

Analysis of Longitudinal Data for Inference and Prediction



- Patrick J. Heagerty PhD
- Department of Biostatistics
- University of Washington

Outline of Short Course

- Inference
 - ▷ How to handle “stochastic” measurements and/or events?
 - ▷ How to address time-dependent covariates?
- Prediction
 - ▷ How to measure ability of marker(s) to predict event time?
 - ▷ What are appropriate time-specific or overall summaries and how do these relate to medical decision making?

Some General Comments / Disclaimers

- Moderately advanced material
- Overview of major issues / depth only with select issues
- First time together / total (!)
- Theory into practice is my goal
- Course teaching ideas rather than SAS / STATA / R implementation

Analysis of Longitudinal Data for Inference and Prediction: Part I – Overview of Issues



- Patrick J. Heagerty PhD
- Department of Biostatistics
- University of Washington

LDA Progress!

- During the last couple of decades statistical methods have been developed (ie. LMM, GEE) that can analyze longitudinal data with:
 - ▷ Unequal number of observations per person (n_i)
 - ▷ Unequally spaced observations (t_{ij})
 - ▷ Time-varying covariates (x_{ij})

- Regression questions:

$$\mu_i(t) = E[Y_i(t) | X_i(t)]$$

- Q: **When** should we directly apply these now standard longitudinal methods to data with the features listed above?

Session One Outline

- Examples
 - ▷ Cystic Fibrosis Foundation (CFF)
 - ▷ Maternal Stress and Child Morbidity (MSCM)
 - ▷ United States Renal Data System (USRDS)
 - ▷ Collaborative Perinatal Project (CPP)
- Time-varying Covariate Processes
 - ▷ Exogenous
 - * Lagged covariates
 - ▷ Endogenous
 - * Fixed vs Dynamic exposure

(* Session One Outline (continued))

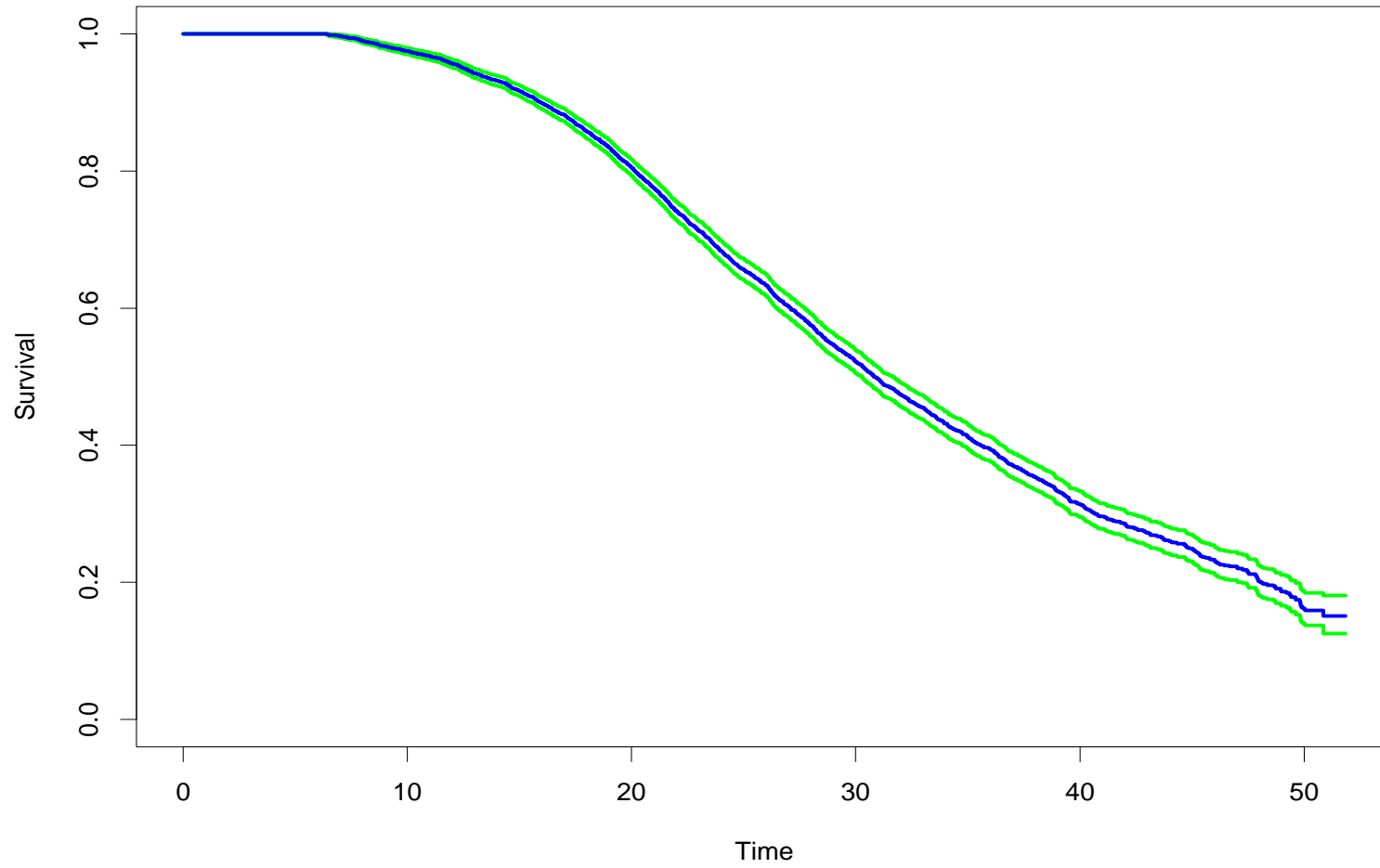
- Analysis with Drop-out
- Analysis with Death
 - ▷ Specification of model
 - ▷ Inference
- Analysis with “drop-in” (scheduling, birth)

Repeated Measures Data

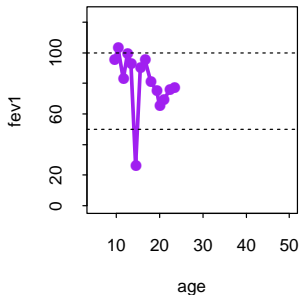
Cystic Fibrosis Data

- $N = 23,530$ subjects, 4,772 deaths, 1986-2000
- $n = 160,005$ longitudinal observations
- Longitudinal measurements: FEV1, weight, height
- Goal: identify factors associated with decline in pulmonary function.
- (Another Goal: predict mortality; transplantation selection)

