Issues in Marginal Recurrent Event Modeling

Alex McLain
Department of Statistics, University of South Carolina
Advisor: Dr. Edsel A. Peña
Talk at Eisai Medical Research

January 31, 2008

Outline

- Introduction to Recurrent Events

Outline

- Introduction to Recurrent Events
- Recurrent Event Modeling

Outline

- Introduction to Recurrent Events
- Recurrent Event Modeling
- Marginal Models

Outline

- Introduction to Recurrent Events
- Recurrent Event Modeling
- Marginal Models
- Properties of Marginal Models

Outline

- Introduction to Recurrent Events
- Recurrent Event Modeling
- Marginal Models
- Properties of Marginal Models
- Concluding Remarks

Examples of Recurrent Events

- Recurrent event data occur in many Biomedical studies.

Examples of Recurrent Events

- Recurrent event data occur in many Biomedical studies.
  1. Rejection episodes in kidney transplant recipients. Studied by Cook & Lawless (1997)

Examples of Recurrent Events

- Recurrent event data occur in many Biomedical studies.
  1. Rejection episodes in kidney transplant recipients. Studied by Cook & Lawless (1997)
  2. Chronic Granulomatous Disease (CGD). Studied by Fleming & Harrington (2005), and Therneau & Hamilton (1997)

Examples of Recurrent Events

- Recurrent event data occur in many Biomedical studies.
  1. Rejection episodes in kidney transplant recipients. Studied by Cook & Lawless (1997)
  2. Chronic Granulomatous Disease (CGD). Studied by Fleming & Harrington (2005), and Therneau & Hamilton (1997)
  3. Epileptic Seizures. Studied by Thall & Vail (1990) and Hougaard, Lee & Whitemore (1997)

Examples of Recurrent Events

- Recurrent event data occur in many Biomedical studies.
  1. Rejection episodes in kidney transplant recipients. Studied by Cook & Lawless (1997)
  2. Chronic Granulomatous Disease (CGD). Studied by Fleming & Harrington (2005), and Therneau & Hamilton (1997)
  3. Epileptic Seizures. Studied by Thall & Vail (1990) and Hougaard, Lee & Whitemore (1997)
  4. Cystic Fibrosis. Studied by Therneau & Hamilton (1997)

Event Time Distributions

- T: the time of the occurrence of an event
- $F(t) = Pr(T \leq t)$ the cumulative distribution function of T.
- $S(t) = 1 - F(t) = Pr(T > t)$ the survivor function of T.

Event Time Distributions

- T: the time of the occurrence of an event
- $F(t) = Pr(T \leq t)$ the cumulative distribution function of T.
- $S(t) = 1 - F(t) = Pr(T > t)$ the survivor function of T.
- Hazard rate/probability and Cumulative Hazards:

  Hazard rate: $\lambda(t) \approx Pr(T < t + dt | T \geq t) = \frac{f(t)}{S(t-)}$

  Cumulative Hazard: $\Lambda(t) = \int_0^t \lambda(w) dw$
- Hazard/Survivor relationship: $S(t) = \exp\{-\Lambda(t)\}$

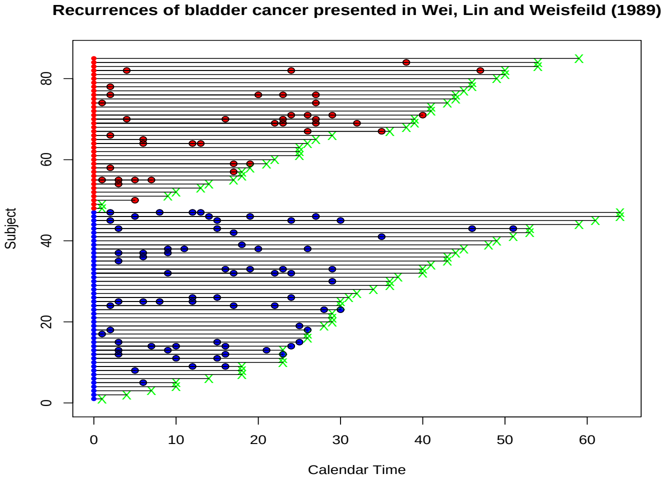
Inference

Estimation of F (S, Λ, λ):

