



武汉大学
Wuhan University

病例队列设计及其应用

Case-Cohort Design and Its Application

项目背景

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

- ◆ 生存分析在许多学科应用广泛；
- ◆ 生存数据删失率高；
- ◆ 有效的有偏随机抽样机制。

病例队列设计

项目背景

病例队列设计

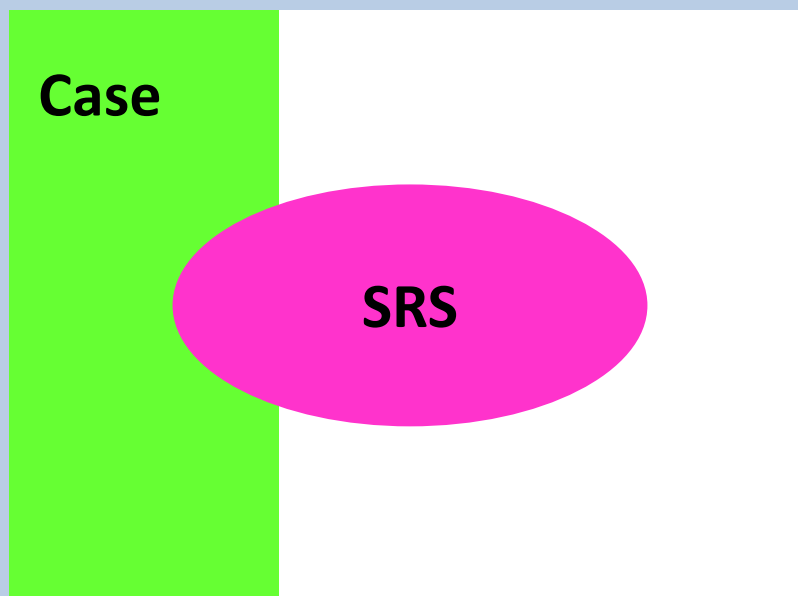
生存分析

统计模型

统计模拟

实例分析

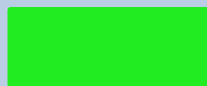
参考文献



所有研究对象（全队列）



简单随机样本（子队列）



所有的病例



病例队列样本

右删失数据

项目背景

病例队列设计

生存分析

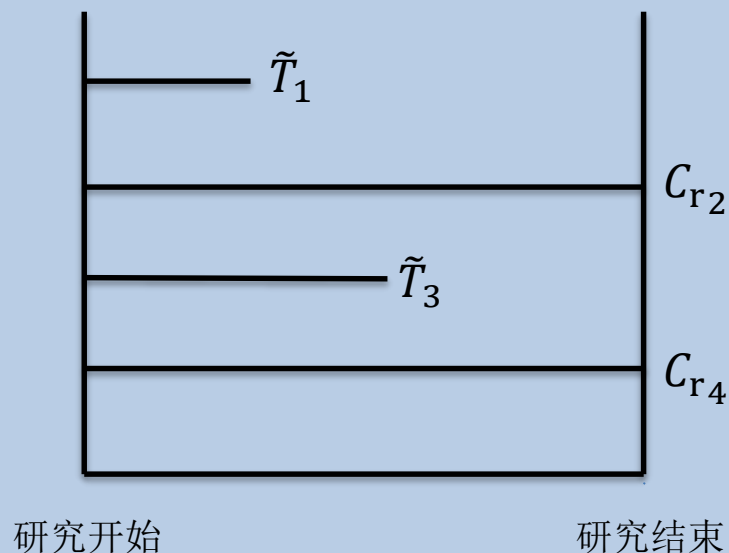
统计模型

统计模拟

实例分析

参考文献

数据表示: (T, δ)



- ◆ $T = \min(\tilde{T}, C_r)$

- ◆ $\delta = I(\tilde{T} \leq C_r)$

- ◆ \tilde{T} : 生存时间

- ◆ C_r : 删失时间

基本函数

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

- ◆ 生存函数

$$S(t) = P(\tilde{T} > t)$$

- ◆ 风险率函数(瞬时死亡率)

$$\lambda(t) = \lim_{\Delta_t \rightarrow 0} \frac{P(t \leq \tilde{T} < t + \Delta_t | \tilde{T} \geq t)}{\Delta_t} = \frac{f(t)}{S(t)}$$