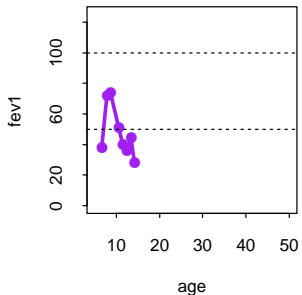
CFF Survival



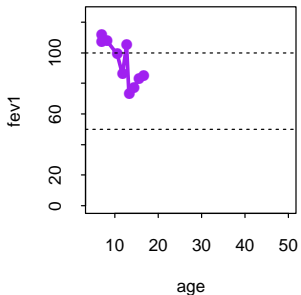
907762



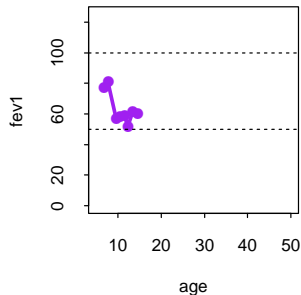
914421



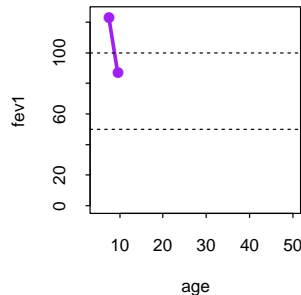
931885



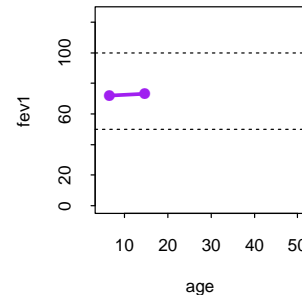
930200



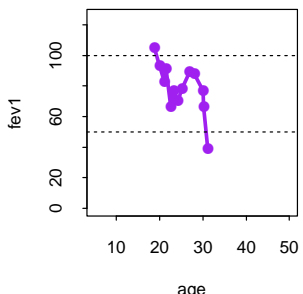
918755



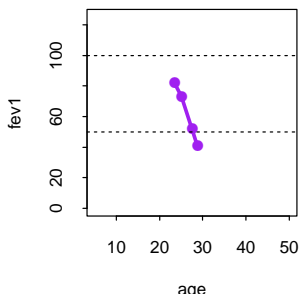
904750



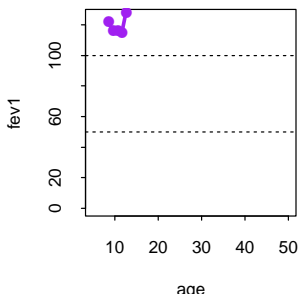
906580



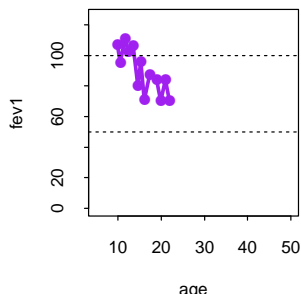
922747



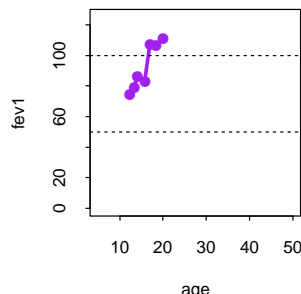
913076



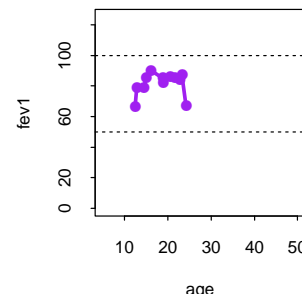
923143



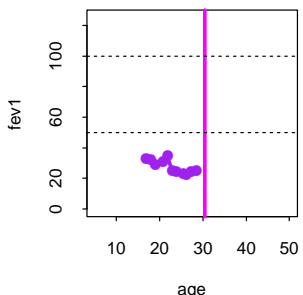
922295



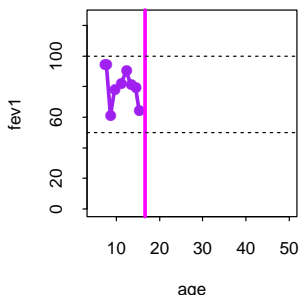
922858



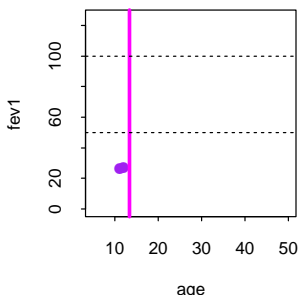
928943



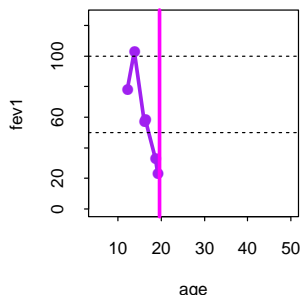
925274



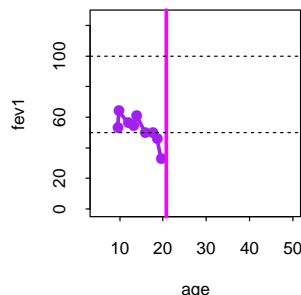
905506



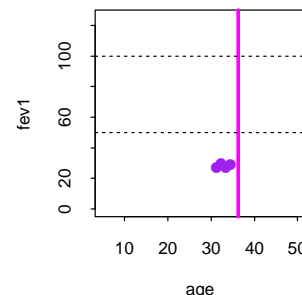
925198



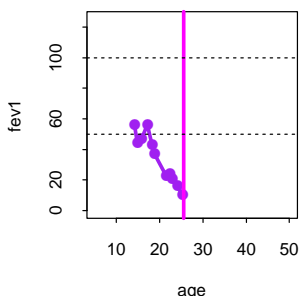
912910



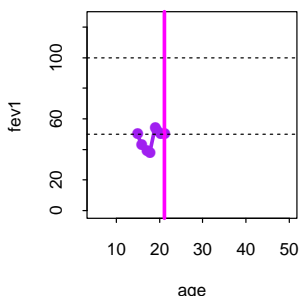
930735



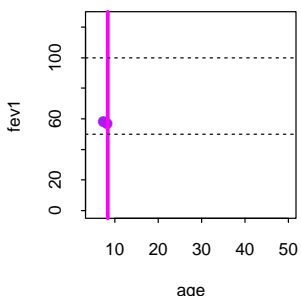
927419



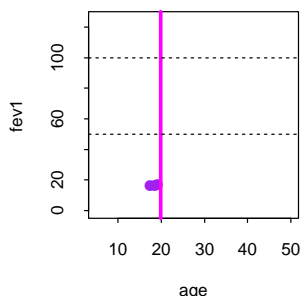
908675



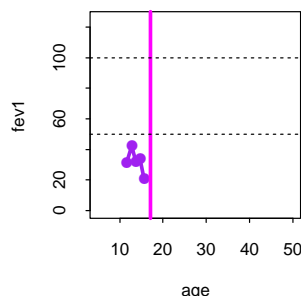
916766



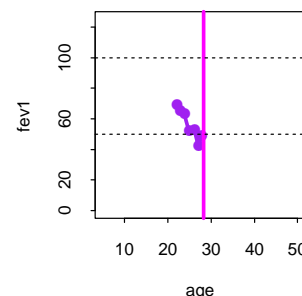
926609



920366



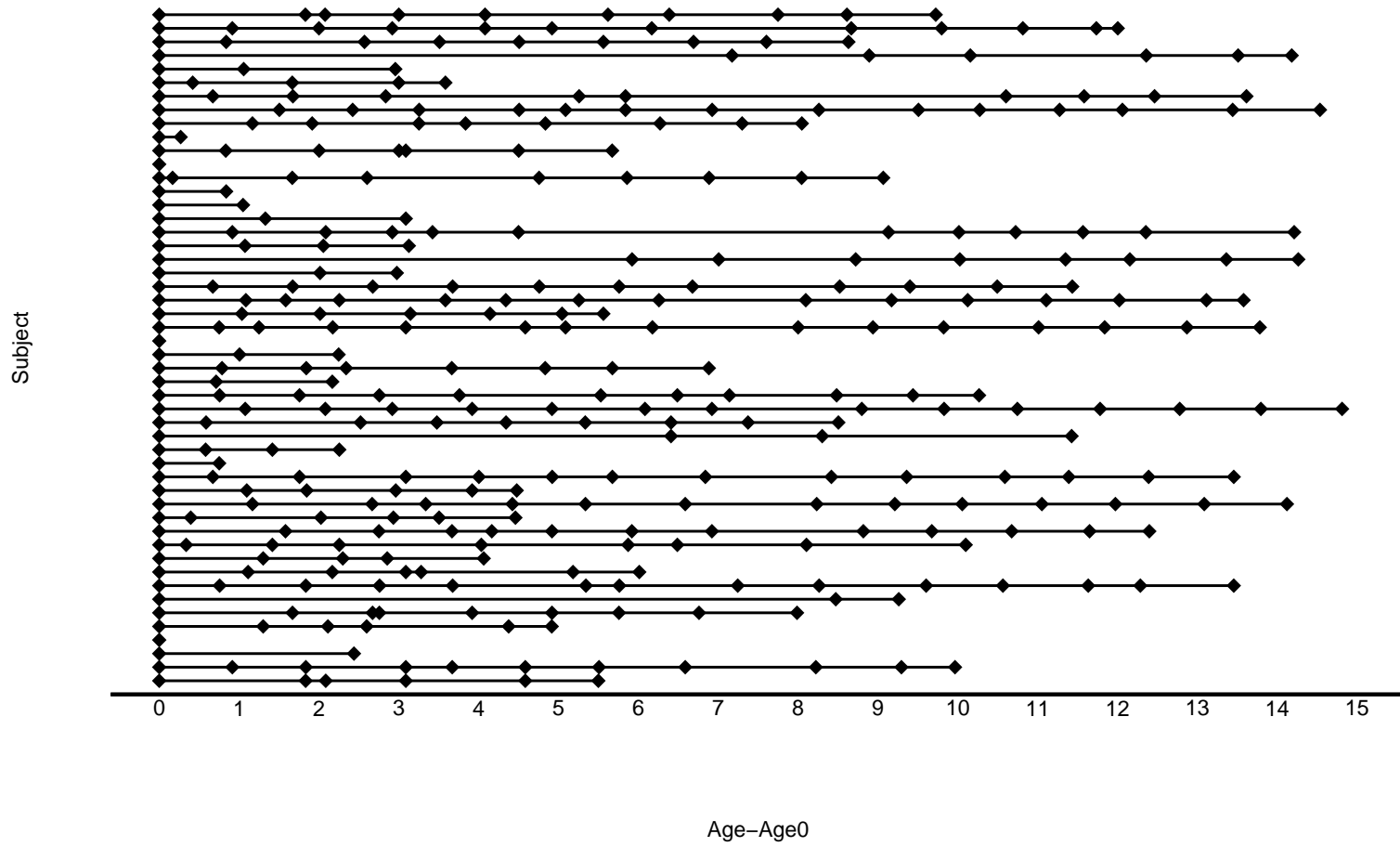
927349



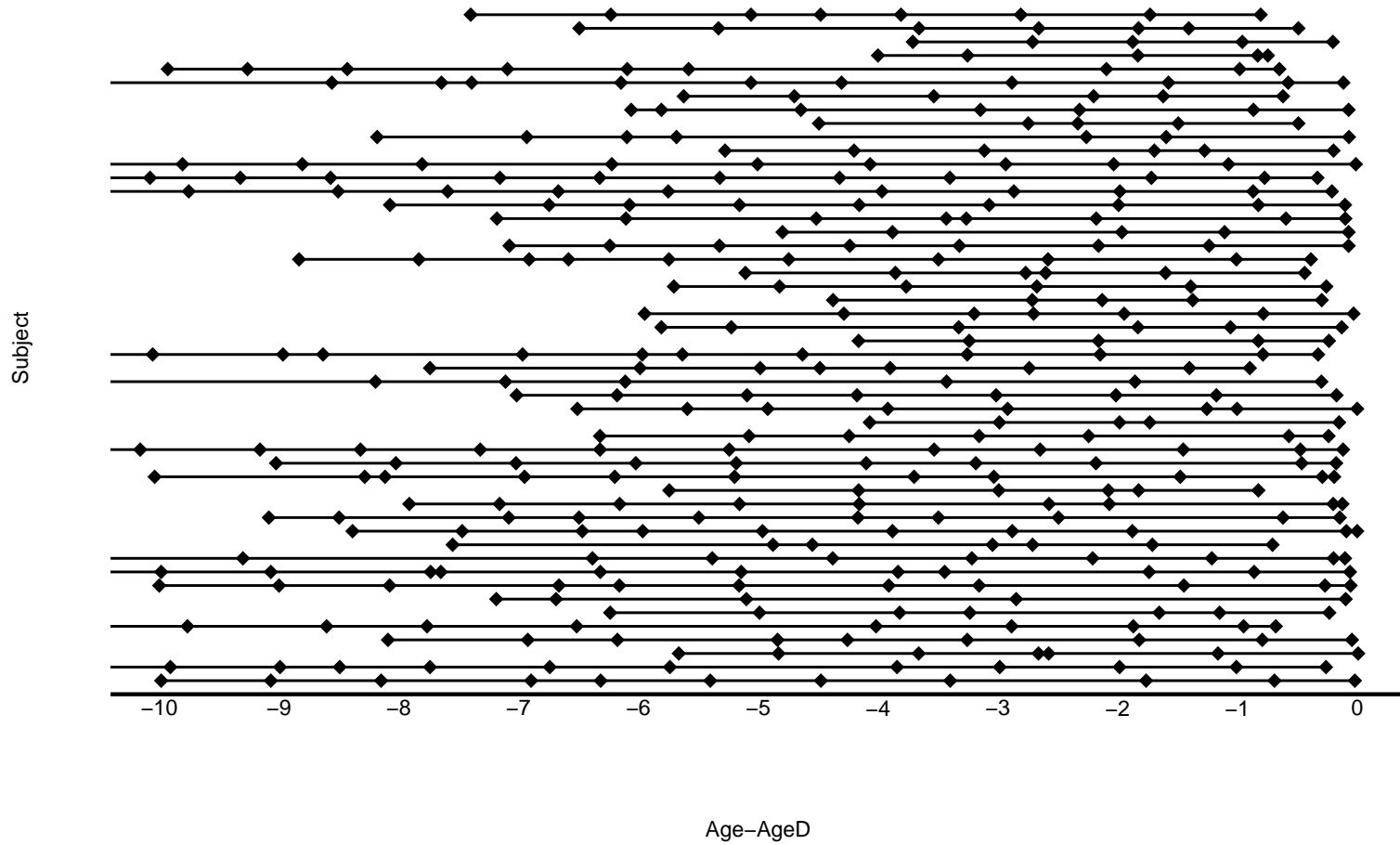
Example: Scientific Goals & CF

- Parad RB, Gerard CJ, Zurakowski D, Nichols DP, Pier GB
“Pulmonary outcome in cystic fibrosis is influenced primarily by mucoid *Pseudomonas aeruginosa* infection and immune status and only modestly by genotype.”
Infect. Immun., **67**(9): 4744-50, 1999.
- Variables:
 - ▷ Measurement time: t_{ij}
 - ▷ Pulmonary function: $Y_i(t_{ij})$
 - ▷ Time-dependent covariate: $X_i(t_{ij})$ – infection status
 - ▷ Death: $D_i(t)$ counting process for T_i

CFF Data and Visit Times



CFF Data and Visit Times -- CASES



Maternal Stress and Child Morbidity

Example 2: Time-dependent covariates

- daily indicators of stress (maternal), and illness (child)
- primary outcome: illness, utilization
- covariates: employment, stress
- **Q**: association between employment, stress and morbidity?
- **Q**: Does stress cause morbidity?