- Parametric Inference: The unknown distribution function F is assumed to be a member of some family of distribution functions (Exponential, Gamma, Weibull).
- Based on $t_1, t_2, \ldots, t_n$, θ is estimated through, for example, Maximum Likelihood. Then $F(\cdot|\theta)$ is estimated by $F(\cdot|\hat{\theta})$.
- Non-Parametric Inference: No assumptions whatsoever are made about the distribution. Example: Empirical Distribution Function (EDF)

$$\hat{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I\{T_i \leq t\}$$

Example


Recurrences of bladder cancer presented in Wei, Lin and Weisfeild (1989)
Subject
0
20
40
60
80
0
10
20
30
40
50
60
Calendar Time

Censoring

- Censoring (left, right, and interval).
- For the *i*th unit there will be possibly random censoring variables $\tau_i \sim Q$.
- The observed variables are (Z_i, δ_i) where $Z_i = min(T_i, C_i)$ and $\delta_i = I(T_i \leq C_i)$.
- Problem: For the observed data (Z_i, δ_i) how do we estimate F.

Data Structure

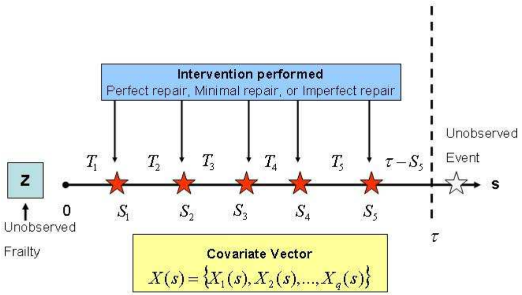


Figure: Graphical view of a unit with recurrent events.

Data Notation

- X(s) - possibly time dependent covariate vector.
- $T_1, T_2, \ldots$ - interevent or gap times.
- $S_1, S_2, \ldots$ - calendar times of occurrences.
- τ - end of the observation period.
- K = $\max\{k; S_k \leq \tau\}$ - number of events observed.
- Z - unobserved frailty variable.

Counting Process Notation

- To accompany the recurrent event data structure we introduce:

$$N_i^{\dagger}(s) = \sum_{k=1}^{\infty} I(S_{ik} \leq s, S_{ik} \leq \tau_i)$$

and

$$Y_i^{\dagger}(s) = I\{\tau_i \geq s\}$$

as the counting process and at risk, for the ith subject, $\{i = 1, 2, \ldots, n\}$

- $Y_i^{\dagger}(s)$ is a predictable process.

Martingales

- Following Gill (1980), we can form:

$$M_i^{\dagger}(s) = N_i^{\dagger}(s) - A_i^{\dagger}(s)$$

which is a zero mean square integrable martingale. Where:

$$A_i^{\dagger}(s) = \int_0^s Y_i^{\dagger}(v)\lambda\left\{\varepsilon_i(v)\right\} dv \tag{1}$$

- With respect to the common history (or filtration):

$$\mathcal{F}_s = \mathcal{F}_0 \vee \left\{ \bigvee_{i=1}^{n} \sigma\left\{ \left(N_i^{\dagger}(w), Y_i^{\dagger}(w+), \varepsilon_i(w+) \right) : 0 < w \leq s \right\} \right\}$$

Martingale Example

The Martingale Central Limit Theorem (MCLT) gives a limiting normal distribution.

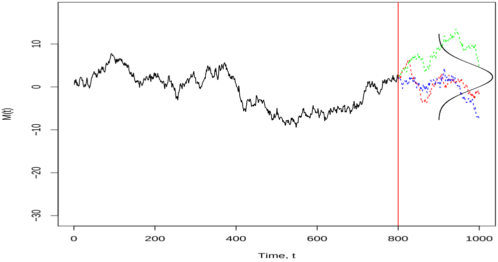


Figure: Example of a fair game process.

Counting Process Style Models

- The adoption of the counting process framework has lead to rapid progress.
- Aalen (1975) originally put recurrent events into a counting process style framework.
- Gill (1980), the texts of Fleming & Harrington (1991) and Andersen, Borgan, Gill & Keiding (1993) are extremely useful.