常用模型

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

$$\lambda(t|Z) = \lambda_0(t) g(\boldsymbol{\beta}'\mathbf{Z})$$

- ◆ $\lambda_0(t)$: 未知的固定的基准风险率函数
- ◆ $g(x)$: 已知的非负函数, 常取 e^x
- ◆ $\mathbf{Z} = (Z_1, \dots, Z_p)'$: 协变量
- ◆ $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$: 协变量参数

Cox比例风险模型

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\beta}' \mathbf{Z}_i)}{\sum_{j \in R_i} \exp(\boldsymbol{\beta}' \mathbf{Z}_j)} \right\}^{\delta_i}$$

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \left\{ \boldsymbol{\beta}' \mathbf{Z}_i - \ln \left[\sum_{j \in R_i} \exp(\boldsymbol{\beta}' \mathbf{Z}_j) \right] \right\}$$

- ◆ $R_i = \{j | T_j \geq T_i\}$: T_i 时刻风险集

病例队列设计

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\beta}' \mathbf{Z}_i)}{\sum_{j \in \tilde{R}_i} \exp(\boldsymbol{\beta}' \mathbf{Z}_j)} \right\}^{\delta_i}$$

$$l(\boldsymbol{\beta}) = \sum_{i=1}^n \delta_i \left\{ \boldsymbol{\beta}' \mathbf{Z}_i - \ln \left[\sum_{j \in \tilde{R}_i} \exp(\boldsymbol{\beta}' \mathbf{Z}_j) \right] \right\}$$

- ◆ $\tilde{R}_i = \{j | T_j \geq T_i\}$: T_i 时刻风险集

模型对比

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

$$\sum_{j \in \tilde{R}_i} g(\boldsymbol{\beta}' \mathbf{Z}_j) = \sum_{j \in R_i} \xi_j g(\boldsymbol{\beta}' \mathbf{Z}_j) + \delta_i (1 - \xi_i) g(\boldsymbol{\beta}' \mathbf{Z}_j)$$

- ◆ $g(\boldsymbol{\beta}' \mathbf{Z}) = \exp(\boldsymbol{\beta}' \mathbf{Z})$;
- ◆ $\xi_i = I(\text{个体 } i \text{ 在子队列中})$;
- ◆ $\delta_i = I(\text{个体 } i \text{ 是病例})$.

渐近性质

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

当满足一定条件时，有如下定理成立：

Cox模型	$\hat{\beta}_n \xrightarrow{p} \beta_0$
	$\sqrt{n} (\hat{\beta}_n - \beta_0) \xrightarrow{L} N(0, \Sigma^{-1})$
病例队列设计	$\tilde{\beta}_n \xrightarrow{p} \beta_0$
	$\sqrt{n} (\tilde{\beta}_n - \beta_0) \xrightarrow{d} N(0, \tilde{\Sigma}^{-1} \Delta \tilde{\Sigma}^{-1})$

模拟实验设计

项目背景

$$\lambda(t) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2)$$

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

- ◆ 样本量 $n = 200,300$
- ◆ 删失率 $cc = 0.5, 0.7, 0.9$
- ◆ 模拟次数 $S = 1000$
- ◆ $\beta_0 = (\ln 2, -0.5)', (\ln 2, 0)', (0, -0.5)'$
- ◆ $Z_1 \sim B(1, 0.5), Z_2 \sim N(0, 1)$
- ◆ $\tilde{T} \sim \text{Exp}(\lambda), C_r \sim U(0, u)$
- ◆ 置信水平 $\alpha = 0.05$
- ◆ SE 的计算: 经验估计 VS Bootstrap估计

$n = 200$ 经验估计模拟结果

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

		病例队列设计				Cox模型			
β_0	cc	Bias	SD	SE	CP	Bias	SD	SE	CP
(ln2, -0.5)	0.5	0.013	0.286	0.184	0.798	0.010	0.204	0.219	0.967
		-0.016	0.159	0.099	0.772	-0.017	0.112	0.116	0.960
	0.7	-0.007	0.275	0.230	0.901	0.013	0.274	0.282	0.961
		-0.015	0.137	0.119	0.921	-0.002	0.138	0.143	0.962
	0.9	0.015	0.439	0.413	0.939	0.041	0.513	0.511	0.971
		-0.018	0.220	0.197	0.922	-0.008	0.238	0.236	0.935
(ln2, 0)	0.5	0.003	0.282	0.181	0.801	0.011	0.208	0.218	0.963
		-0.006	0.143	0.090	0.791	0.001	0.108	0.104	0.943
	0.7	0.010	0.256	0.230	0.925	0.006	0.266	0.279	0.959
		0.001	0.132	0.112	0.909	0.003	0.135	0.134	0.945
	0.9	0.023	0.471	0.458	0.962	0.045	0.581	0.574	0.966
		-0.002	0.224	0.207	0.937	-0.020	0.269	0.252	0.932

