Statistical Methods for Alzheimer's Disease Studies

Rebecca A. Betensky, Ph.D.

Department of Biostatistics,

Harvard T.H. Chan School of Public Health

July 19, 2016

OUTLINE

Statistical collaborations in Alzheimer's disease research

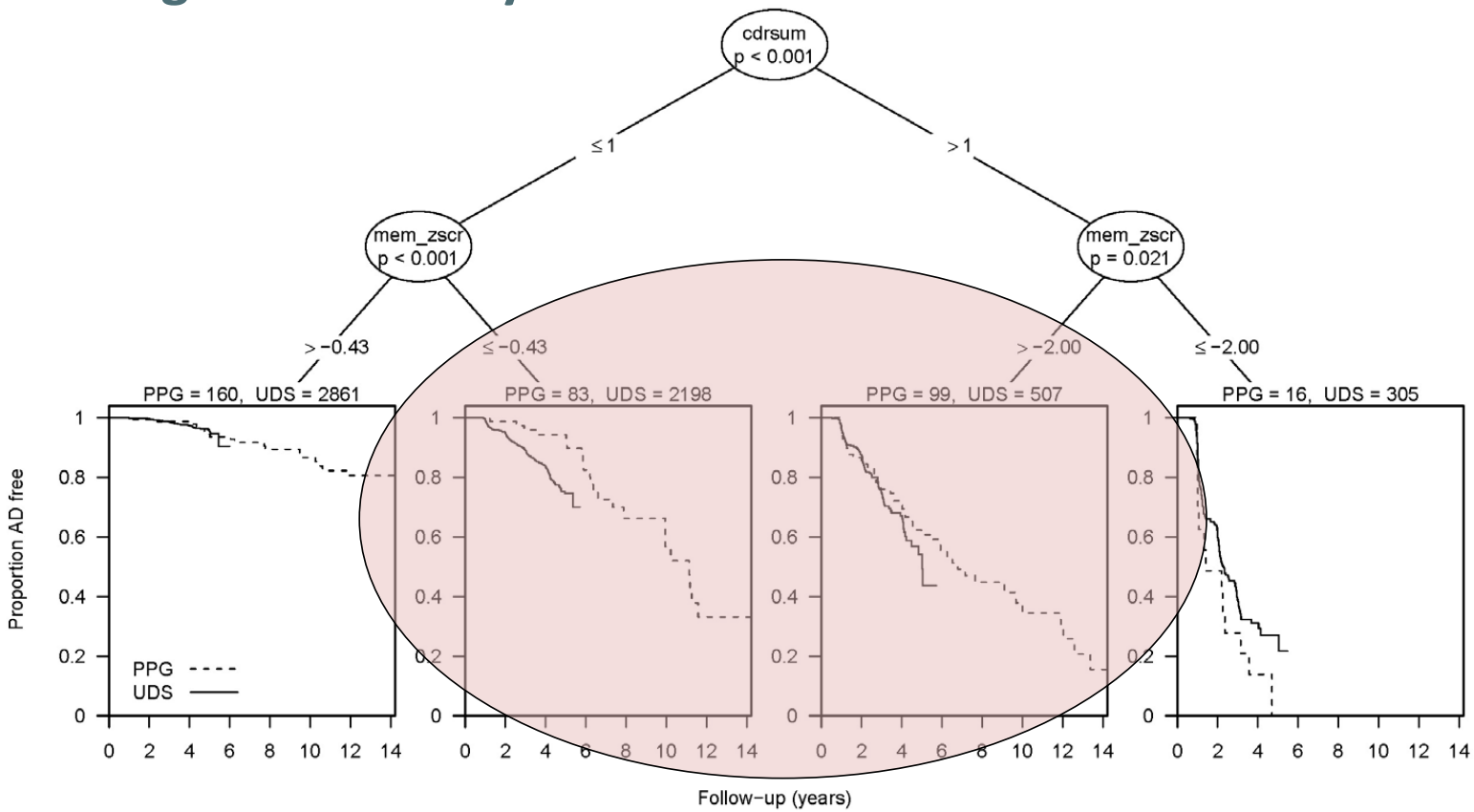
- Analysis of autopsy studies (selection bias)
- Selecting subjects and endpoints for efficient clinical trials (clinical trial design)
- Combining amyloid PET values across data sets (latent class analysis)
- Regression models with age of dementia onset: censored covariates