Methods

Methods of Analysis of Recurrent Event Data:

- Parametric (Weibull regression, exponential, lognormal, Gamma process prior etc.). Using the likelihood of Jacod (1975)

$$L_i(\theta|s) = \left\{ \prod_{w=0}^{s} [dA_i^\dagger(w)]^{dN_i^\dagger(w)} \right\} \exp\{-A_i^\dagger(s)\}$$

Methods

Methods of Analysis of Recurrent Event Data:

- Parametric (Weibull regression, exponential, lognormal, Gamma process prior etc.). Using the likelihood of Jacod (1975)

$$L_i(\theta|s) = \left\{ \prod_{w=0}^{s} [dA_i^\dagger(w)]^{dN_i^\dagger(w)} \right\} \exp\{-A_i^\dagger(s)\}$$

- Nonparametric (Kaplan-Meier, Generalized Nelson-Aalen, Dirichlet Process)
- Semi-parametric (Cox (1972) proportional hazards regression model), using partial likelihood.

Dynamic Models

A General Class of Full Models was proposed by Peña and Hollander (2004):

$$\lambda(s|Z) = Z\lambda_0(\mathcal{E}(s))\rho[N(s-);\alpha]\psi[\beta^t X(s)]$$

Dynamic Models

A General Class of Full Models was proposed by Peña and Hollander (2004):

$$\lambda(s|Z) = Z\lambda_0(\mathcal{E}(s))\rho[N(s-);\alpha]\psi[\beta^t X(s)]$$

Where

$$\begin{aligned} Z \quad &- \quad \textit{Unobserved frailty} \\ \lambda_0(\cdot) \quad &- \quad \textit{Baseline hazard rate function} \\ \rho[\cdot;\alpha] \quad &- \quad \textit{nonnegative function with } \rho[0;\alpha] = 1 \\ \psi[\cdot] \quad &- \quad \textit{nonnegative link function} \end{aligned}$$

Examples are $\rho[k;\alpha] = \alpha^k$ and $\psi[\beta^t X(s)] = \exp\{\beta^t X(s)\}$.

Effective Age Process $\mathcal{E}(s)$

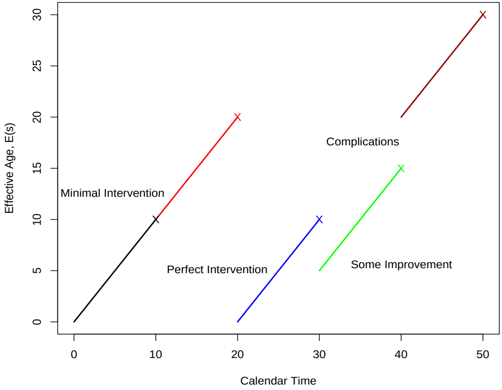


Figure: Examples of possible effective age's.

Examples of Recurrent Events

Recall the earlier examples mentioned:

- Rejection episodes in kidney transplant recipients. Studied by Cook & Lawless (1997)
- Chronic Granulomatous Disease (CGD). Studied by Fleming & Harrington (2005), and Therneau & Hamilton (1997)
- Epileptic Seizures. Studied by Thall & Vail (1990) and Hougaard, Lee & Whitemore (1997)
- Cystic Fibrosis. Studied by Therneau & Hamilton (1997)

Introduction to Cox Proportional Hazards Model

- In the Cox proportional hazards model, we assume that the hazard rate of T is:

$$\lambda(t|x) = \lambda_0(t)\exp\{\beta^t X\}$$

where β is a q x 1 vector of parameters and X is a q x 1 vector of covariates.

Introduction to Cox Proportional Hazards Model

- In the Cox proportional hazards model, we assume that the hazard rate of T is:

$$\lambda(t|x) = \lambda_0(t)\exp\{\beta^t X\}$$

  where β is a q x 1 vector of parameters and X is a q x 1 vector of covariates.
- The goal of this model is to make inference about λ_0 and/or β.
- $\hat{\beta}$ maximizes the partial likelihood function of β:

$$L_p(\beta) = \prod_{i=1}^{n} \prod_{t \leq \infty} \left[\frac{\exp\{\beta^t X_i\}}{\sum_{j=1}^{n} Y_j(t)\exp\{\beta^t X\}} \right]^{dN_i(t)}$$

Cox Proportional Hazards Methods

- First Event Analysis: Inefficient, though valid.
- Full Modeling Approach: Andersen & Gill (82); Lindqvist & co-workers; Peña & Hollander (04).
- Marginal Modeling Approach: Pioneered by Wei, Lin & Weissfeld (89), hence called the WLW approach.
- Conditional Modeling Approach: Pioneered by Prentice, Williams & Peterson (81), hence referred to as the PWP approach.