UTILIZATION OF PEDIATRIC AMBULATORY CARE

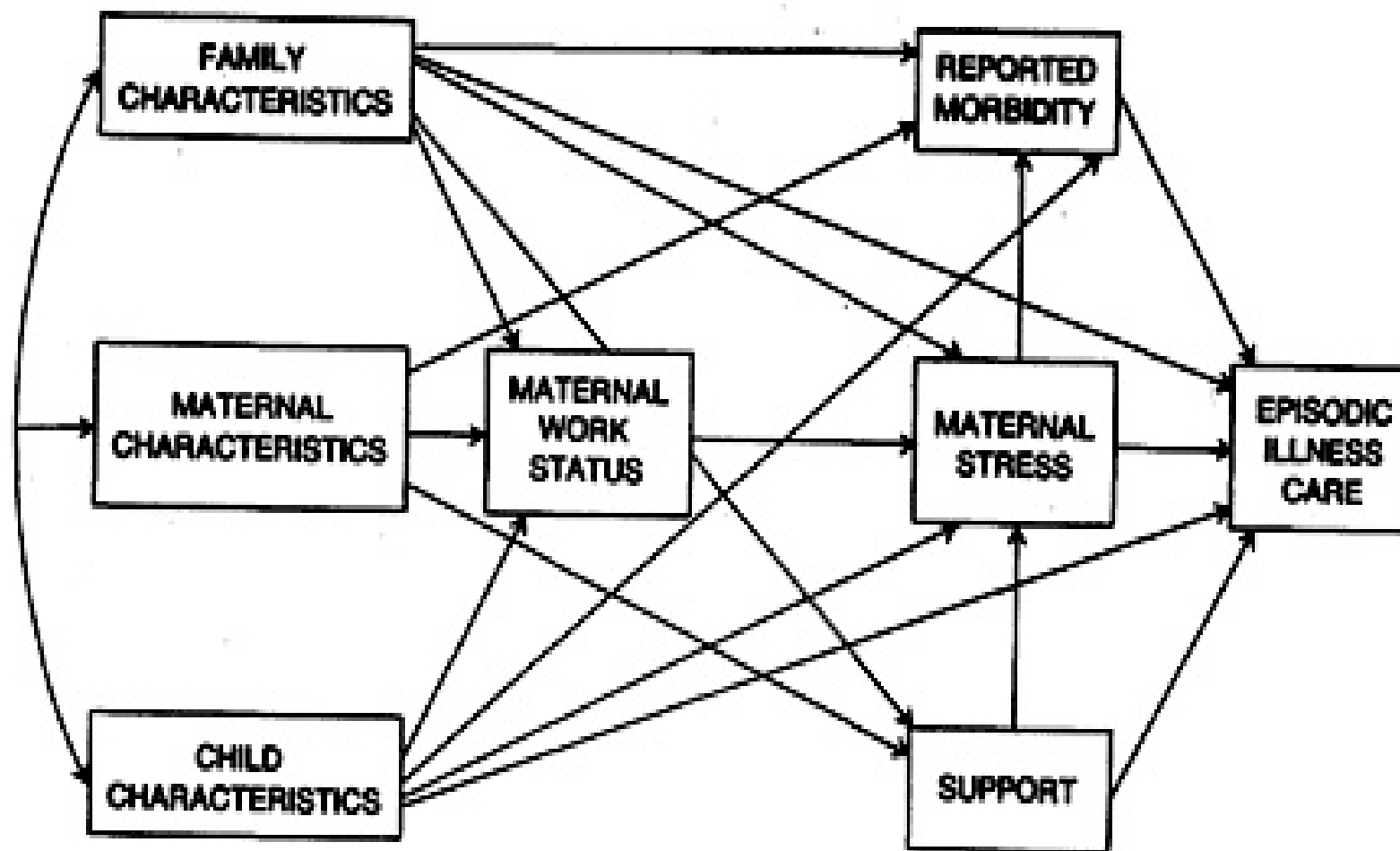
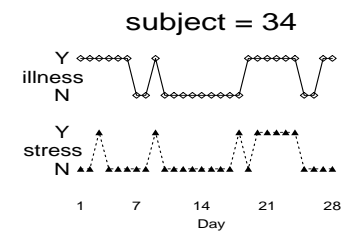
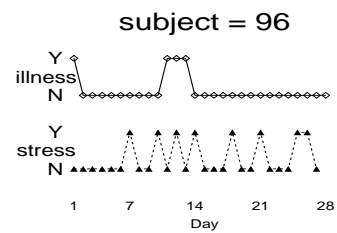
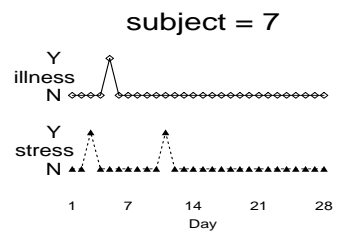
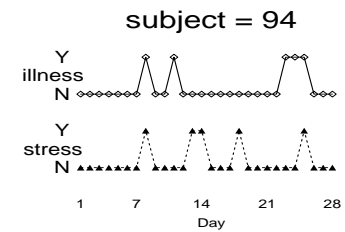
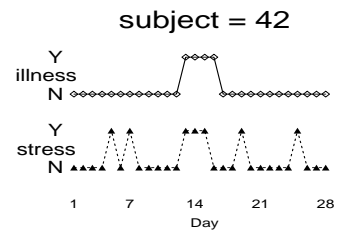
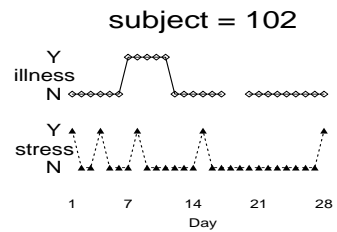
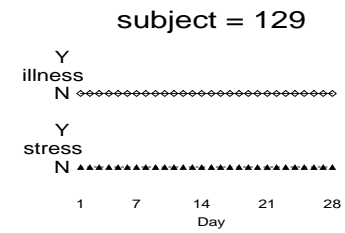
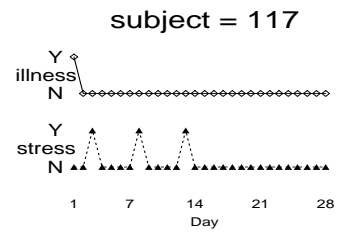
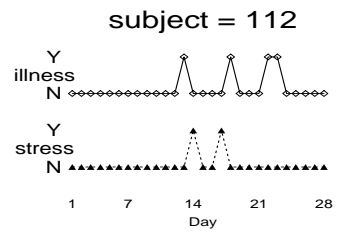
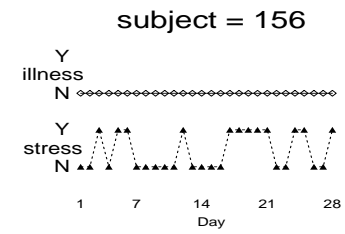
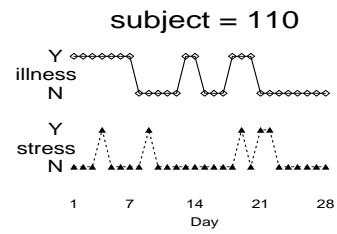
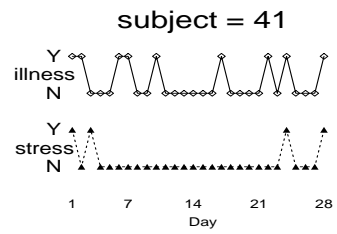


FIG. 1. Determinants of episodic illness care utilization.

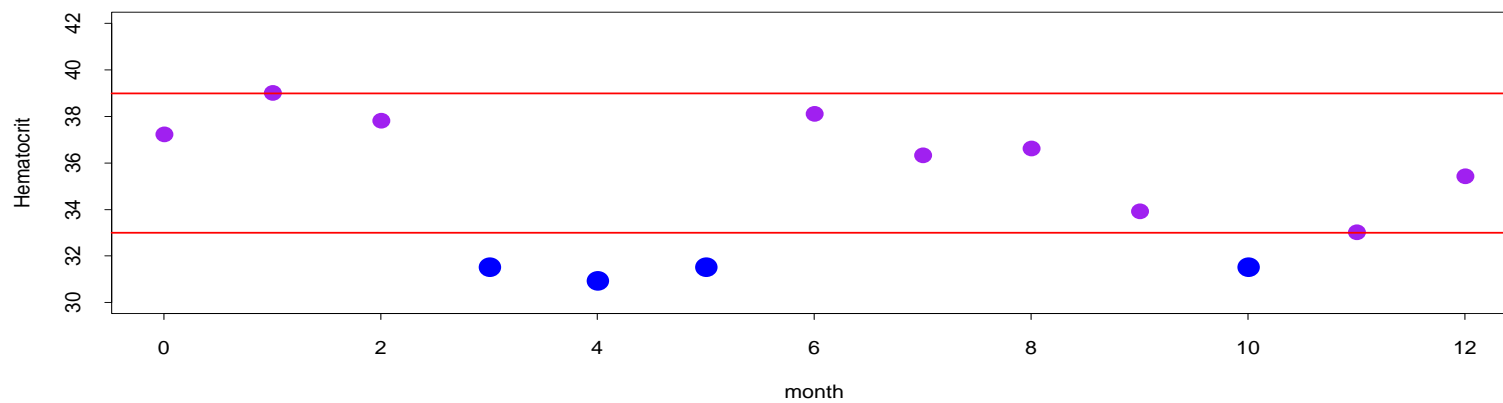
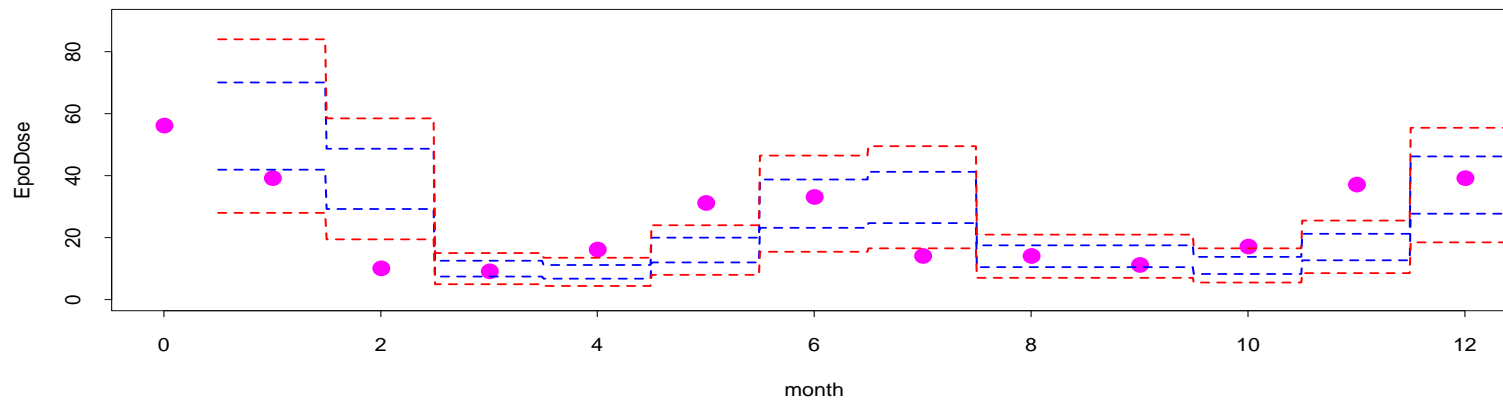


USRDS Data: Safety of ESAs?

- End Stage Renal Disease (ESRD)
 - ▷ Poor kidney function
 - ▷ Dialysis
 - ▷ Fail to stimulate formation of red blood cells
- Epoetin
 - ▷ Anemia treatment
 - ▷ \$3 billion Medicare / year
- Studies show an association between high dose and risk of death
 - ▷ Adverse outcomes?
 - ▷ Confounding by indication?