IPW model

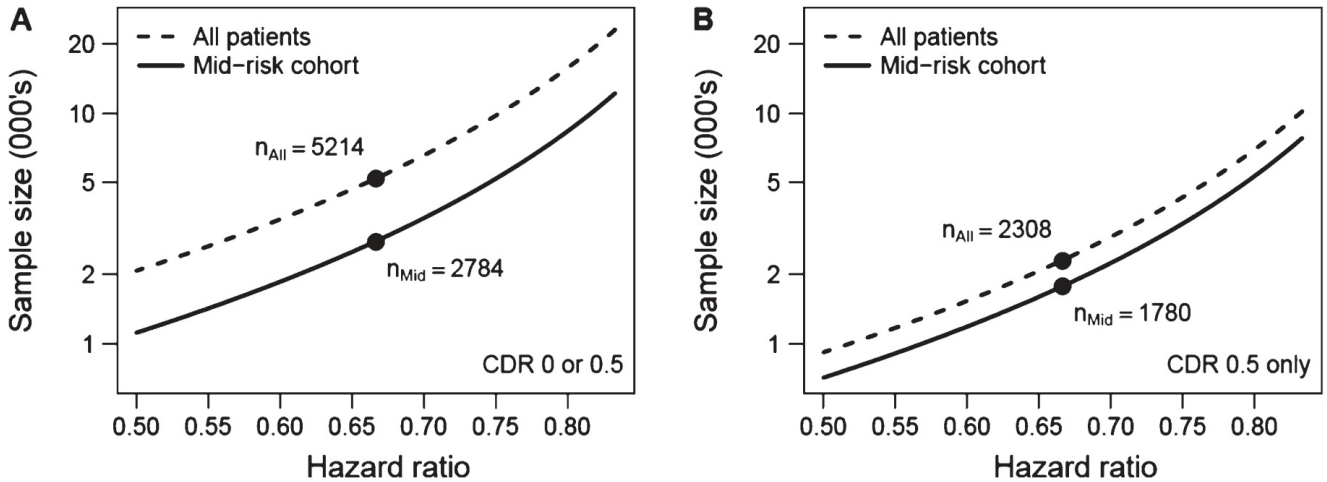
	Model 1 (only demographics)		Model 2 (Model 1+ neuritic plaques+NFTs)		Model 2 (Model 1 + neuritic plaques+NFTs)	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Sex (female as ref.)						
at *k=1,2,3,4*	1.12 (0.99, 1.26)	0.063	1.12 (0.96, 1.30)	0.148	1.13 (1.03, 1.24)	0.010
at *k=5*	0.55 (0.38, 0.79)	0.001	0.53 (0.36, 0.76)	0.001	0.51 (0.39, 0.66)	<0.001
Age of death (in 5-years unit)	0.89 (0.87, 0.92)	<0.001	0.91 (0.88, 0.94)	<0.001	0.90 (0.88, 0.92)	<0.001
Education (in 4-years unit)	0.90 (0.85, 0.95)	<0.001	0.85 (0.79, 0.92)	<0.001	0.85 (0.81, 0.89)	<0.001
CERAD (neuritic plaques) (none/sparse as ref.)						
moderate			1.45 (1.21, 1.74)	<0.001	1.44 (1.29, 1.61)	<0.001
frequent			1.85 (1.51, 2.25)	<0.001	1.78 (1.58, 2.00)	<0.001
NFTs (Braak) (none/I/II as ref.)						
Stage III/IV						
at *k=1*			4.75 (2.55, 8.85)	<0.001	4.82 (3.46, 6.73)	<0.001
at *k=2,3,4,5*			1.06 (0.81, 1.39)	0.660	1.02 (0.87, 1.21)	0.794
Stage V/VI						
at *k=1*			9.84 (3.41,	<0.001	8.16 (4.74, 14.04)	<0.001
at *k=2,3,4,5*			2.15 (1.63, 2.83)	<0.001	2.22 (1.87, 2.63)	<0.001

Derivation of “mid-risk” cohort from existing longitudinal study

a



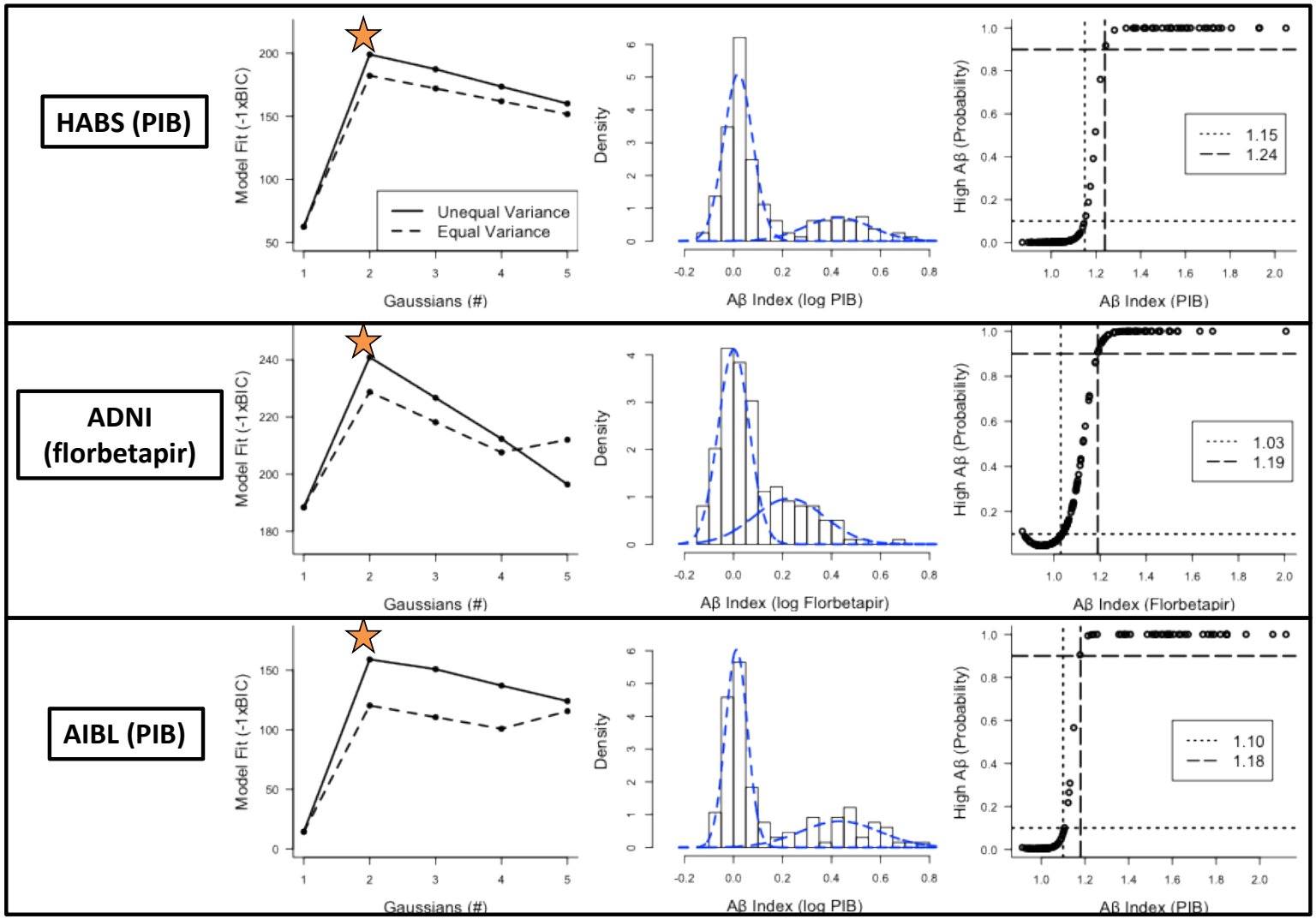
Selection of mid-risk patients could decrease sample size required