Marginal Models

- Marginal models specify a distribution on the kth event T_{ik} or S_{ik} for all subjects.

Marginal Models

- Marginal models specify a distribution on the kth event T_{ik} or S_{ik} for all subjects.

Why are marginal models used?

- The relationship a subject's rate of failure will have as events continue to occur, is difficult to specify.
- The dependence between the events is not an interesting aspect of the analysis.
- Marginal models ignore the dependence for estimation and correct for it the variance.
- Easier to interpret and analyze.

Marginal Modeling with the Cox Model

The two methods of marginal analysis differ in their at-risk function:

Marginal Modeling with the Cox Model

The two methods of marginal analysis differ in their at-risk function:

1. *Conditional (PWP) Model*

The PWP model what proposed in Prentice et al. (1981):

$$\lambda_{ik}(t|\beta, x_i) = Y_{ik}^{p}(t)\lambda_0(t)e^{\beta x_i}$$

where $Y_{ik}^{p}(t) = I(S_{ik-1} < t \leq S_{ik} \text{ and } t \leq \tau_i)$.

Marginal Modeling with the Cox Model

The two methods of marginal analysis differ in their at-risk function:

1. *Conditional (PWP) Model*

The PWP model what proposed in Prentice et al. (1981):

$$\lambda_{ik}(t|\beta, x_i) = Y^p_{ik}(t)\lambda_0(t)e^{\beta x_i}$$

where $Y^p_{ik}(t) = I(S_{ik-1} < t \leq S_{ik} \text{ and } t \leq \tau_i)$.

2. *WLW Model*:

The WLW model, proposed in Wei et al. (1989):

$$\lambda_{ik}(t|\beta, x_i) = Y^w_{ik}(t)\lambda_0(t)e^{\beta x_i}$$

where $Y^w_{ik}(t) = I(t \leq S_{ik} \text{ and } t \leq \tau_i)$.

Differences in Strata

Differences in Strata

- *Andersen Gill Model:* All events are in the same strata, models yields a single estimate for β.

For marginal models there are k_0 strata:

- *PWP Model:* Subjects progress in strata $1 \rightarrow 2 \rightarrow 3 \rightarrow$ etc., and are in 1 strata at a time.
- *WLW Model:* At time t subject is in strata $\{N(t)+1, N(t)+2, \ldots, k_0\}$, and are in $k_0 - N(t)$ strata at time t.

Marginal Modeling with the Cox Model

- $\hat{\beta}_k^w$ (WLW) and $\hat{\beta}_k^p$ (PWP) are obtained via maximum likelihood, from the partial score processes $U_{pk}^W(\alpha)$ and $U_{pk}^P(\alpha)$ where:

$$U_{pk}^W(\alpha) = \frac{\partial}{\partial\alpha} l_{pk}(\alpha) = \sum_{i=1}^{n} \int_0^{\infty} \left\{ x_i - \frac{S_K^{w(1)}(v;\alpha)}{S_K^{w(0)}(v;\alpha)} \right\} dN_{ik}(v)$$

where $S_k^{w(m)}(t;\alpha) = \sum_{i=1}^{n} x_i^m Y_{ik}^w(t) e^{\alpha x_i}$.

Marginal Modeling with the Cox Model

- $\hat{\beta}_k^w$ (WLW) and $\hat{\beta}_k^p$ (PWP) are obtained via maximum likelihood, from the partial score processes $U_{pk}^W(\alpha)$ and $U_{pk}^P(\alpha)$ where:

$$U_{pk}^W(\alpha) = \frac{\partial}{\partial\alpha} l_{pk}(\alpha) = \sum_{i=1}^{n} \int_0^\infty \left\{ x_i - \frac{S_K^{w(1)}(v;\alpha)}{S_K^{w(0)}(v;\alpha)} \right\} dN_{ik}(v)$$

  where $S_k^{w(m)}(t;\alpha) = \sum_{i=1}^n x_i^m Y_{ik}^w(t) e^{\alpha x_i}$.

- An overall estimate may be obtained via:
  1. $\hat{\beta^w} = \sum_{i=1}^{k_0} c_i \hat{\beta}_i^w$ with c_i being 'optimal' weights. See Wei et al. (1989).
  2. $U_p^w(\alpha) = \sum_{i=1}^{k_0} U_{pi}^W(\alpha)$. 'Working' independence estimate.