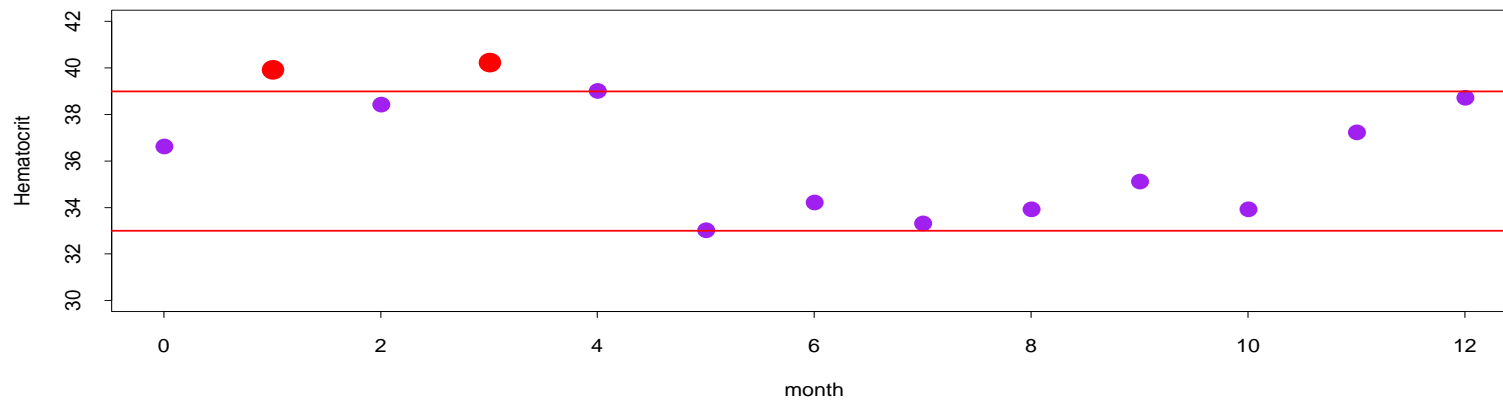
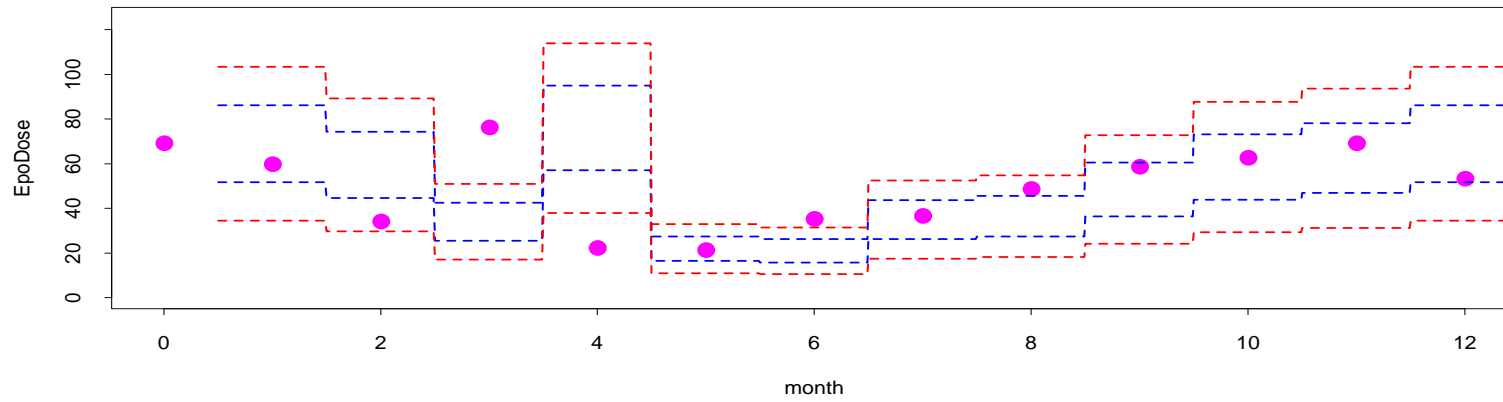
USRDS Dialysis Data

ID = 69366



USRDS Dialysis Data

ID = 71650



“Repeated Measures” Data

CPP Data

- Subset of $N = 8467$ women
- A total of $8467 + 1960 + 527$ births
- Recruitment during 1959-1965
- Prospective measurements: smoking, weight, birth outcomes.
- Goal: risk factors for poor pregnancy outcome?
 - ▷ Exposures: Infection, Drugs, Smoking.
 - ▷ Outcomes: Weight, SGA, malformations.
- Multiple births/woman recorded if within 1959-1965.

The Processes

Primary Process

$Y_i(t)$ The response process

Secondary Processes

$X_i(t)$ The covariate process

$S_i(t)$ The scheduling process (**not today**)

$R_i(t)$ The recording process (**not today**)

$D_i(t)$ The death process (**not today**)

$B_i(t)$ The birth process (**not today**)

LDA and Regression

- Most statistical representations focus on discussion of

$$\mu_i(t) = E[Y_i(t) | X_i(t)]$$

- But what about the other processes? Do we mean:

$$\text{CFF} : E[Y_i(t) | X_i(t), X_i(s), S_i(t) = 1, R_i(t) = 1, D_i(t) = 0]$$

$$\text{USRDS} : E[Y_i(t) | X_i(t), X_i(s), R_i(t) = 1, D_i(t) = 0]$$

$$\text{MSCM} : E[Y_i(t) | X_i(t), X_i(s), R_i(t) = 1]$$

$$\text{CPP} : E[Y_i(t) | X_i(t), X_i(s), dB_i(t) = 1, R_i(t) = 1]$$

Time-dependent Covariates

- **Exogenous** $Y_i(t)$ doesn't predict $X_i(s)$, $s > t$.
 - ▷ Appropriate lags: $E[Y_i(t) | X_i(s), s < t]$
 - ▷ Weighted estimation and bias (DHLZ section 12.3)
- **Endogenous** $Y_i(t)$ does predict $X_i(s)$, $s > t$.
 - ▷ Causal inference targets
 - * Model covariate process – MSMs
 - See: Robins, Hernán & Brumback (2000)
 - * Model response process – G-comp
 - See: Robins, Greenland, & Hu (1999)
 - ▷ Association
 - * Model covariate process
 - See: Miglioretti & Heagerty (2004)

Motivation: Hospitalization and EPO Dose?

- Background:

- ▷ **NEJM – November 2006**

- * RCTs target high versus low hemoglobin
- * Higher target → higher Epo dose
- * Higher target associated with AEs

- ▷ **FDA – March 2007**

Issued a “black box warning” which indicated that aggressive use of erythropoiesis-stimulating agents to raise hemoglobin to a target of 12 g/dL or higher was associated with “serious and life-threatening side-effects and/or death.”

- General Question:

- ▷ **Q:** Are higher doses of EPO associated with greater rates of adverse events such as hospitalization?

Motivation: Full Data History

- Regression:

$$E[\text{Hosp}(t) \mid \text{Dose}(t - 1), \text{Dose}(t - 2), \dots, X]$$

- Statistical Issues:

- ▷ What aspects of **exposure history** are associated with current hosp?
- ▷ What is the role of the **outcome history**
 $\text{Hosp}(t - 1), \text{Hosp}(t - 2), \dots?$
- ▷ What is the role of **intermediate history**
 $\text{Hem}(t - 1), \text{Hem}(t - 2), \dots?$

Time-dependent Covariates: Lagged Covariates

- **Exogenous** – future covariates are not influenced by current / past outcomes.

$$[X(t + 1) | Y(t), X(t)] \sim [X(t + 1) | X(t)]$$

- **Analysis Issues:**

- ▷ Include single lagged covariates (current, cumulative)

- * **MSCM**: $E[\text{Sick}(t) | \text{Stress}(t - k)]$

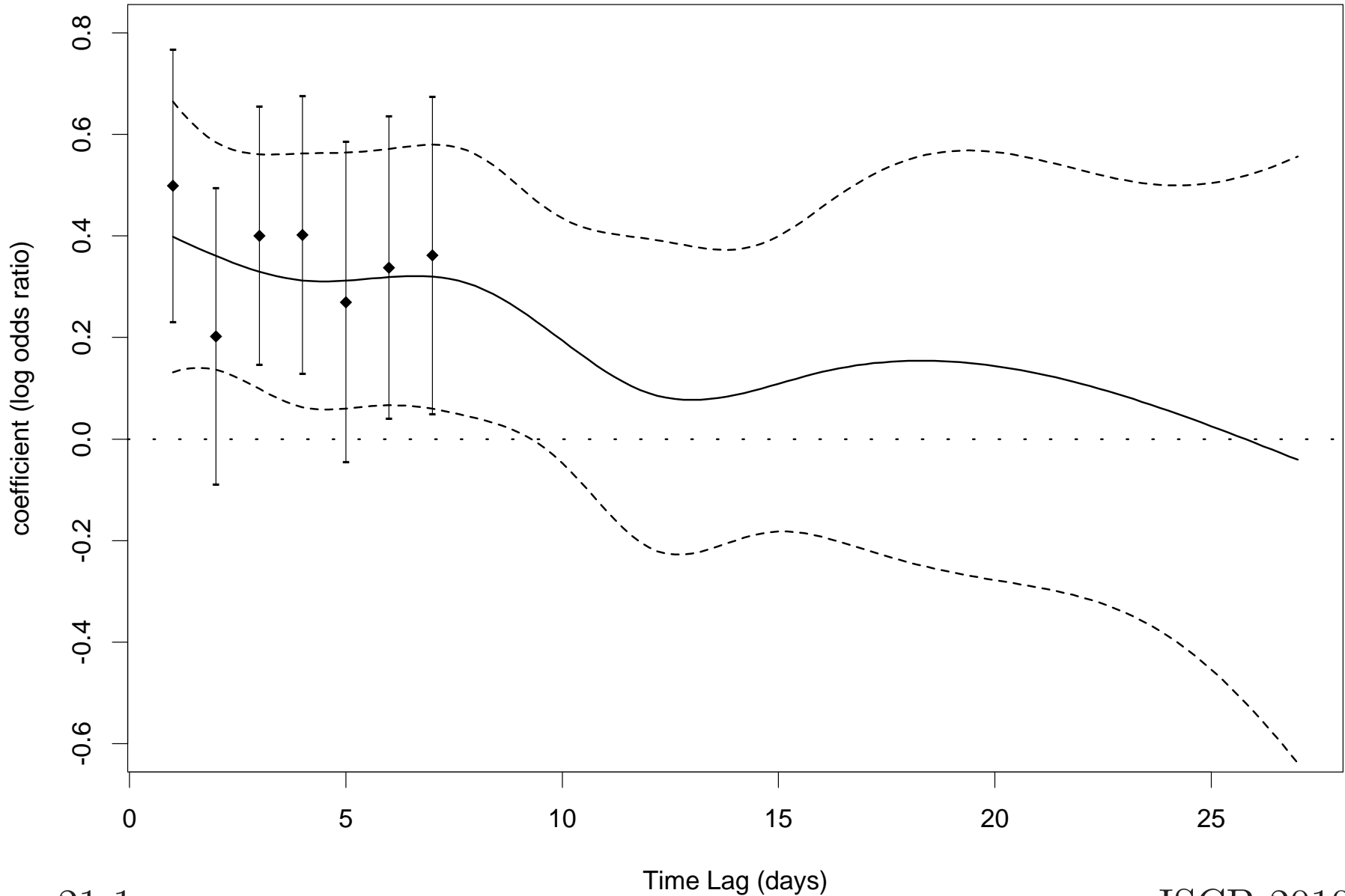
- * **USRDS**: $E[\text{Hosp}(t) | \text{Dose}(t - k)]$