- Sample sizes for 80% power to detect hazard ratio of 0.67
- Without assumption that subjects within 2 years of dementia do not benefit, required sample sizes for unselected population would be 3598 and 2402 (all non-demented) and 1408 and 1370 (CDR 0.5)

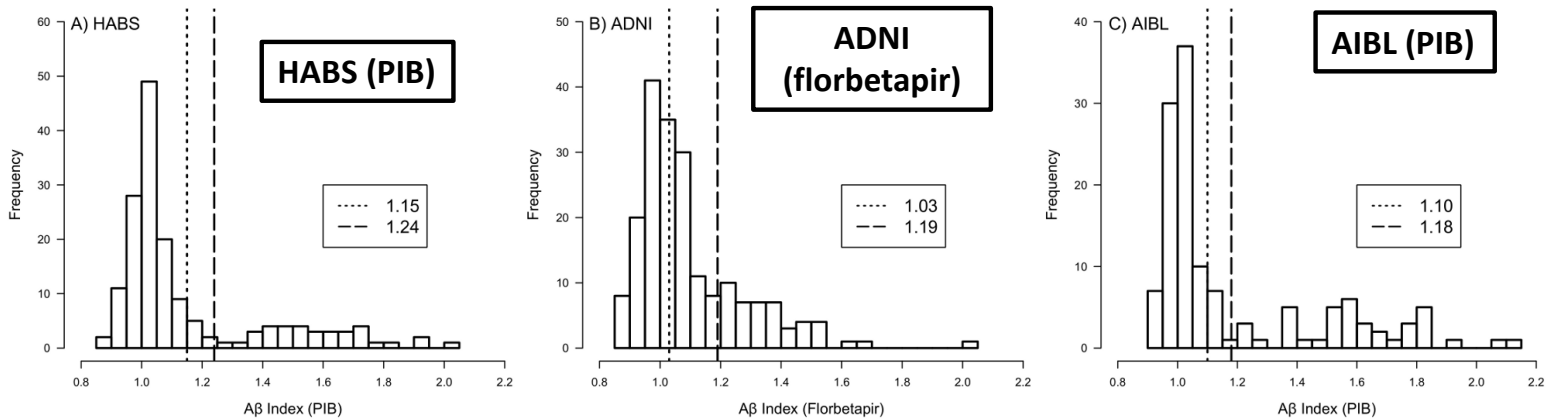
67

Applying Gaussian Mixture Modeling



2 distributions selected for each cohort

Many more “ambiguous” cases in ADNI/florbetapir



	HABS	ADNI	AIBL
N	161	198	131
Age*#⊗	74.3 (6.1)	76.2 (6.5)	72.3 (6.9)
Low Education (%)*#	21 (13.0%)	20 (10.1%)	56 (42.7%)
Female (%)	88 (55%)	99 (50%)	68 (52%)
APOE4+ (%)	41 (25%)	51 (25.8%)	46 (35%)
MMSE#	29.1 (0.9)	29.2 (1.0)	28.9 (1.1)
Logical Memory, Immediate Recall*#	15.1 (3.4)	15.0 (3.1)	13.0 (3.6)
Logical Memory, Delayed Recall*#	13.9 (3.3)	14.2 (3.4)	11.5 (3.7)
Baseline-Final Session (years) *#	1.46 (0.63)	1.41 (0.58)	2.7 (0.72)
Baseline session-PET separation (years) *#⊗	0.29 (0.21)	0.20 (0.28)	0.41 (0.30)
Aβ Index#	1.15 (0.25)	1.10 (0.18)	1.19 (0.30)
High Aβ	36 (22.3%)	48 (24.2%)	40 (30.5%)
Low Aβ #⊗	119 (73.9%)	95 (48.0%)	86 (65.6%)
Ambiguous Aβ #⊗	6 (3.7%)	55 (27.8%)	5 (3.8%)

High Aβ: >90% belonging to high distribution

Low Aβ: >90% belonging to low distribution

Ambiguous Aβ: everyone else

Amyloid and maternal age of onset

- The risk of AIzheimer's disease (AD) is known to increase dramatically with age
- Another major risk factor for AD: family history (FH)
- beta-amyloid (Aβ) deposition early event in pathological progression of AD; measurable via PET scan imaging
- A study was conducted at Mass General Hospital and Brigham and Women's Hospital to investigate the relationship between maternal age of onset of dementia and beta-amyloid deposition in cognitively normal older offspring (Maye et al, 2016).