$n = 200$ Bootstrap模拟结果

项目背景
病例队列设计
生存分析
统计模型
统计模拟
实例分析
参考文献

		病例队列设计				Cox模型			
β_0	cc	Bias	SD	SE	CP	Bias	SD	SE	CP
(ln2, -0.5)	0.5	0.015	0.294	0.298	0.955	0.011	0.212	0.214	0.952
		-0.015	0.163	0.164	0.956	-0.003	0.113	0.111	0.952
	0.7	0.001	0.267	0.278	0.957	0.019	0.281	0.285	0.951
		-0.015	0.144	0.148	0.960	-0.009	0.137	0.143	0.960
	0.9	0.047	0.442	0.457	0.973	0.050	0.499	0.545	0.970
		-0.013	0.213	0.219	0.961	-0.013	0.231	0.253	0.962
(ln2, 0)	0.5	0.014	0.281	0.289	0.955	0.012	0.214	0.216	0.951
		-0.003	0.144	0.152	0.966	0.004	0.105	0.106	0.946
	0.7	0.028	0.262	0.271	0.962	0.012	0.272	0.285	0.966
		-0.001	0.137	0.137	0.960	0.002	0.134	0.138	0.957
	0.9	0.061	0.475	0.507	0.973	0.039	0.585	0.600	0.952
		-0.004	0.218	0.228	0.961	-0.001	0.266	0.271	0.957