Questions of marginal models

What is the overall effect of using marginal models?

Questions of marginal models

What is the overall effect of using marginal models?

- Is there loss of efficiency?

Questions of marginal models

What is the overall effect of using marginal models?

- Is there loss of efficiency?
- Should marginal estimates be combined?

Questions of marginal models

What is the overall effect of using marginal models?

- Is there loss of efficiency?
- Should marginal estimates be combined?
- What assumptions do marginal models make?

Questions of marginal models

What is the overall effect of using marginal models?

- Is there loss of efficiency?
- Should marginal estimates be combined?
- What assumptions do marginal models make?
- What is the bias of these models when incorrectly specified?

Assumptions

In order to discuss properties let
$N_{ik}(t) = I(S_{ik} \leq t \; and \; S_{ik} \leq \tau_i)$.

- The gap times $T_{i1}, T_{i2}, \ldots, T_{ik_i} \stackrel{iid}{\rightarrow} \bar{G_1}$ with hazard:

$$\lambda_1(t|\beta, x) = \lambda_0(t)e^{\beta^t x}$$

- The calendar times $S_{1k}, S_{2k}, \ldots, S_{nk} \stackrel{iid}{\rightarrow} \bar{G_k}$ with hazard $\lambda_k(t|\beta, x)$.

Assumptions

In order to discuss properties let
$N_{ik}(t) = I(S_{ik} \leq t \textit{ and } S_{ik} \leq \tau_i)$.

- The gap times $T_{i1}, T_{i2}, \ldots, T_{ik_i} \overset{iid}{\rightarrow} \bar{G_1}$ with hazard:

$$\lambda_1(t|\beta, x) = \lambda_0(t)e^{\beta^t x}$$

- The calendar times $S_{1k}, S_{2k}, \ldots, S_{nk} \overset{iid}{\rightarrow} \bar{G_k}$ with hazard $\lambda_k(t|\beta, x)$.
- Under these assumptions:

$$M_{ik}(t) = N_{ik}(t) - \int_0^t Y_{ik}^w(v)\lambda_k(v|\beta, x_i)dv$$

Forms a zero mean square integrable martingale.

Properties

- Recall the score process:

$$U_{pk}^{W}(\alpha) = \sum_{i=1}^{n} \int_{0}^{\infty} \{H_{ik}^{w}(v; x, \alpha)\} \, dN_{ik}(v)$$

where $H_{ik}^{w}(v; x, \alpha) = x_i - \frac{S_{ik}^{w(1)}(v;\alpha)}{S_{ik}^{w(0)}(v;\alpha)}$

- Which by the above assumptions can be written as:

$$U_{pk}^{w}(\alpha) = \sum_{i=1}^{n} \left\{ \int_{0}^{\cdot} H_{ik}^{w}(v; x, \alpha) dM_{ik}(v) + \int_{0}^{\infty} H_{ik}^{w}(v; x, \alpha) Y_{ik}^{w}(v) \lambda_k(v; \beta, x_i) dv \right\}$$

- Since M_{ik} is a is a martingale $\int_0^{\cdot} H_{ik}^{w}(v; x, \alpha) dM_{ik}(v)$ is a martingale.

Score Function

- Since the first part is a martingale:

$$\frac{1}{n}U^W_{pk}(\alpha) \longrightarrow u^W_k(\alpha) = E\left\{\int_0^\infty \left[x_i - e^w_{ik}(v;x,\alpha)\right] Y^w_{ik}(v)\lambda_k(v;\beta,x_i)dv\right\}$$

as $n \to \infty$, where

$$e^w_{ik}(v;x,\alpha) = \frac{E[S^{w(1)}(t;\alpha)]}{E[S^{w(0)}(t;\alpha)]} = \frac{E_x[xe^{\alpha x}\bar{G}_K(v;\beta,x)]}{E_x[e^{\alpha x}\bar{G}_K(v;\beta,x)]}$$

Bias

- Solving $u_k(\beta)$ will give the asymptotic estimate of β when n$\rightarrow \infty$.
- Under the PWP model it can be shown that $u_k^p(\beta)$= 0 for a case control study. Hence the PWP model is unbiased under the assumptions for any F.
- For the WLW model $u_k^w(\alpha)$=0 does not necessarily occur when $\alpha = \beta$.
- When this is the case $u_k^w(\alpha)$ can be used to calculate the asymptotic bias.