The study and the statistical problem:

- The family history study of dementia:
  - 147 participants
  - cognitively normal or mildly impaired
  - maternal onset of dementia: ascertained using parental history questionnaire, 70% censoring
- Standard linear regression analysis:
  - Y: beta-amyloid deposition
  - X: maternal age of onset of dementia
  - controlling for $\mathbf{Z}$: age of offspring, education, gender

  Problem: *random right censoring* of age of onset means that X is not observed for every subject

Starting point: available methods

- Most of the literature on censored covariates addresses limit of detection (type I censoring)
  - **Parametric**: MLE (May et al, 2011); multiple imputation (Lynn, 2001)
  - **Nonparametric**: imputation (Schisterman et al, 2006); multiple imputation (Wang and Feng, 2012)
- Methods for random censoring are lacking; recent developments using multiple imputation (Atem et al, 2016).
- Use of censored covariate, without adjustment, leads to *bias* and *inflated* type I error (Austin & Brunner 2003).
- Complete-case analysis:
  - simplest approach and most commonly done
  - omits individuals with censored covariate
  - *valid* under some assumptions, but typically *inefficient* with moderate or heavy censoring

The Model

- Consider the linear regression model,

$$Y = \alpha_0 + \alpha_1 X + \boldsymbol{\alpha}_2^T \mathbf{Z} + \epsilon, \tag{1}$$

  where X (the covariate of interest) is right censored by C, and $\mathbf{Z}$ is a completely observed $p \times 1$ covariate vector

- Observable: Y, Z, $U = \min(X, C)$ and $\delta = I(X < C)$

- Model assumptions:
  - $(X, C, \mathbf{Z}^T)^T \perp \epsilon$
  - ϵ has mean 0 and finite variance σ^2

- Our primary scientific interest: α_1, which captures the association between Y and X

- We would like to test $H_0 : \alpha_1 = 0$ and obtain a consistent estimator of α_1

- Two threshold regression approaches:
  - deletion threshold regression
  - complete threshold regression

Deletion Threshold Regression

- Using a threshold t^*, define

$$X^* = \begin{cases} 1, & \text{if } X > t^*, C > t^*; \\ 0, & \text{if } X \leq t^*, X < C \end{cases}$$

and delete non-informative observations that have $C \leq t^*, X > C$

- The linear regression model implies that

$$E(Y|X^* = 0, \mathbf{Z}) = \alpha_0 + \alpha_1 E(X|\mathbf{Z}, X^* = 0) + \boldsymbol{\alpha}_2^T \mathbf{Z},$$

$$E(Y|X^* = 1, \mathbf{Z}) = \alpha_0 + \alpha_1 E(X|\mathbf{Z}, X^* = 1) + \boldsymbol{\alpha}_2^T \mathbf{Z}.$$

These equations justify fitting the following model conditional on the X^*'s:

$$E[Y|X^*, \mathbf{Z}] = \beta_0(t^*, \mathbf{Z}) + \beta_1(t^*, \mathbf{Z})X^* + \boldsymbol{\beta}_2^T \mathbf{Z} \qquad (2)$$

Hypothesis Test $H_0 : \alpha_1 = 0$

- It follows that

$$\beta_1(t^*, \mathbf{Z}) = \alpha_1 \{E(X|\mathbf{Z}, X^* = 1) - E(X|\mathbf{Z}, X^* = 0)\}.$$

- Under independence of X and $\mathbf{Z}$ given X^*,

$$\beta_1(t^*) = \alpha_1\{E(X|X^* = 1) - E(X|X^* = 0)\} \equiv \alpha_1\mu(t^*) \quad (3)$$

  Since $\mu(t^*) > 0$, it follows that

  A test of $H_0' : \beta_1(t^*) = 0$ is a valid test of $H_0 : \alpha_1 = 0$

- Remarks:
  - the test is valid even if C is dependent on X
  - the choice of t^* impacts the power of the hypothesis test

Consistent estimation of α_1: $(X^* \perp \mathbf{Z})$

- First, obtain consistent estimator of $\beta_1(t^*)$ fitting model (2).
- Then, estimate the bias-correction term $\mu(t^*)$ in equation (3).
- The conditional mean $E(X|X^* = 0)$ can be estimated empirically by

$$\sum_{i=1}^{n} \delta_i U_i I(U_i < t^*) / \sum_{i=1}^{n} \delta_i I(U_i < t^*)$$

- Since $\int_t^\infty S_X(u)du = \int_t^\infty u f_X(u)du - tS_X(t)$,

$$\begin{aligned} E(X|X^* = 1) &= E(X|X > t^*) \\ &= \frac{\int_{t^*}^\infty u f_X(u)du}{S_X(t^*)} = \frac{\int_{t^*}^\infty S_X(u)du}{S_X(t^*)} + t^* \quad (4) \end{aligned}$$

Consistent estimation of α_1: $(X \perp C)$

- $\hat{S}_X(x)$: Kaplan-Meier estimator of $S_X(x)$. $\int_{t^*}^{\infty} S_X(u)du$ can be estimated by $\int_{t^*}^{\infty} \hat{S}_X(u)du$, which can be approximated using the trapezoidal rule.