- ▷ Include multiple lagged covariates

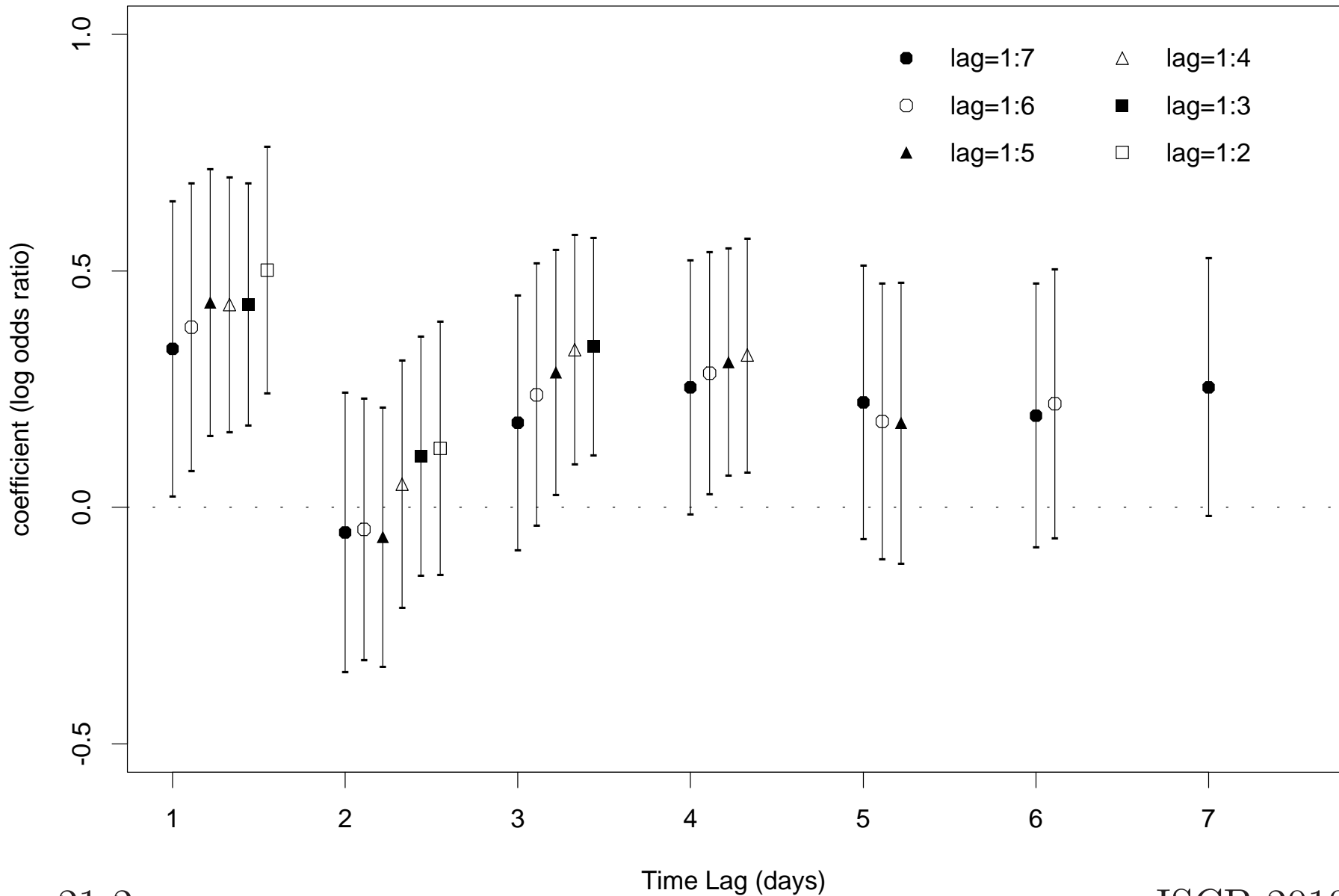
- * **MSCM**: $E[\text{Sick}(t) | \text{Stress}(t - 1), \text{Stress}(t - 2)]$

- * **USRDS**: $E[\text{Hosp}(t) | \text{Dose}(t - 1), \text{Dose}(t - 2)]$

Lag Coefficient Function



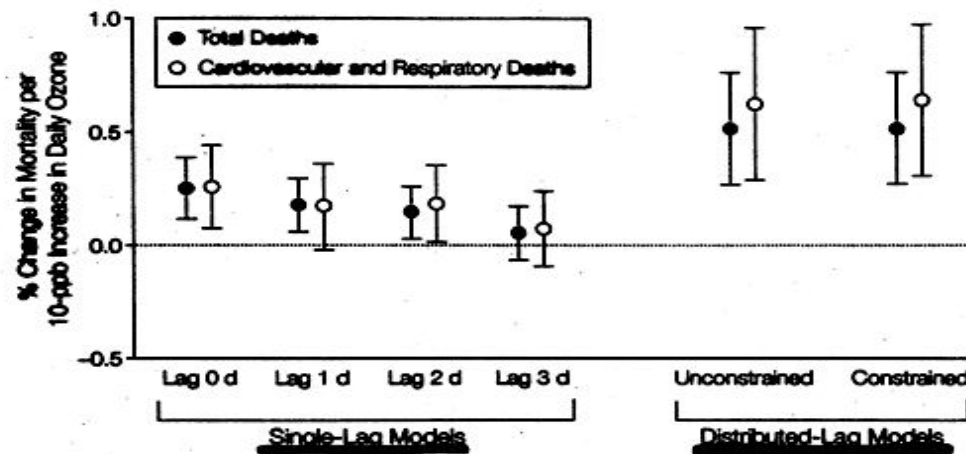
Multivariate models with different lags



Bell et al. *JAMA* (2004)

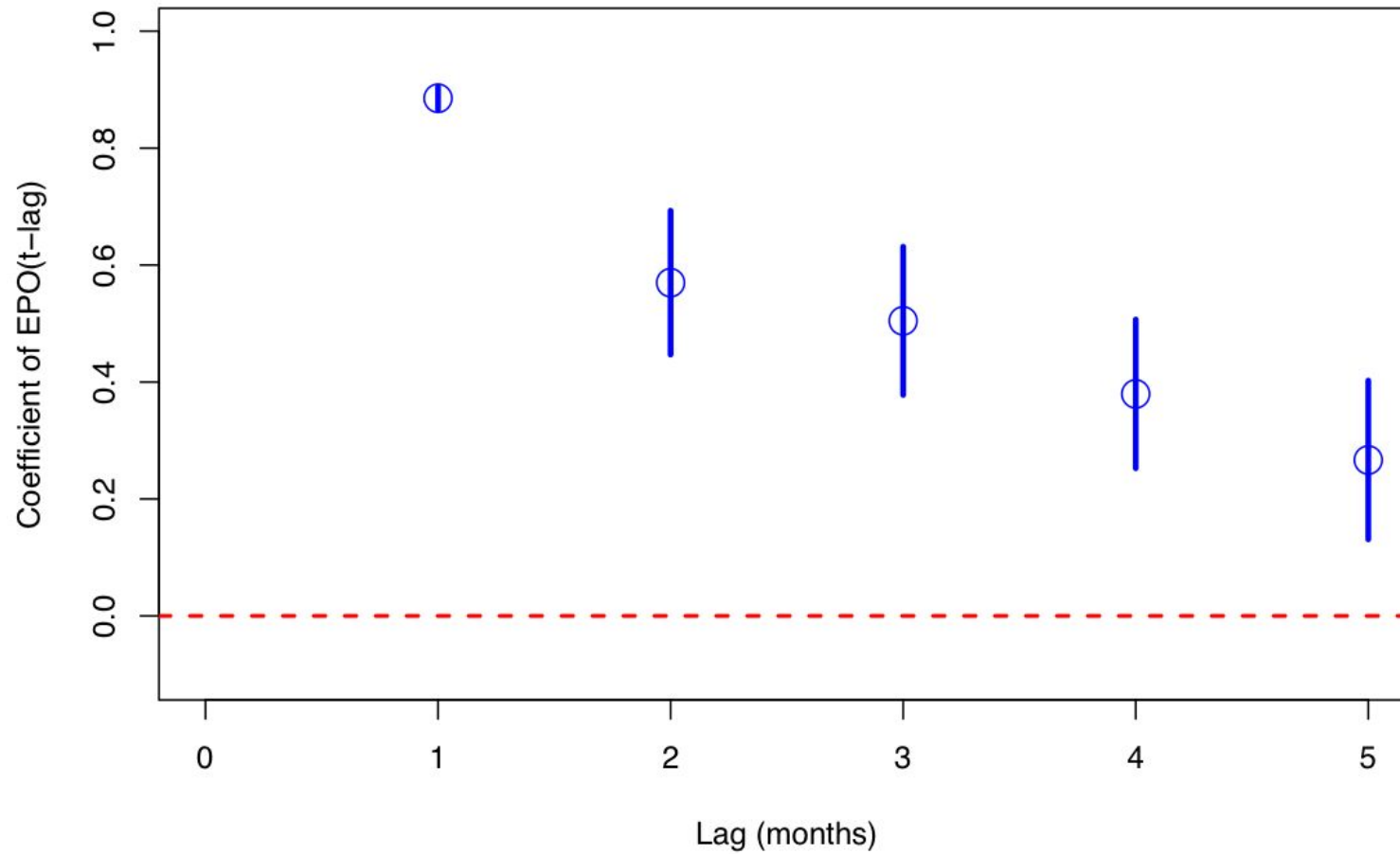
OZONE AND MORTALITY IN US URBAN COMMUNITIES

Figure 1. Percentage Change in Daily Mortality for a 10-ppb Increase in Ozone for Total and Cardiovascular Mortality, for Single-Lag and Distributed-Lag Models