Homogeneous Poisson Process

- Understanding simple models might provide important insights.
- The simplest for of a proportional hazard model is when $\lambda_0(t) = \theta$ or the Homogeneous Poisson Process.
- In this case T_{ik} ∼Exponential($\theta e^{\beta x}$), and S_{ik} ∼Gamma(k, $\theta e^{\beta x}$).
- Assume that X ∼ Ber(p), and $\tau_i = B_i$ for i=1,...,n.
- Goal is to make inference on β.

Full Model Analysis

Full analysis of HPP($\theta e^{\beta x}$)

- In this case $\hat{\beta}$ solves:

$$\frac{\sum X_i K_i}{\sum K_i} = \frac{\sum \tau_i X_i e^{\beta^t X_i}}{\sum \tau_i e^{\beta^t X_i}}$$

- Note: $\hat{\beta}$ does not directly depend on T_{ik} or S_{ik}.
- Sufficiency: (K_i, τ_i)'s contain all the information about β.
- Since we know the distribution of S_{ik} we can calculate the value of α such that $u_k^w(\alpha) = 0$.

Simulated Bias of β_k^W

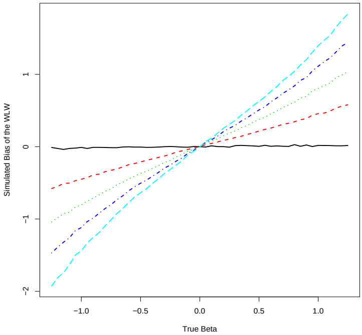


Figure: Simulated bias of the WLW model (for k=1, 2, 3, 4, and 5) under exponential interevent times, for n=50 and M=2000

Calculated Bias using $u_k^w(\alpha)$

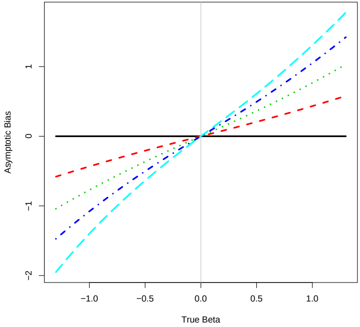


Figure: Calculated bias of the WLW method when the interevent times follow an Exponential ($e^{\beta x}$).

Unspecified G_k

- Goal: To infer about β_k^w for an unspecified distribution.
- In a case control study, $u_k^w(\alpha)$ can be written as:

$$u_k^w(\alpha) = \int_0^\infty H(v; p, \bar{G}, k, \alpha) \{\lambda_k(v|1) - e^\alpha \lambda_k(v|0)\} \, dv$$

  where $H(v; p, \bar{G}, K, \alpha) \geq 0$
- $\lambda_1(v|1) = e^\beta \lambda_1(v|0)$, so first event is unbiased.
- Will $|\beta_k^w| \geq |\beta|$ for k > 1?

Hazard Rate Ordering

Lemma: For iid random variable Y_1, and Y_2 with hazard $\lambda_Y(t|\beta) = e^{\beta}\lambda_0(t)$ and iid random variables Z_1 and Z_2 with hazard $\lambda_Z(t|\beta) = \lambda_0(t)$. Under certain conditions (including $\lambda_0(t) \in$ IFR) we have that:

$$\lambda_{Y_1+Y_2}(t|\beta) \geq e^{\beta}\lambda_{Z_1+Z_2}(t)\ \forall t \in [0, \infty],\ \beta > 0$$

Consequences:

1. Under the conditions $|\beta_2^W| \geq |\beta|$, or the WLW method has positive absolute bias.
2. Full estimates β^W will also be bias.
3. Significance tests for β_k^W will have false power.

Summary

- An introduction to recurrent event analysis was given.
- Properties of Marginal Models, were looked into.
- Use of the WLW method in the iid gap time case appears not to be valid.
- Under these assumptions estimates should not be combined.
- Marginal model assumptions.

Acknowledgements

- Research support provided by NIH Grant 2 R01 GM56182
- Thank you to Irene Hueter and Carlos Perdomo.

Thank You

Thanks for listening.
Please contact me at alex.mclain@gmail.com with any questions!!