- An estimator for α_1 is thus given by $\hat{\alpha}_1 = \hat{\beta}_1(t^*)/\hat{\mu}(t^*)$, where

$$\hat{\mu}(t^*) = \frac{1}{2\,\hat{S}(t^*)} \left(\sum_{j=1}^{k} I\left\{X_{(j)} > t^*\right\} \left[X_{(j)} \left\{X_{(j-1)} \vee t^*\right\}\right] \right.$$

$$\left. - \left[\hat{S}\{X_{(j-1)} \vee t^*\} + \hat{S}\{X_{(j)} \vee t^*\}\right] + t^* \right)$$

$$- \frac{\sum_{i=1}^{n} \delta_i U_i I(U_i \leq t^*)}{\sum_{i=1}^{n} \delta_i I(U_i \leq t^*)},$$

$X_{(1)} < X_{(2)} < \ldots < X_{(k)}$ are the observed, uncensored failure times in the sample, and $X_{(0)} = 0$.

Estimating the tail of $S_X(x)$

- **Problem**: If the largest observations are censored, the tail of S_X cannot be estimated, but is required for accurate estimation of $E(X)$ and $E(X|X > t^*)$.
- **Strategy 1**: treat the largest observation of X as an observed failure even if it is censored (Efron, 1967)

  – underestimates $E(X|X > t^*)$ if X has much heavier tail than C.
- **Strategy 2**: approximate the tail of $S_X(x)$ using a parametric function (Gong & Fang, 2012)

  – parametric assumptions may not hold
- **Strategy 3**: increase the observed time to the upper limit of the support of X and consider it to be an event

  – requires knowledge of the upper limit of the support of X

Asymptotic Properties

- Strong consistency: $\hat{\alpha}_1 \to \alpha_1$ almost surely, as $n \to \infty$;
- Asymptotic normality: $n^{1/2}(\hat{\alpha}_1 - \alpha_1) \to N(0, \Sigma)$.
- Prove using the empirical processes theory.
  - show that $\hat{\alpha}_1$ is a plug-in estimator in a map from the distribution of $\{Y, \mathbf{Z}, U, \Delta\}$ to α_1, and the mapping is compactly differentiable.
  - Glivenko-Cantelli theorem plus continuous mapping theorem $\longrightarrow$ strong consistency.
  - Donsker theorem plus functional delta method $\longrightarrow$ asymptotic normality.

Selection of threshold, t^*

- t^* impacts the power of the hypothesis test $H_0' : \beta_1(t^*) = 0$.
- The test of $H_0' : \beta_1(t^*) = 0$ is essentially a two-sample test comparing the means of two normal distributions with equal variances, with power function

$$\Phi\left(-z_{1-\alpha/2} + \frac{|\mu_1(t^*) - \mu_2(t^*)|}{\sigma\sqrt{1/n_1(t^*) + 1/n_2(t^*)}}\right)$$

$$= \Phi\left(-z_{1-\alpha/2} + \frac{|\alpha_1\mu(t^*)|}{\sigma\sqrt{1/n_1(t^*) + 1/n_2(t^*)}}\right),$$

  where $\mu_1(t^*) = E(Y|X^* = 1)$, $\mu_2(t^*) = E(Y|X^* = 0)$.
- Select t^* to maximize $\psi_1(t^*) = |\mu(t^*)|/\sqrt{1/n_1(t^*) + 1/n_2(t^*)}$.
- Does not require a correction for maximal selection since $\mu(t^*)$ is unrelated to the association between X and Y and depends only on the distributions of X and C.

Consistent estimation of α_1: $(X^* \not\perp \mathbf{Z})$

- When $\mathbf{Z}$ is categorical with K categories, $\mathbf{Z}^{(1)}$, ..., $\mathbf{Z}^{(K)}$, stratify estimation on values of $\mathbf{Z}$ and estimate α_1 as

$$\hat{\alpha}_1 = \frac{1}{K} \sum_{k=1}^{K} \frac{\hat{\beta}_1(t^*, \mathbf{Z}^{(k)})}{\hat{E}(X|\mathbf{Z}^{(k)}, X^* = 1) - \hat{E}(X|\mathbf{Z}^{(k)}, X^* = 0)}.$$

- When $\mathbf{Z}$ is continuous
  - discretize it into K distinct categories and stratify
  - fit a Cox model for X given $\mathbf{Z}$ to calculate $E(X|X^*, \mathbf{Z})$ and similarly average to estimate α_1

Complete Threshold Regression

- Alternative complete threshold regression method that retains all observations.
  - gains efficiency through use of all observations, especially when adjusting for censored covariate X.
  - sacrifices efficiency due to potential misclassification of indeterminate observations.
- Derive binary covariate that indicates whether $U = \min(X, C) \leq t^*$ or $U > t^*$.
- U is completely observed: no indeterminate observations and thus no deletions.
- Fit the derived model, conditional on the thresholded U's:

$$E(Y|I(U > t^*), \mathbf{Z}) = \gamma_0(t^*, \mathbf{Z}) + \gamma_1(t^*, \mathbf{Z})I(U > t^*) + \boldsymbol{\gamma}_2^T \mathbf{Z}.$$

Complete Threshold Regression ...

- Under independence of X and $\mathbf{Z}$ given X^*,

$$\gamma_1(t^*) = \alpha_1 \{E(X|U > t^*) - E(X|U \leq t^*)\} = \alpha_1 \nu(t^*), \quad (5)$$

and

$$\nu(t^*) = \left[\int_0^\infty S_X(u)du - \left\{ \frac{\int_{t^*}^\infty S_X(u)du}{S_X(t^*)} + t^* \right\} \right] \Big/ \Pr(U \leq t^*),$$

where the integrals are estimated using tail approximations for S_X

- test of $H_0' : \gamma_1(t^*) = 0$ is a valid test of $H_0 : \alpha_1 = 0$.

