0)^t.
\]
On Marginal Modeling: WLW and PWP

- $k_0$ specified (usually the maximum value of the observed $K$'s).
- Assume a Cox PH-type model for each $S_k$, $k = 1, \ldots, k_0$.
- Counting Processes ($k = 1, 2, \ldots, k_0$):
  \[ N_k(s) = I\{S_k \leq s; S_k \leq \tau \} \]
- At-Risk Processes ($k = 1, 2, \ldots, k_0$):
  \[ Y_k^{WLW}(s) = I\{S_k \geq s; \tau \geq s \} \]
  \[ Y_k^{PWP}(s) = I\{S_{k-1} < s \leq S_k; \tau \geq s \} \]
Working Model Specifications

WLW Model

\[
\begin{cases} 
N_k(s) - \int_0^s Y_k^{WLW}(v) \lambda_{0k}^{WLW}(v) \exp\{\beta_k^{WLW} X(v)\} dv \\
\end{cases}
\]

PWP Model

\[
\begin{cases} 
N_k(s) - \int_0^s Y_k^{PWP}(v) \lambda_{0k}^{PWP}(v) \exp\{\beta_k^{PWP} X(v)\} dv \\
\end{cases}
\]

are assumed to be zero-mean martingales (in s).
Parameter Estimation

See Therneau & Grambsch’s book *Modeling Survival Data: Extending the Cox Model.*

\( \hat{\beta}_k^{WLW} \) and \( \hat{\beta}_k^{PWP} \) obtained via partial likelihood (Cox (72) and Andersen and Gill (82)).

Overall \( \beta \)-estimate:

\[
\hat{\beta}_{WLW} = \sum_{k=1}^{k_0} \hat{c}_k \hat{\beta}_k^{WLW};
\]

\( c_k \)'s being ‘optimal’ weights. See WLW paper.

\( \hat{\Lambda}_{0k}^{WLW}(\cdot) \) and \( \hat{\Lambda}_{0k}^{PWP}(\cdot) \): Aalen-Breslow-Nelson type estimators.
Two Relevant Questions

**Question 1:** When one assumes marginal models for $S_k$s that are of the Cox PH-type, does there exist a full model that actually induces such PH-type marginal models?

**Answer:** YES, by a very nice paper by Nang and Ying (Biometrika:2001). BUT, the joint model obtained is rather ‘limited’.

**Question 2:** If one assumes Cox PH-type marginal models for the $S_k$s (or $T_k$s), but the true full model does not induce such PH-type marginal models *[which may usually be the case in practice]*, what are the consequences?
True Full Model: for a unit with covariate $X = x$, events occur according to an HPP model with rate:

$$\lambda(t|x) = \theta \exp(\beta x).$$

For this unit, inter-event times $T_k, k = 1, 2, \ldots$ are IID exponential with mean time $1/\lambda(t|x)$.

Assume also that $X \sim BER(p)$ and $\mu_T = E(\tau)$.

Main goal is to infer about the regression coefficient $\beta$ which relates the covariate $X$ to the event occurrences.
Full Model Analysis

\[ \hat{\beta} \text{ solves} \]
\[
\frac{\sum X_i K_i}{\sum K_i} = \frac{\sum \tau_i X_i \exp(\beta X_i)}{\sum \tau_i \exp(\beta X_i)}.
\]

\[ \hat{\beta} \text{ does not directly depend on the } S_{ij}. \text{ Why?} \]

**Sufficiency:** \((K_i, \tau_i)\)s contain all information on \((\theta, \beta)\).

\[
(S_{i1}, S_{i2}, \ldots, S_{iK_i}) | (K_i, \tau_i) \overset{d}{=} \tau_i (U_{(1)}, U_{(2)}, \ldots, U_{(K_i)}).
\]

**Asymptotics:**

\[
\hat{\beta} \sim AN \left( \beta, \frac{1}{n \mu \tau \theta [(1 - p) + pe^\beta]} \right).
\]
Some Questions