模拟结果分析

项目背景

病例队列设计

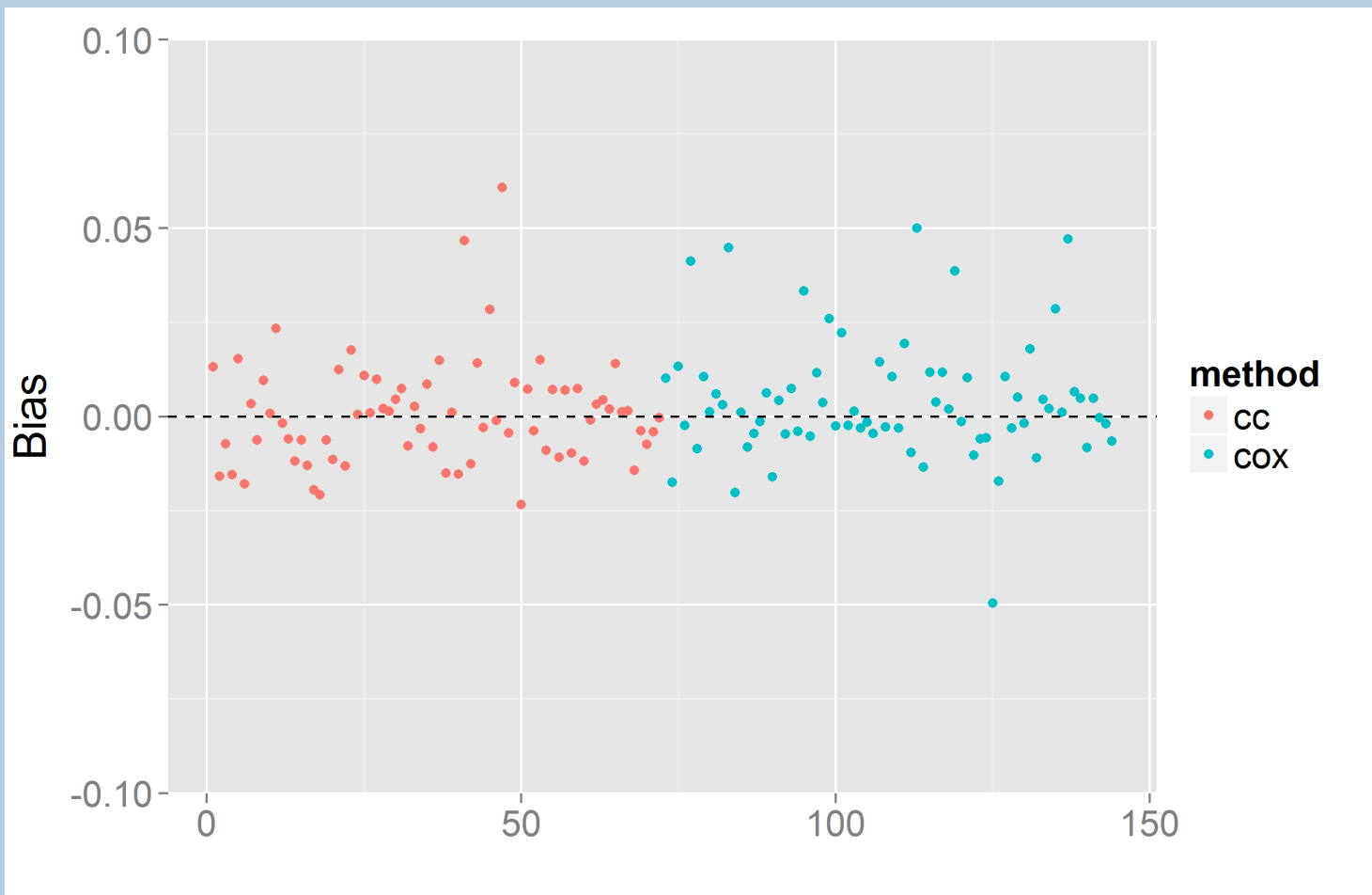
生存分析

统计模型

统计模拟

实例分析

参考文献



模拟结果分析

项目背景

病例队列设计

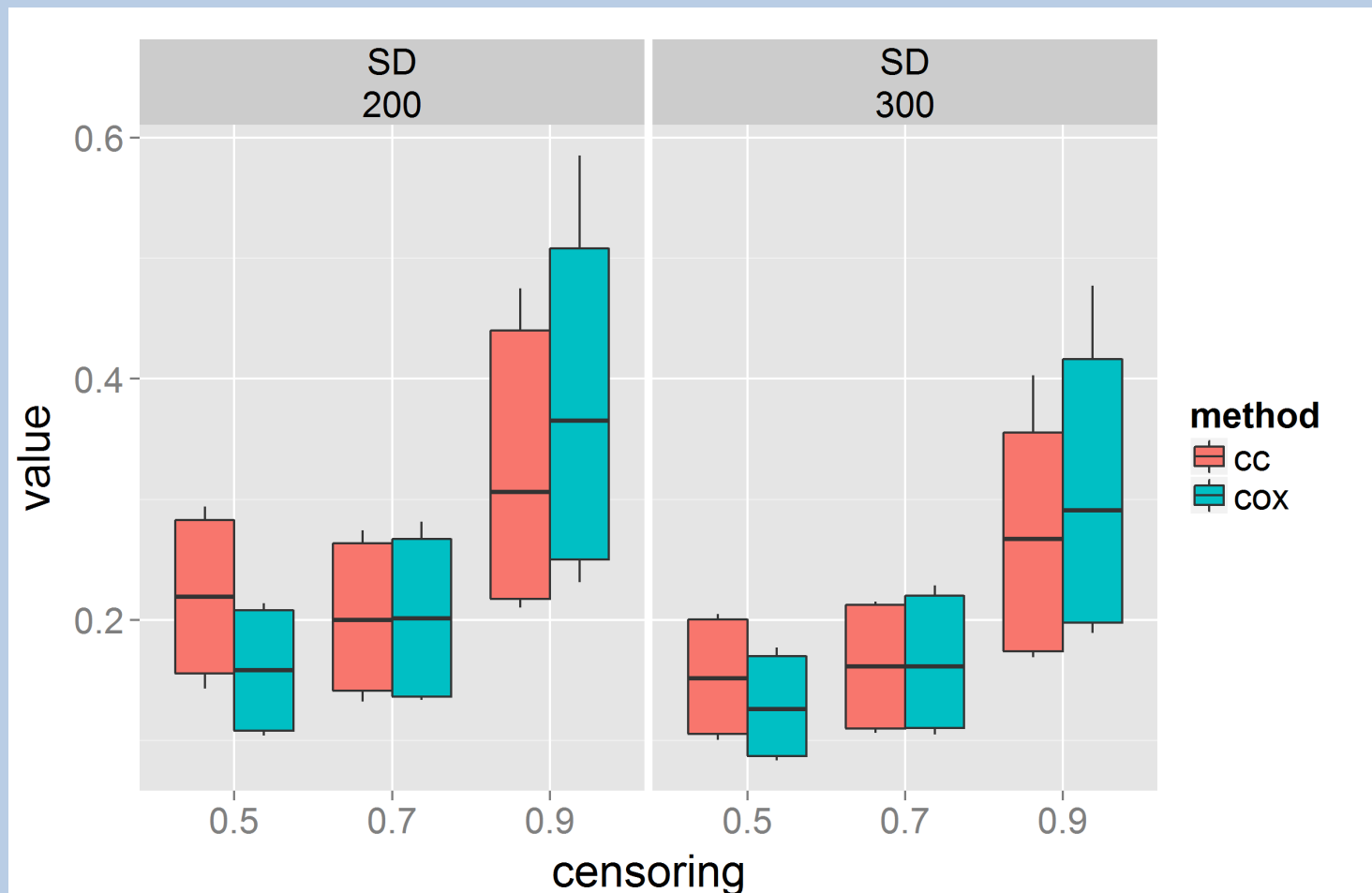