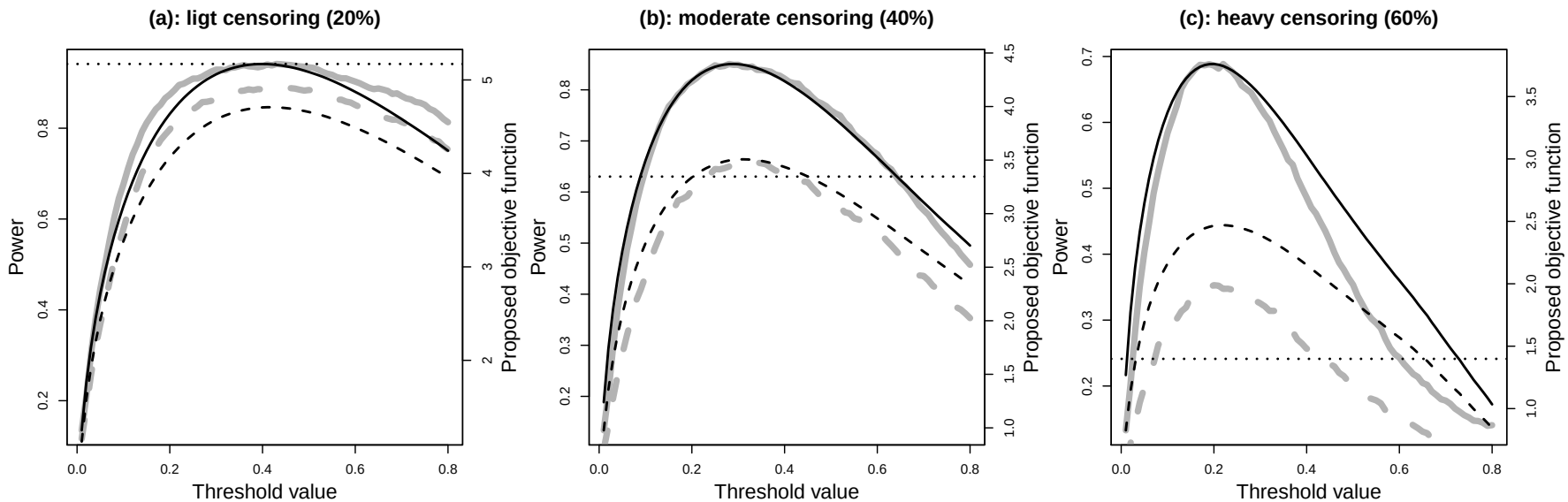
Reverse Survival Regression

- An alternative approach to testing the association between Y and X, adjusting for $\mathbf{Z}$.
- We use the Cox proportional hazards model with outcome X and covariates Y and $\mathbf{Z}$, i.e., $h(x|y, \mathbf{z}) = h_0(x) \exp(\tilde{\alpha}_1 y + \tilde{\boldsymbol{\alpha}}_2 \mathbf{z})$.
- We show that the test of $H_0 : \tilde{\alpha}_1 = 0$ based on the Cox model that reverses the natural roles of Y and X yields a valid test for $H_0 : \alpha_1 = 0$.
- However, it does not yield an estimator for α_1.

Simulation Set-up

- Simulate from model (1) with $\alpha_0 = 0.5$, $\alpha_1 = 0.5$, $\alpha_2 = -0.5$.
- Generate X from an exponential distribution $Exp(3)$, Z from an uniform distribution $Unif(1, 6)$, ϵ_i from a normal distribution $N(0, 0.75^2)$.
- Generate C from an exponential distribution: light, moderate or heavy censoring with censoring rate of 20%, 40%, or 60%.
- Sample sizes of 200 and 500; 1000 replications.
- Estimate type I error (setting $\alpha_1 = 0$ in our data generation model) and power (setting $\alpha_1 = 0.5$). Compare to Wald tests based on $\hat{\beta}_1(t^*)$.

Selection of threshold to optimize power



Simulation Results: light censoring

method	τ_x	t^*	bias $\hat{\beta}_1(\hat{\gamma}_1)$	bias $\hat{\alpha}_1$	SD $\hat{\alpha}_1$	SE $\hat{\alpha}_1$	CP(%) $\hat{\alpha}_1$	SD $\hat{\alpha}_2$	% del	% $\leq t^*$	% $> t^*$
light censoring rate of 20%											
cc				0.0031	0.227	0.228	95.2	0.0423	20.0		
m1	1.75	0.40	-0.208	0.0158	0.238	0.236	94.6	0.0412	15.5	62.1	22.4
m1	1.50	0.40	-0.208	0.0198	0.240	0.239	94.7	0.0412	15.5	62.1	22.4
m1	2.00	0.40	-0.208	0.0109	0.236	0.234	94.4	0.0412	15.5	62.1	22.4
m1	obs	0.40	-0.208	0.0205	0.240	0.241	94.7	0.0412	15.5	61.1	22.4
m2	1.75	0.42	-0.236	0.0112	0.260	0.260	94.6	0.0381	0.0	79.2	20.8
m2	1.50	0.42	-0.236	0.0153	0.262	0.263	94.5	0.0381	0.0	79.2	20.8
m2	2.00	0.42	-0.236	0.0061	0.257	0.257	94.5	0.0381	0.0	79.2	20.8
m2	obs	0.42	-0.236	0.0161	0.262	0.265	94.6	0.0381	0.0	79.2	20.8

τ_x: the guess of the upper support of X;
t^*: threshold value;
cc: complete-case regression;
m1: deletion threshold regression;
m2: complete threshold regression;
obs: treating the largest observation of X as an observed failure.