- Under WLW or the PWP: how are $\beta_k^{WLW}$ and $\beta_k^{PWP}$ related to $\theta$ and $\beta$?
- Impact of event position $k$?
- Are we ignoring that $K_i$s are informative? Why not also a marginal model on the $K_i$s?
- Are we violating the Sufficiency Principle?
- Results simulation-based: Therneau & Grambsch book (’01) and Metcalfe & Thompson (SMMR, ’07).
- Comment by D. Oakes that PWP estimates less biased than WLW estimates.
Properties of $\hat{\beta}_{k}^{W LW}$

Let $\hat{\beta}_{k}^{W LW}$ be the partial likelihood MLE of $\beta$ based on at-risk process $Y_{k}^{W LW}(v)$.

**Question:** Does $\hat{\beta}_{k}^{W LW}$ converge to $\beta$?

- $g_{k}(w) = w^{k-1}e^{-w}/\Gamma(k)$: standard gamma pdf.
- $\bar{G}_{k}(v) = \int_{v}^{\infty} g_{k}(w)dw$: standard gamma survivor function.
- $\bar{G}(\cdot)$: survivor function of $\tau$.
- $E(\cdot)$: denotes expectation wrt $X$. 
Limit Value (LV) of $\hat{\beta}_{k}^{W LW}$

- **Limit Value** $\beta_{k}^* = \beta_{k}^*(\theta, \beta)$ of $\hat{\beta}_{k}^{W LW}$: solution in $\beta^*$ of

$$\int_{0}^{\infty} E(X\theta e^{\beta X} g_k(v\theta e^{\beta X}))\bar{G}(v)dv =$$

$$\int_{0}^{\infty} e_{k}^{W LW}(v; \theta, \beta, \beta^*) E(\theta e^{\beta X} g_k(v\theta e^{\beta X}))\bar{G}(v)dv$$

where

$$e_{k}^{W LW}(v; \theta, \beta, \beta^*) = \frac{E(X e^{\beta^* X} \bar{g}_k(v\theta e^{\beta X}))}{E(e^{\beta^* X} \bar{g}_k(v\theta e^{\beta X}))}$$

- **Asymptotic Bias of** $\hat{\beta}_{k}^{W LW} = \beta_{k}^* - \beta$
Bias Plots for WLW Estimator

Colors pertain to value of \( k \), the Event Position

\( k = 1 \): Black; \( k = 2 \): Red; \( k = 3 \): Green; \( k = 4 \): DarkBlue;
\( k = 5 \): LightBlue

Theoretical

Simulated
On PWP Estimators

- Main Difference Between WLW and PWP:

\[
E(Y_k^{WLW} (v)|X) = \tilde{G}(v)\tilde{G}_k(v\theta \exp(\beta X));
\]

\[
E(Y_k^{PWP} (v)|X) = \tilde{G}(v)\frac{g_k(v\theta \exp(\beta X))}{\theta \exp(\beta X)}.
\]

- Leads to: \(u_k^{PWP} (s; \theta, \beta) = 0\) for \(k = 1, 2, \ldots\).

\(\hat{\beta}_k^{PWP}\) are asymptotically unbiased for \(\beta\) for each \(k\) (at least in this HPP model).

- Theoretical result consistent with observed results from simulation studies and D. Oakes’ observation.
Concluding Remarks

- Recurrent events prevalent in many scientific areas.
- **Dynamic models**: accommodate unique aspects of recurrent data.
- Inference for dynamic models need to be examined at a deeper level.
- **Current limitation**: keeping track of effective age. Current data schemes ignore this.
- GOF and residual analysis (Quiton’s dissertation).
- General studies on marginal modeling approaches!
- **Dynamic recurrent event modeling**, a challenge and a fertile research area.
Acknowledgements

Thanks to NIH grants that partially support research.

Thanks to research collaborators: Myles Hollander, Rob Strawderman, Elizabeth Slate, Juan Gonzalez, Russ Stocker, Akim Adekpedjou, and Jonathan Quiton.

Thanks to current students: Alex McLain, Laura Taylor, Josh Habiger, and Wensong Wu.

Thanks to all of you!