生存分析

统计模型

统计模拟

实例分析

参考文献



模拟结果分析

项目背景

病例队列设计

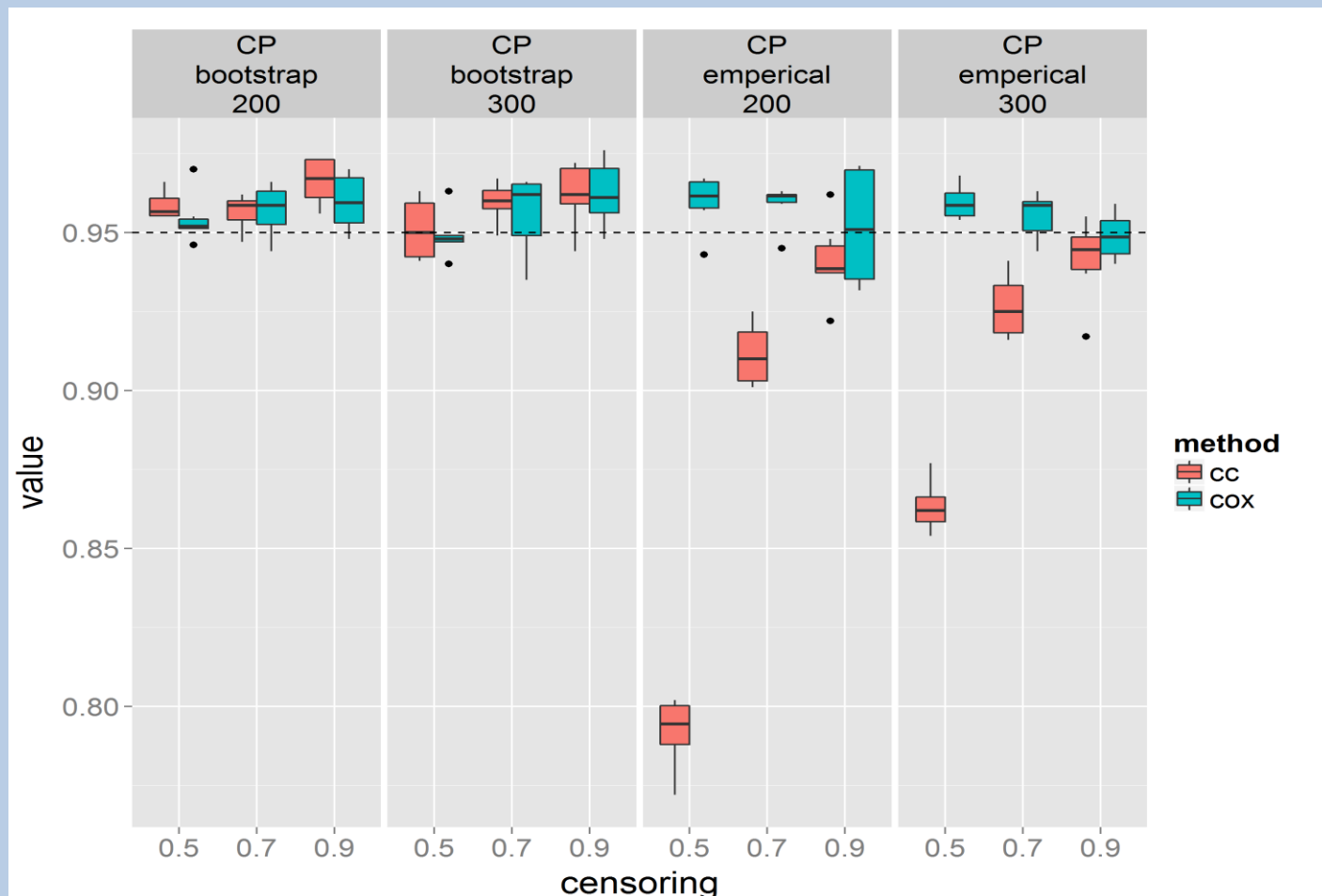
生存分析

统计模型

统计模拟

实例分析

参考文献



模拟结果分析

项目背景