Simulation Results: moderate censoring

method	τ_x	t^*	bias $\hat{\beta}_1(\hat{\gamma}_1)$	bias $\hat{\alpha}_1$	SD $\hat{\alpha}_1$	SE $\hat{\alpha}_1$	CP(%) $\hat{\alpha}_1$	SD $\hat{\alpha}_2$	% del	% $\leq t^*$	% $> t^*$
				moderate censoring rate of 40%							
cc				-0.0092	0.340	0.355	95.1	0.0484	40.0		
m1	1.75	0.29	-0.244	0.0106	0.276	0.277	96.2	0.0453	30.6	45.9	23.4
m1	1.50	0.29	-0.244	0.0246	0.283	0.284	96.1	0.0455	30.6	45.9	23.4
m1	2.00	0.29	-0.244	-0.0029	0.268	0.269	96.2	0.0455	30.6	45.9	23.4
m1	obs	0.29	-0.244	0.0409	0.293	0.298	96.1	0.0453	30.6	45.9	23.4
m2	1.75	0.31	-0.301	0.0123	0.346	0.347	95.7	0.0385	0.0	78.8	21.2
m2	1.50	0.31	-0.301	0.0273	0.356	0.357	95.8	0.0385	0.0	78.8	21.2
m2	2.00	0.31	-0.301	-0.0022	0.336	0.337	95.9	0.0385	0.0	78.8	21.2
m2	obs	0.31	-0.301	0.0451	0.369	0.376	95.8	0.0385	0.0	78.8	21.2