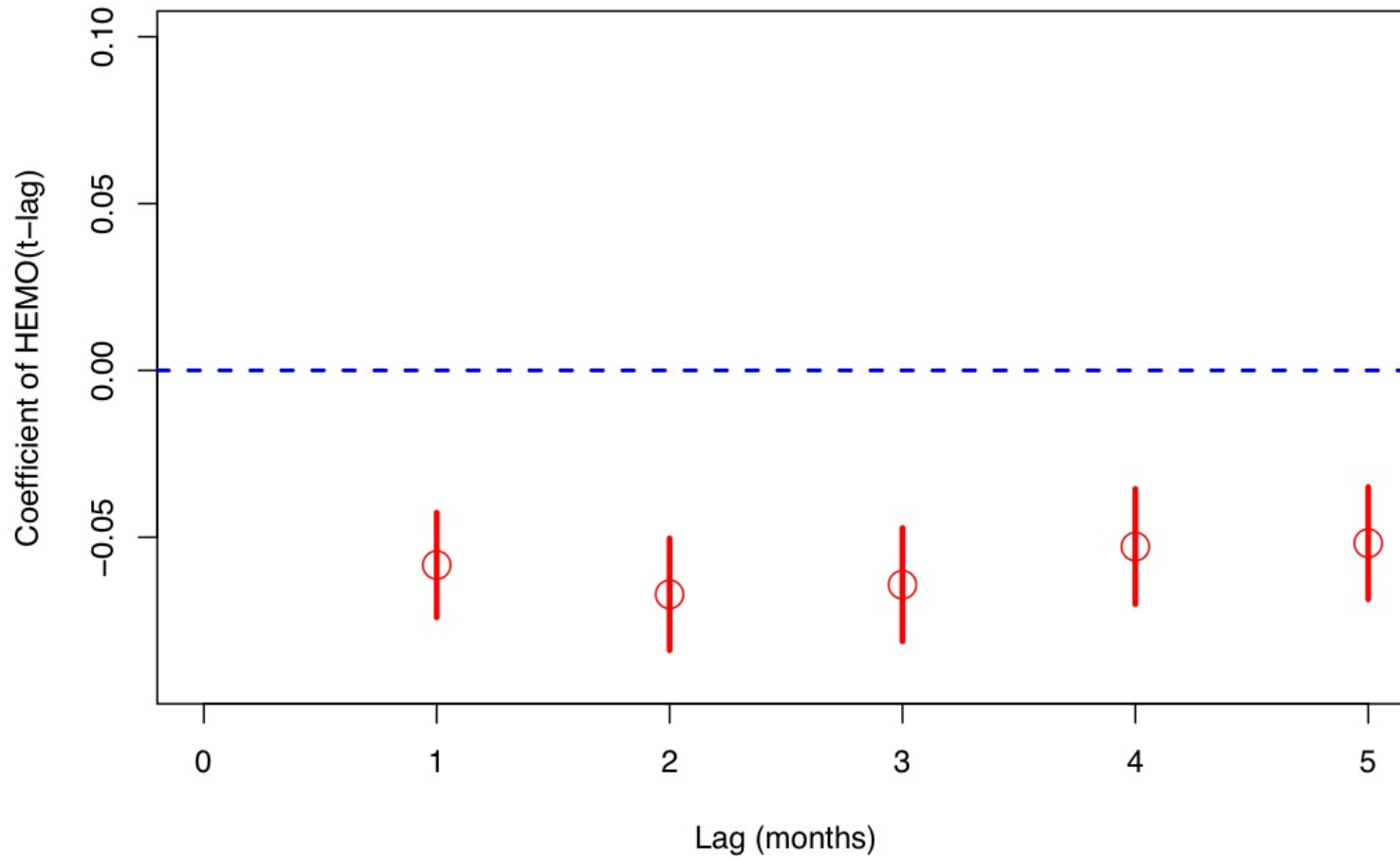


The single-lag model reflects the percentage increase in mortality for a 10-ppb increase in ozone on a single day. The distributed-lag model reflects the percentage change in mortality for a 10-ppb increase in ozone during the previous week. Error bars indicate 95% posterior intervals.

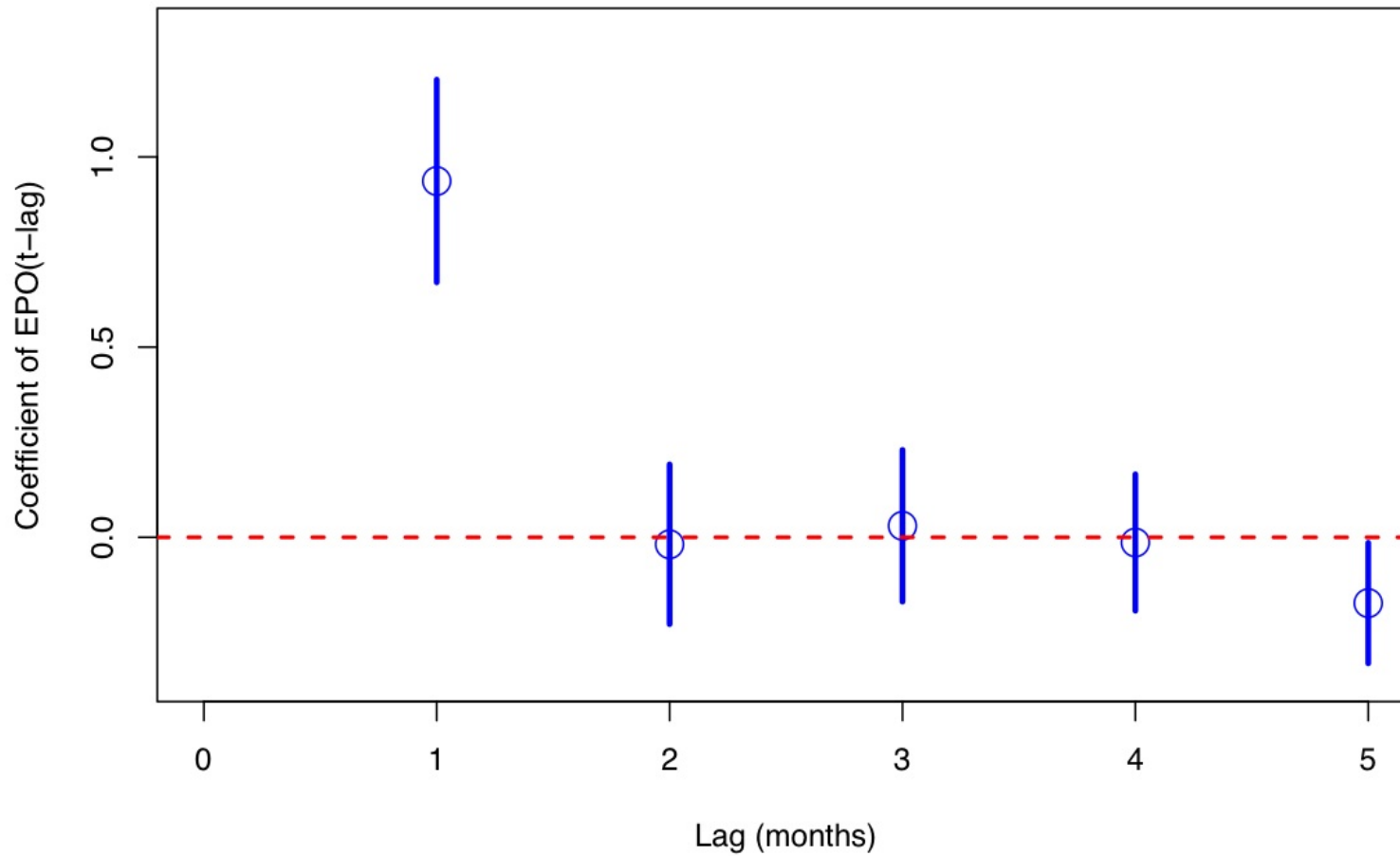
Single lag coefficients for $\text{Death}(t) \sim \text{Epo}(t-\text{lag})$



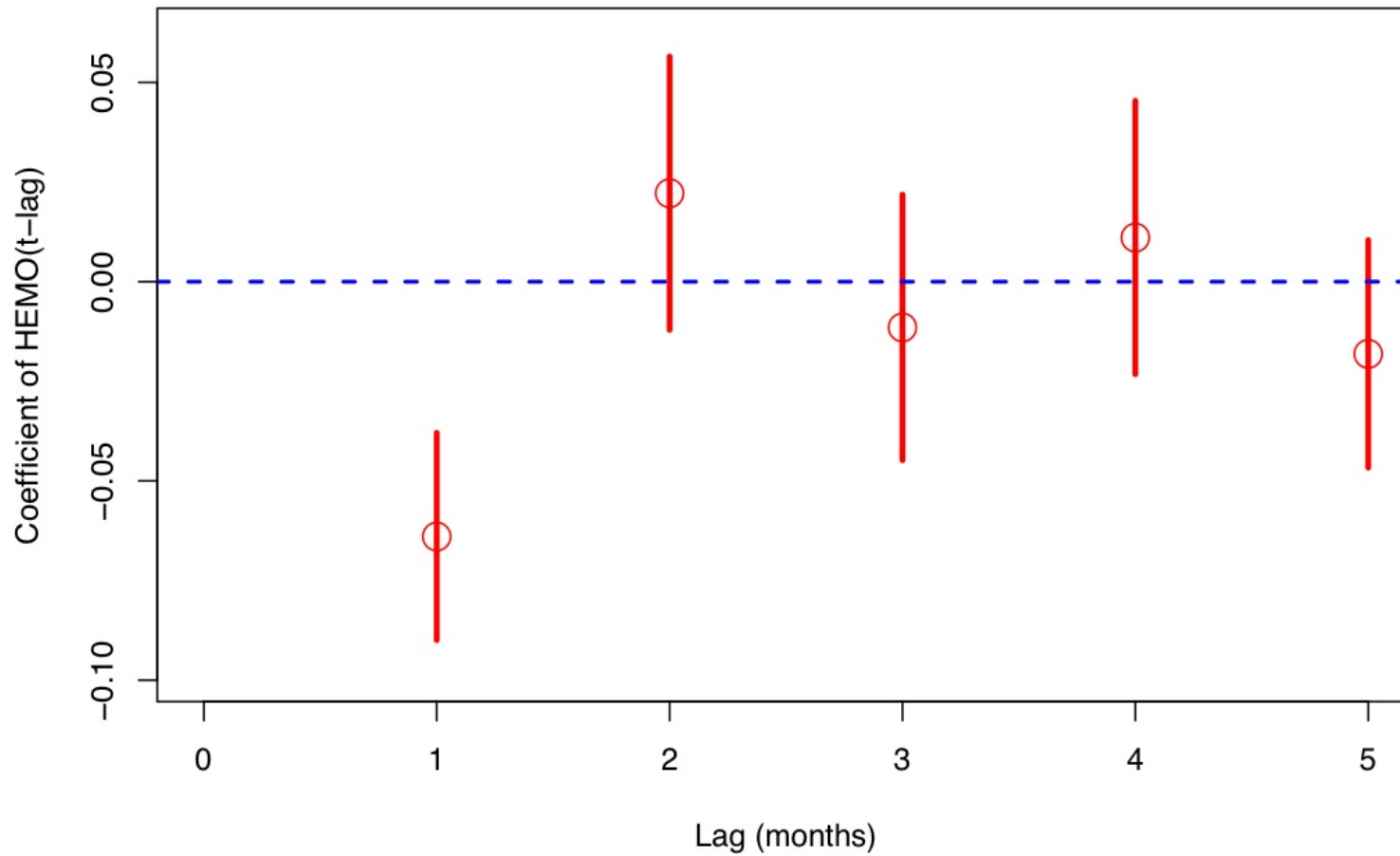
Single lag coefficients for $\text{Death}(t) \sim \text{Hemo}(t-\text{lag})$



Coefficients for $\text{Death}(t) \sim \text{Hemo}(t-1) + \text{Epo}(t-1) + \dots$



Coefficients for $\text{Death}(t) \sim \text{Hemo}(t-1) + \text{Epo}(t-1) + \dots$



Endogenous: Analysis

- **Definition:** – The covariate is influenced by past outcomes (or intermediate variables)

$$Y(t) \rightarrow X(t + 1)$$

- **Implication:**

$$E[Y_i(t) \mid X_i(1), \dots, X_i(n)]$$

depends on $X_i(s)$ for $s > t$ (future values of covariate).

- Role for causal inference concepts.
- See: DHLZ (2002) section 12.5 for introduction.