病例队列设计

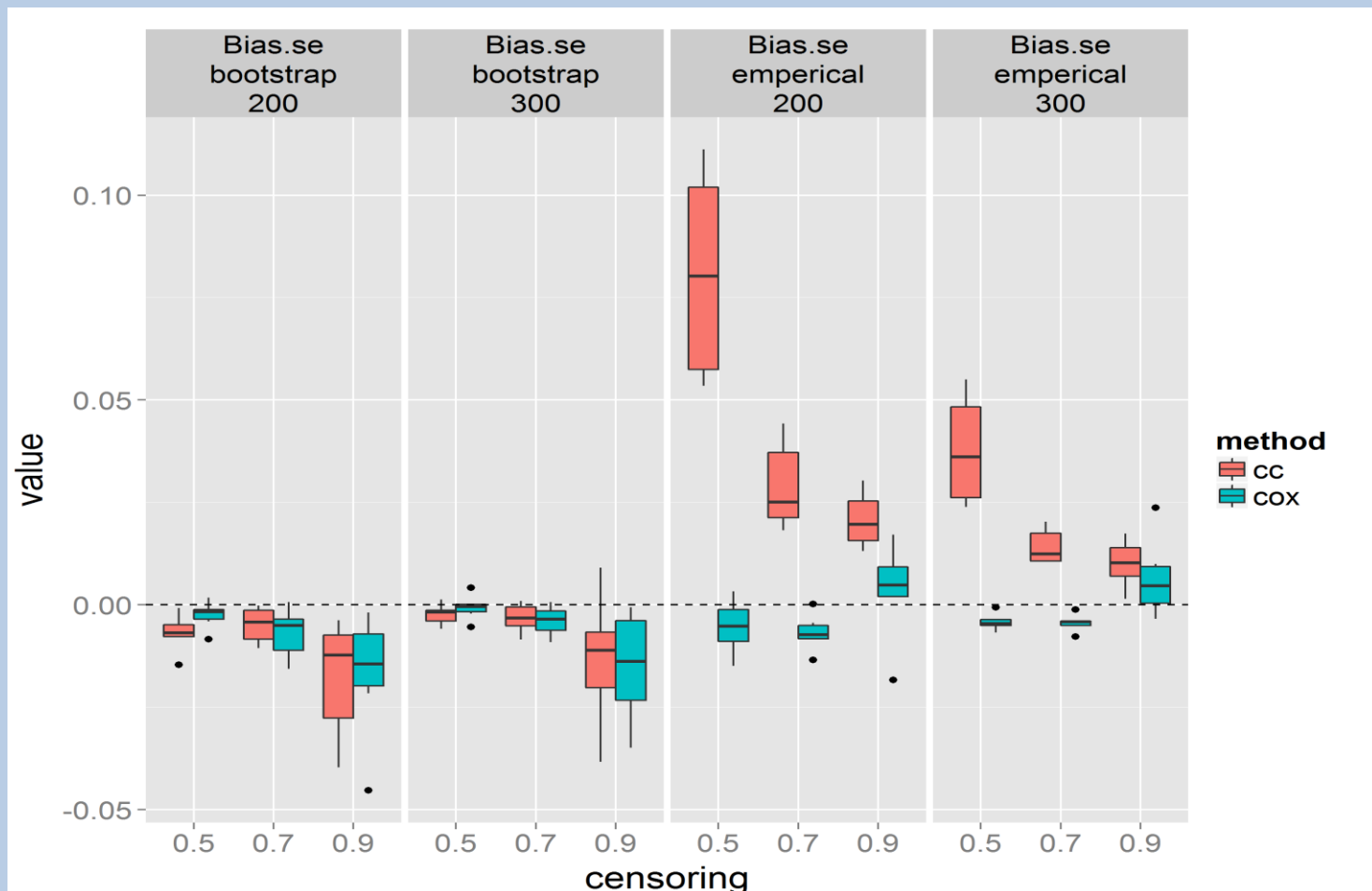
生存分析

统计模型

统计模拟

实例分析

参考文献



美国肾母细胞瘤研究数据

项目背景

病例队列设计

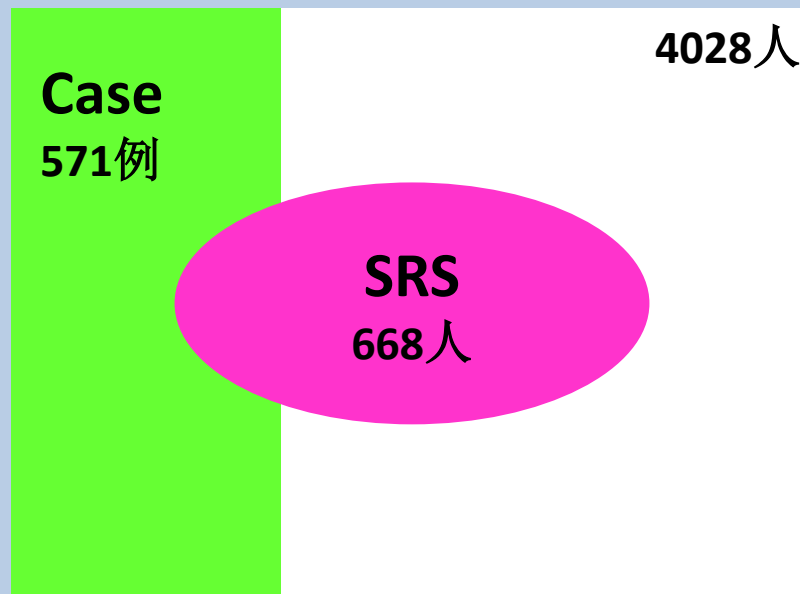
生存分析

统计模型

统计模拟

实例分析

参考文献



删失率: 0.86
发病率: 0.14

Z_1 : Histology (组织学类型) = FH (预后良好)
UH (预后不良)

Z_2 : Stage (肿瘤发展阶段) = I/II/III/IV

Z_3 : Age (年龄)