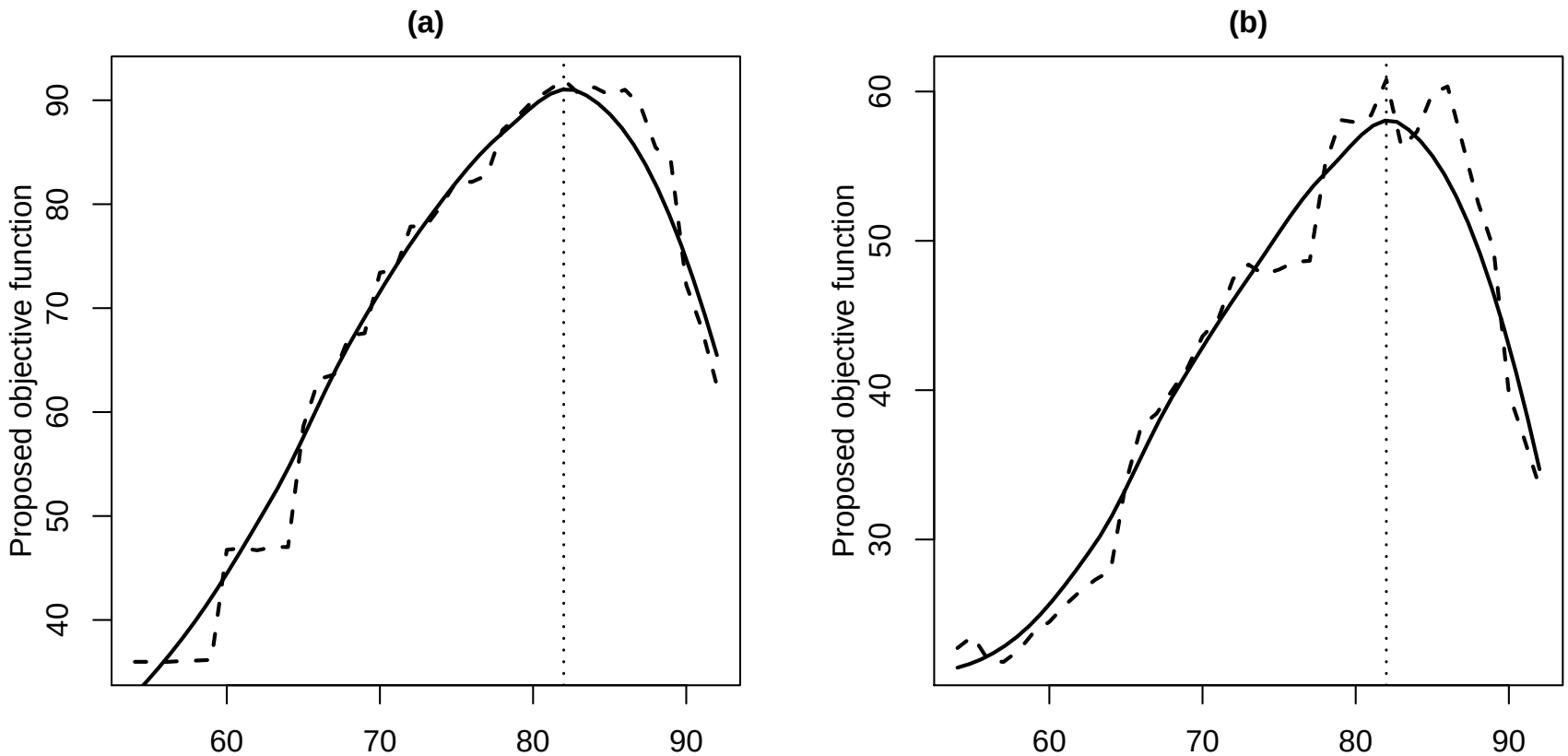
Simulation Results: heavy censoring

method	τ_x	t^*	bias $\hat{\beta}_1(\hat{\gamma}_1)$	bias $\hat{\alpha}_1$	SD $\hat{\alpha}_1$	SE $\hat{\alpha}_1$	CP(%) $\hat{\alpha}_1$	SD $\hat{\alpha}_2$	% del	% $\leq t^*$	% $> t^*$
				heavy censoring rate of 60%							
cc				-0.0046	0.668	0.650	95.1	0.0606	59.7		
m1	1.75	0.20	-0.271	-0.0221	0.316	0.325	95.9	0.0527	46.3	31.2	22.5
m1	1.50	0.20	-0.271	0.0075	0.334	0.344	96.2	0.0527	46.3	31.2	22.5
m1	2.00	0.20	-0.271	-0.0480	0.301	0.310	95.3	0.0527	46.3	31.2	22.5
m1	obs	0.20	-0.271	0.1033	0.396	0.419	96.5	0.0527	46.3	31.2	22.5
m2	1.75	0.22	-0.364	-0.0305	0.482	0.499	96.3	0.0387	0.0	80.7	19.3
m2	1.50	0.22	-0.364	0.0004	0.512	0.529	96.1	0.0387	0.0	80.7	19.3
m2	2.00	0.22	-0.364	-0.0574	0.457	0.473	96.3	0.0387	0.0	80.7	19.3
m2	obs	0.22	-0.374	0.1049	0.621	0.656	96.4	0.0387	0.0	80.7	19.3

Simulations: power and type I error

n		m1		m2		m1		m2	
		light censoring				heavy censoring			
hypothesis test based on:		β_1	α_1	γ_1	α_1	β_1	α_1	γ_1	α_1
		Power							
200	*complete-case*	60.4%				13.0%			
	optimal threshold	57.5%	58.4%	50.9%	51.7%	32.7%	30.4%	14.9%	13.8%
500	*complete-case*	94.6%				27.5%			
	optimal threshold	93.0%	93.2%	87.8%	88.0%	69.8%	69.2%	35.5%	35.9%
		Type I error							
200	*complete-case*	5.14%				5.74%			
	optimal threshold	5.42%	5.44%	5.32%	5.10%	5.48%	4.22%	5.34%	4.06%
500	*complete-case*	5.36%				4.70%			
	optimal threshold	5.26%	5.26%	5.52%	5.62%	4.92%	4.72%	5.56%	5.06%

Alzheimer's study: identification of optimal threshold for maternal age of dementia onset



Alzheimer's Study Results

method	τ_x	$\hat{\alpha}_1$ (SE)	p-value on $\hat{\alpha}_1$	$\hat{\beta}_1$ or $\hat{\gamma}_1$ (SE)	p-value on $\hat{\beta}_1$ or $\hat{\gamma}_1$	%del
the variable of interest: Maternal age of demential onset (in years)						
cc		-0.00981 (0.00498)	0.057	–	–	70.21%
m1	105	-0.00379 (0.00245)	0.122	-0.0775 (0.0499)	0.121	34.75%
m1	obs (100)	-0.00367 (0.00245)	0.135	-0.0775 (0.0499)	0.121	34.75%
m2	105	-0.00309 (0.00323)	0.339	-0.0306 (0.0347)	0.378	0%
m2	obs(100)	-0.00304 (0.00335)	0.365	-0.0306 (0.0347)	0.378	0%
rs			0.001			0%

Conclusions

- Threshold regression is simple and avoids extensive modeling.
- It allows for estimation of the regression coefficient of censored covariate, as well as efficient hypothesis testing of censored covariate effect.
- The optimal threshold can be easily identified through an objective function.
- Deletion threshold regression versus complete threshold regression: comparable type I error; higher power (especially under heavy censoring).
- Deletion threshold regression versus complete case analysis: higher power under moderate or heavy censoring.

Conclusions

- Assumes censoring mechanism is independent of X.
- Extend to the case of multiple censored covariates.
- Extend to generalized linear regression model.

Acknowledgments

- Deborah Blacker, M.D.
- Bradley Hyman, M.D., Ph.D.
- Keith Johnson, M.D.
- Eric Macklin, Ph.D.
- Jacqueline Maye
- Jing Qian, Ph.D.

Alternative Approach: Multiple Imputation

Under linear model (1), we develop a proper multiple imputation approach that also does not impose distributional assumptions on the X, but rather uses a Cox model for the distribution of X given other covariates in the model.

1. Sample with replacement from the original data.
2. Fit model (1) using the uncensored observations to sample from the distribution of the coefficients $(\alpha_0, \alpha_1, \alpha_2)$ and obtain estimates $(\hat{\alpha}_0^c, \hat{\alpha}_1^c, \hat{\alpha}_2^c)$.
3. Fit a model to the sampled data for X given Z to estimate β and $f_\beta(x|z)$, the model based estimate of the density of X given Z, with corresponding survivor function, $S_\beta(x|z)$.
4. Generate X from its predictive distribution, $P(X = x \,|\, C = c, X > c, Y = y, Z = z)$.

5. Fit a linear regression model of Y on the completed data (X, Z) and estimate the parameters of the model, $(\hat{\alpha}_0^m, \hat{\alpha}_1^m, \hat{\alpha}_2^m)$, where the superscript m labels the estimates from the mth imputation.
6. Repeat Steps 1-5 M times.
7. Obtain multiple imputation estimates and variances.

We have extended this multiple imputation method to logistic regression analysis (Atem et al, 2016).