Causal Targets of Inference

- Longitudinal Treatment

$$\text{vec}(X_0) \equiv [X(1) = \mathbf{0}, X(2) = \mathbf{0}, \dots, X(n) = \mathbf{0}]$$

$$\text{vec}(X_1) \equiv [X(1) = \mathbf{1}, X(2) = \mathbf{1}, \dots, X(n) = \mathbf{1}]$$

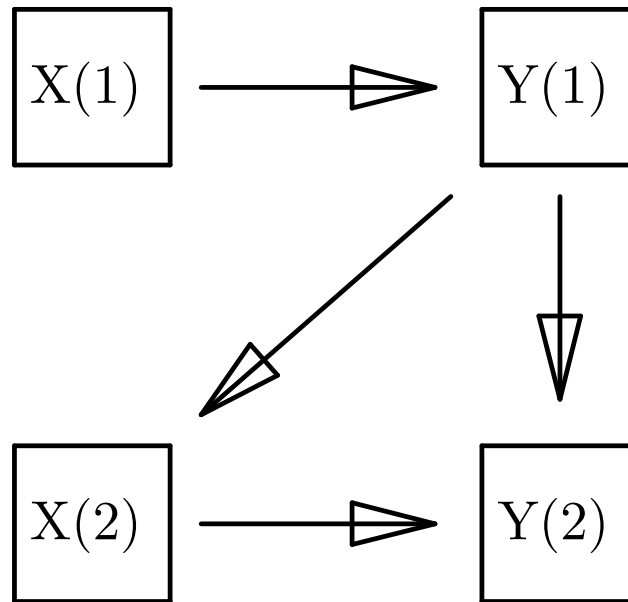
- Population Means

- ▷ Mean of population if all subjects had $X = 1$ at all times, and similar population mean if $X = 0$ at all times.

$$\mu_0(n) \equiv E[Y(n) \mid \text{vec}(X_0)]$$

$$\mu_1(n) \equiv E[Y(n) \mid \text{vec}(X_1)]$$

Endogenous Covariates

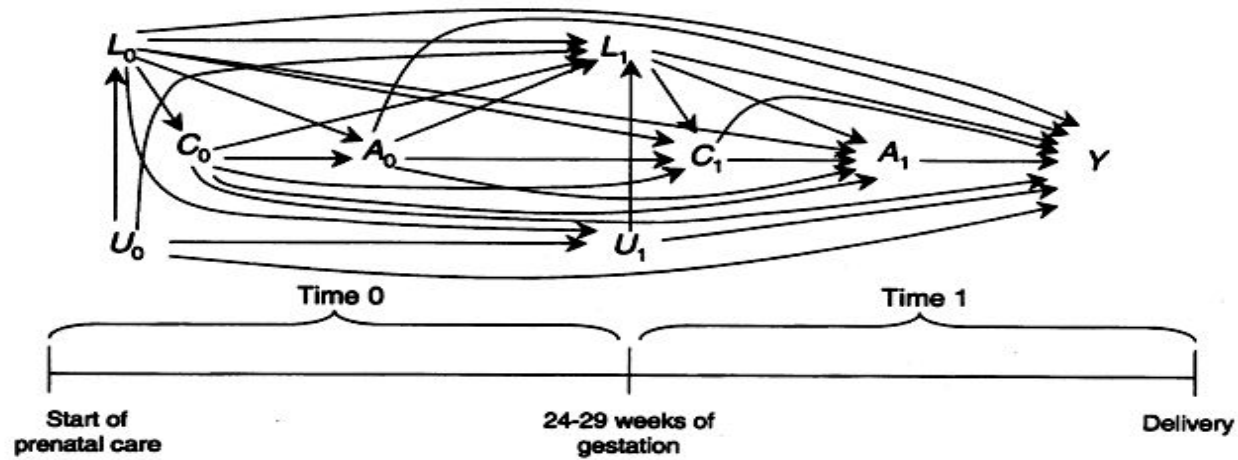


**treatment /
exposure**

response

Bodnar et al. *AJE* (1997)

MSMs for Causal Effects of Time-dependent Treatment



Model / Estimation

- **G-computation**

- ▷ **Model:** model **outcome** given past outcomes / exposure.

$$P[Y(t) | X(t), \{Y(s), X(s)\} s < t] \quad : \quad \text{outcome}$$

$$P[X(t) | \{Y(s), X(s)\} s < t] \quad : \quad \text{exposure}$$

- ▷ **Compute:** compute means of interest by allowing intermediate effects, $Y(s)$, to occur naturally, but controlling exposure.

$$\mu_1(t) = E_t \{ E_s [Y(t) | \mathbf{X}(t)=1, \{Y(s), \mathbf{X}(s)=1\} s < t] \}$$

Model / Estimation

- **Marginal Structural Models**

- ▷ **Model:** model **exposure** given past outcomes / exposure.

$$[X(t) \mid \{Y(s), X(s)\} s < t]$$

- ▷ **Compute:** compute a regression of the outcome using inverse probability weights (IPW) to control for exposure selection bias.

Table 1: Regression of stress, S_{it} , on illness, I_{it-k} $k = 0, 1$, and previous stress, S_{it-k} $k = 1, 2, 3, 4+$ using GEE with working independence.

	est.	s.e.	Z
(Intercept)	-1.88	(0.36)	-5.28
I_{it}	0.50	(0.17)	2.96
I_{it-1}	0.08	(0.17)	0.46
S_{it-1}	0.92	(0.15)	6.26
S_{it-2}	0.31	(0.14)	2.15
S_{it-3}	0.34	(0.14)	2.42
mean($S_{it-k}, k \geq 4$)	1.74	(0.24)	7.27
employed	-0.26	(0.13)	-2.01
married	0.16	(0.12)	1.34
maternal health	-0.19	(0.07)	-2.83
child health	-0.09	(0.07)	-1.24
race	0.03	(0.12)	0.21
education	0.42	(0.13)	3.21
house size	-0.16	(0.12)	-1.28

Table 2: MSM estimation of the effect of stress, S_{it-k} $k \geq 1$, on illness, I_{it} .

	est.	s.e.	Z
(Intercept)	-0.71	(0.40)	-1.77
S_{it-1}	0.15	(0.14)	1.03
S_{it-2}	-0.19	(0.18)	-1.05
S_{it-3}	0.18	(0.15)	1.23
mean(S_{it-k} , $k \geq 4$)	0.71	(0.43)	1.65
employed	-0.11	(0.21)	-0.54
married	0.55	(0.17)	3.16
maternal health	-0.13	(0.10)	-1.27
child health	-0.34	(0.09)	-3.80
race	0.72	(0.21)	3.46
education	0.34	(0.22)	1.57
house size	-0.80	(0.18)	-4.51

method	logOR
GEE cross-sectional association	0.66
GEE with seven days lagged	1.38
Transition model (direct effect)	0.50
G-computation	0.80
MSM	0.85

Summary of Endogenous

- Interest in exposure over time – more than simply the acute (most recent) exposure.
- A variable (perhaps outcome) is both a **consequence** of exposure at early times, and a **cause** of exposure at later times.
- Intermediate and confounder.
- **G-computation**
- **MSM**
- Interest in outcomes under a controlled and **static** treatment plan.

EPO: November 2006 NEJM

- **Drüeke** CREATE
 - ▷ Control **Hemoglobin** rather than fix the dose.
 - * Low group (11.0-12.5)
 - * Normal group (13.0-15.0)
- **Singh** CHOIR
 - ▷ Control **Hemoglobin** rather than fix the dose.
 - * Low group (11.3)
 - * Normal group (13.5)
- **Research Question(s)**

Q: What target hemoglobin should be used? How to use observational data to compare different targets and/or compare mortality experience to RCT data?

Analysis of Dynamic Treatment

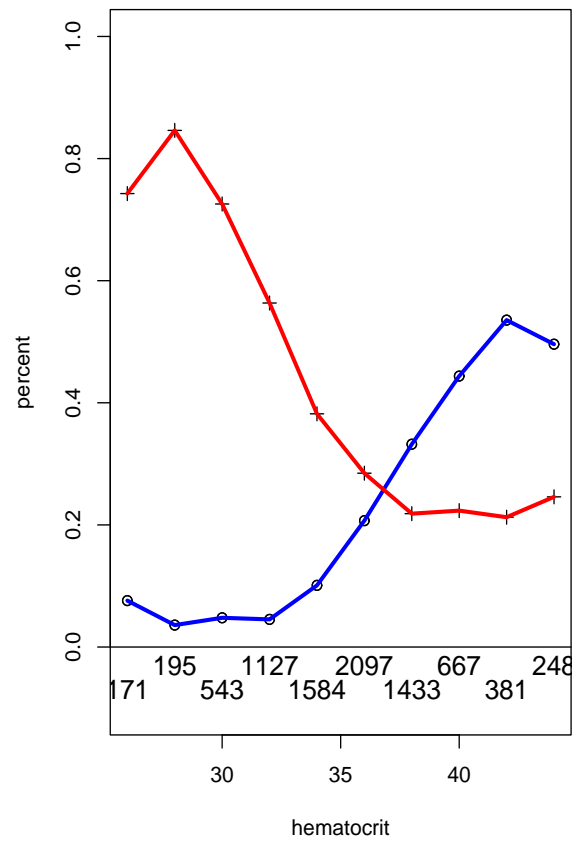
- **Note:** The guidelines for Epo do not suggest a **static** dose be administered. Rather, dose is driven by the state of the intermediate (Hb):

$$X(t+1) = \begin{cases} 1.25 \times X(t) & \text{if } Z(t) \leq 11 \\ X(t) & \text{if } 11 < Z(t) \leq 13 \\ 0.75 \times X(t) & \text{if } Z(t) > 13 \end{cases}$$

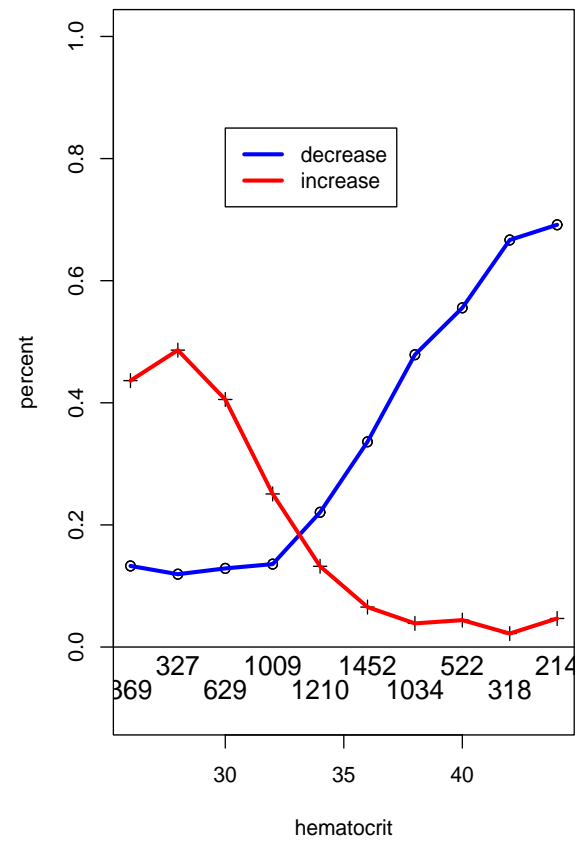
- This corresponds to a **dynamic treatment guideline**, \mathcal{G}_1 .
- **Q:** How to formulate DOSE questions in this setting?
 - ▷ \mathcal{G}_1 corresponds to correction of $\pm 25\%$ at Hb=(11,13).
 - ▷ Compare to a \mathcal{G}_2 which uses alternative target Hb threshold(s).

USRDS Data (2003 sample)

25% or more Change -- LOW Epo



25% or more Change -- HIGH Epo



Some Literature on Dynamic Treatment

- Hernán et al. (2006) “Comparison of dynamic treatment regimes via inverse probability weighting” *Basic & Clin Pharm & Tox*, 98; 237-242.
- Robins, Orellana, and Rotnitzky (2008) “Estimation and extrapolation of optimal treatment and testing strategies” *Statistics in Medicine*, 27: 4678-4721.
- Cotton and Heagerty (2010) “Inference for the comparison of dynamic treatment regimens with application to epoetin dosing strategies” (submitted)

(*) Drop-out

- Extensive literature has been developed. See summaries:
 - ▷ Book: DHLZ (2002) – Chapter 13
 - ▷ Book: (Geert)² – Chapters 14-21(!)