疾病复发时间分布

项目背景

病例队列设计

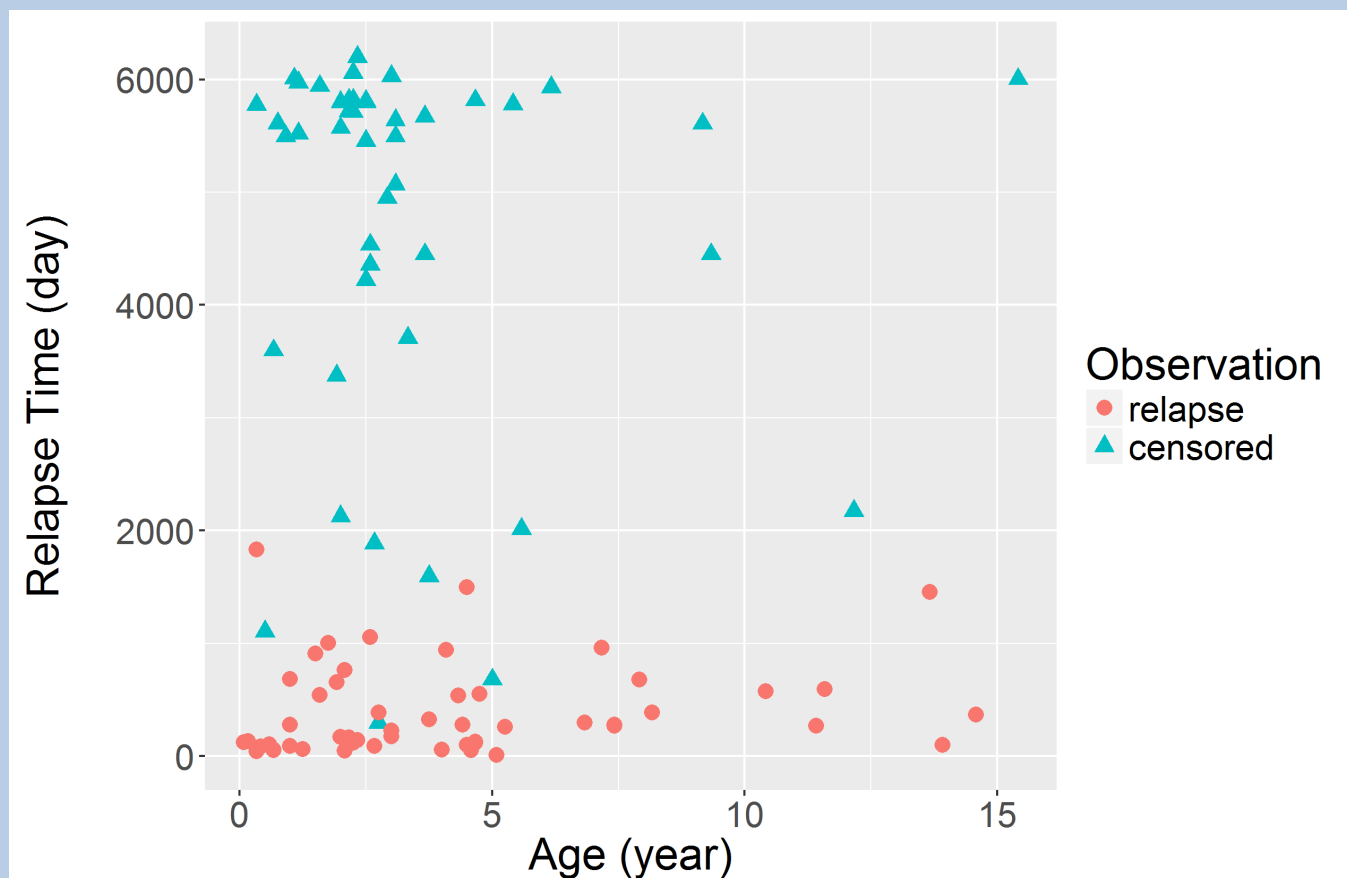
生存分析

统计模型

统计模拟

实例分析

参考文献



比例风险模型拟合

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

$$\lambda(t|Z) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_3)$$

	Parameter	Estimate	Standard Error	p value
Histology	β_1	1.429	0.153	<0.0001
Stage	β_2	0.378	0.065	<0.0001
Age	β_3	0.041	0.026	0.124

注：SE为标准差的Bootstrap估计

剔除 β_3 后模型拟合

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

$$\lambda(t|Z) = \lambda_0(t) \exp(\beta_1 Z_1 + \beta_2 Z_2)$$

	Parameter	Estimate	Standard Error	p value
Histology	β_1	1.438	0.150	<0.0001
Stage	β_2	0.401	0.062	<0.0001

注：SE为标准差的Bootstrap估计

实例分析结果

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

◆ 相对风险比

变量	相对风险	95%置信区间
UH	4.211	[3.136, 5.655]
Stage II	1.493	[1.322, 1.686]
Stage III	2.228	[1.747, 2.842]
Stage IV	3.326	[2.310, 4.790]

参考文献

项目背景

病例队列设计

生存分析

统计模型

统计模拟

实例分析

参考文献

- ◆ KLEIN & MOESCHBERGER. (2003) Survival Analysis Techniques for Censored and Truncated Data[M].
- ◆ Cox. (1972) Regression Model and Life-Tables[J].
- ◆ SELF & PRENTICE. (1988) Asymptotic Distribution Theory and Efficiency Results for Case-cohort Studies[J].
- ◆ Breslow & Chatterjee. (1998) Design and Analysis of Two-phase Studies with Binary Outcome Applied to Wilms Tumor Prognosis.