- Main issue is selection bias (Little & Rubin, 1987):

$$E[Y_i(t) | X_i, R_i(t) = 1] \neq E[Y_i(t) | X_i]$$

- Main approaches to analysis with MAR, (NI sensitivity):
 - ▷ **Inverse Probability Weighting (IPW)**
 - * See: Robins, Rotnitzky & Zhao (1995)
 - ▷ **Correctly specified likelihood**
 - * See: Laird (1988)

(*) LDA with Death

- Different than drop-out

- With Drop-out:

$$E[Y_i(t) | X_i] = E[Y_i(t) | X_i, R_i(t) = 1] \times P[R_i(t) = 1 | X_i] + E[Y_i(t) | X_i, R_i(t) = 0] \times P[R_i(t) = 0 | X_i]$$

- Linear Mixed Models (LMM) applied to the observed data where $R_i(t) = 1$ can validly estimate parameters in the mean $E[Y_i(t) | X_i]$ when data are MAR.

- With Death:

$$E[Y_i(t) | X_i] = E[Y_i(t) | X_i, D_i(t) = 0] \times P[D_i(t) = 0 | X_i] + E[Y_i(t) | X_i, D_i(t) = 1] \times P[D_i(t) = 1 | X_i]$$

(*) LDA with Death: Analysis

- Analysis conditional on death information:

- ▷ Full (future) stratification:

$$E[Y_i(t) \mid X_i(t), \mathbf{T}_i = \mathbf{s}] \quad s > t$$

- * See: Pauler, McCoy & Moinpour (2003)

- ▷ Partial (current status) conditioning:

$$E[Y_i(t) \mid X_i(t), \mathbf{T}_i > \mathbf{t}]$$

- * See: Kurland and Heagerty (2004)

- ▷ Conditional on principal strata (potential status):

$$E[Y_i(t \mid 1) - Y_i(t \mid 0) \mid \{\mathbf{T}_i(\mathbf{0}) > \mathbf{t}, \mathbf{T}_i(\mathbf{1}) > \mathbf{t}\}]$$

- * See Frangakis and Rubin (2002), Rubin (2007)

(*) LDA with Death: Comments on Analysis

- Full stratification using $[\mathbf{T}_i = s] \quad s > t$
 - ▷ Compares groups defined by X_i comparable in terms of death.
 - ▷ Conditions on future (not yet observed) information.
- Partial (current status) conditioning: $[\mathbf{T}_i > t]$
 - ▷ Conditions on observed vital status.
 - ▷ Compares groups defined by X_i after selection by death.
- Principal stratification: $[\{\mathbf{T}_i(\mathbf{0}) > t, \mathbf{T}_i(\mathbf{1}) > t\}]$
 - ▷ Compares subgroups defined by X_i comparable in terms of death.
 - ▷ Conditions on unobservable potential status.

(* Measurement Process

- Health status (current) may influence the scheduling, $S_i(t)$, of administrative data collection.
- Does past/current/future outcome (health status) predict time-until next visit?

$$\lambda_k(t \mid t_{i1}, \dots, t_{ik-1}, Y_i(t), X_i)$$

- Potential for bias if visit intensity process is not independent of response process.

(* Measurement Process: Analysis

- Estimating Equations: Lin, Scharfstein & Rosenheck (2004)
 - ▷ Assume λ_k independent of current/future $Y_i(t)$, but may depend on observed past (including auxiliary variables).
 - ▷ Model λ_k and use inverse intensity weights (IIW) GEE.
 - ▷ Semiparametric.
- Likelihood: Lipsitz et al. (2002)
 - ▷ Assume λ_k only depends on $Y_i(t-)$ and not on t_{i1}, \dots, t_{ik} (given Y).
 - ▷ Fully model the response and use maximum likelihood.
 - ▷ Parametric.

(* Birth Processes

- When discussion of missing data or measurement processes we assume $Y_i(t)$ is measurable at any time t .
- For the Collaborative Perinatal Project (CPP) research focuses on measurements that are associated with a (recurrent) event time.
- Birth Process:
 - ▷ $B(t)$ is counting process for number of births for subject i through time t .
 - ▷ The birth outcome, $Y_i(t)$, is only defined at the time of birth: when $dB(t) = 1$.
- Regression analysis for (recurrent) marked point process data

$$E[Y_i(t) \mid X_i(t), dB_i(t) = 1]$$

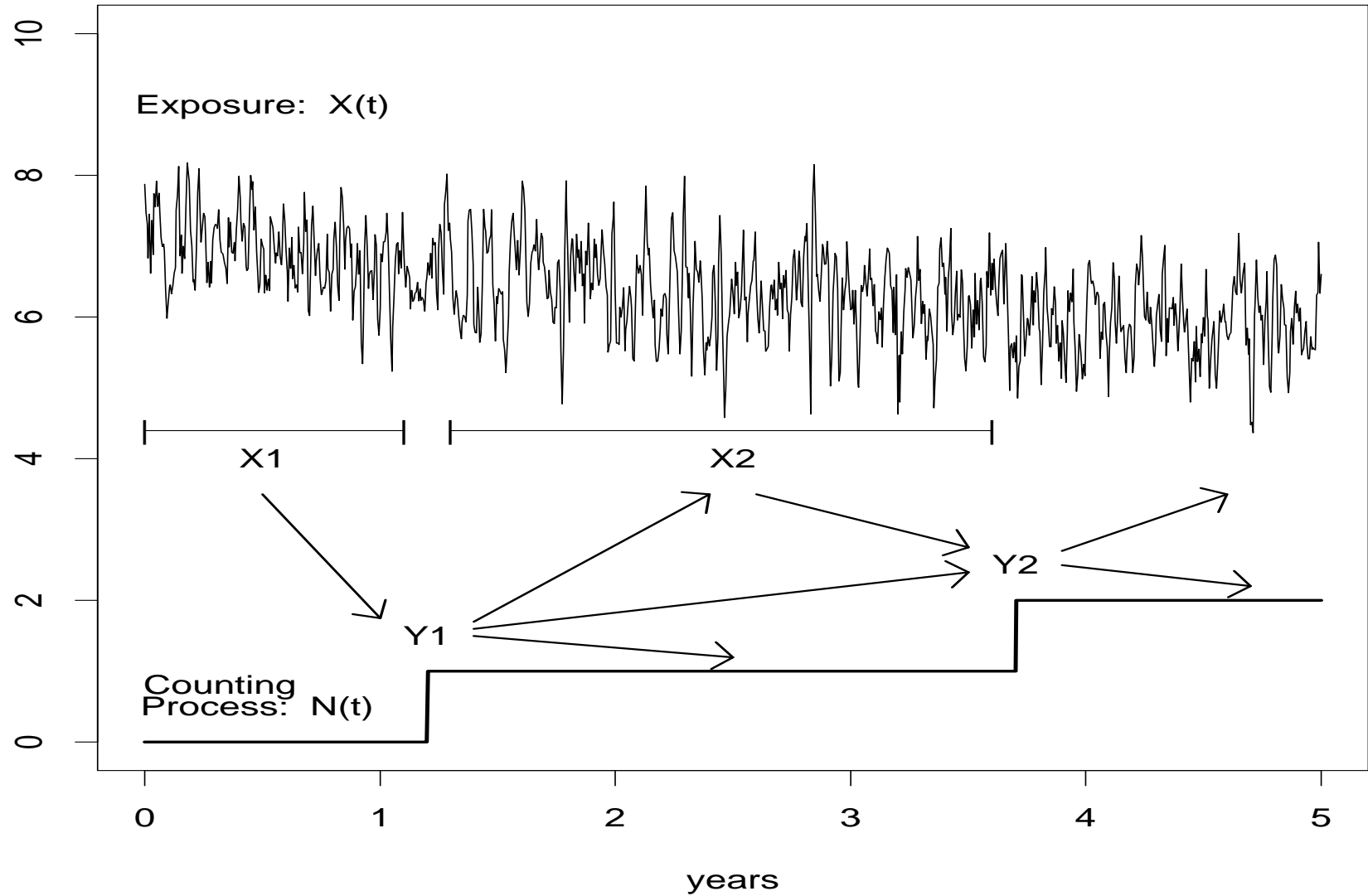
(*) Birth Processes: Mark Regression Analysis

- For some analyses the past response is an intermediate outcome and interest is in marginal means:

$$E[Y_i(t) \mid X_i(t), dB_i(t) = 1, H_i^X(t-), H_i^B(t-)]$$

- If $Y_i(t)$ predicts time-until future birth, $B_i(t+)$, then this endogeneity implies standard non-independence GEE or LMM will give biased estimates.
- If endogeneity then IEE valid characterization of association between birth outcome and prior exposure (and parity).
- French and Heagerty (2009)

Birth Data and Exposure



(*) Birth Processes: simulations

- Scenario = probability of the next birth depends on the last birth outcome if there was one. Babies per mom is still between 1 and 5 with mean 2-3.

Estimation method	Intercept est.		Slope est.	
	mean	(std dev)	mean	(std dev)
OLS using first born	1.048	(0.352)	1.019	(0.181)
GEE with working independence	1.014	(0.041)	1.004	(0.038)
Linear Mixed ML random intercept only	0.741	(0.063)	0.914	(0.039)

(*) Inclusion: CPP Data

- Cox model for time-until-second birth

covariate	est.	exp(est.)	s.e.(est.)	Z	p
weight(2)	0.1649	1.179	0.0861	1.914	0.056
weight(3)	0.1825	1.200	0.0798	2.285	0.022
weight(4)	-0.0152	0.985	0.1501	-0.101	0.920
site(2)	-0.3497	0.705	0.0895	-3.906	<0.001
site(3)	-0.6911	0.501	0.0841	-8.218	<0.001
site(4)	-0.0943	0.910	0.0847	-1.114	0.271
race(black)	0.0885	1.093	0.0844	1.048	0.293
race(other)	-0.0289	0.972	0.1501	-0.192	0.851
smoke	0.0594	1.061	0.0462	1.287	0.200

	First born (n=8467)		Second born (n=1960)		Third born (n=527)		All births (n=10954)		All births (n=10954)	
	Linear Regn		Linear Regn		Linear Regn		GEE		Mixed Model	
	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.	est.	s.e.
Intercept	3280.5	(11.7)	3372.8	(23.2)	3334.0	(44.7)	3284.4	(10.9)	3281.7	(11.1)
SMOKE	-161.6	(13.1)	-179.5	(26.3)	-179.2	(49.7)	-164.7	(12.3)	-157.6	(11.8)
site(2)	33.7	(23.6)	-124.7	(51.1)	113.1	(136.7)	10.9	(21.1)	17.1	(22.4)
site(3)	45.9	(19.6)	-36.2	(51.6)	96.4	(105.7)	36.6	(17.8)	39.8	(18.8)
site(4)	-84.8	(25.5)	-68.4	(56.5)	-146.3	(113.5)	-85.4	(22.7)	-88.9	(24.2)
race(black)	-212.6	(24.5)	-221.9	(56.4)	-110.2	(113.3)	-208.3	(22.6)	-210.1	(24.2)
race(other)	-180.3	(40.7)	-200.1	(88.6)	-248.7	(197.9)	-184.3	(36.3)	-179.7	(38.7)
parity(2)							65.5	(14.8)	64.1	(12.3)
parity(3)							63.7	(26.5)	81.9	(23.5)

Some recommendations

- In applications we should identify factors that influence the **secondary stochastic processes** and choose appropriate statistical techniques in order to validly answer the scientific question.
- In statistical research reports we should be explicit about the **assumptions** we are making regarding the secondary stochastic processes.
- For **time-dependent covariates** ask about associations with both past and future covariate values – consider the factors that drive the covariate.