Bayesian Nonparametric Statistics for Fostering Innovation and Discovery in Biomedical Research

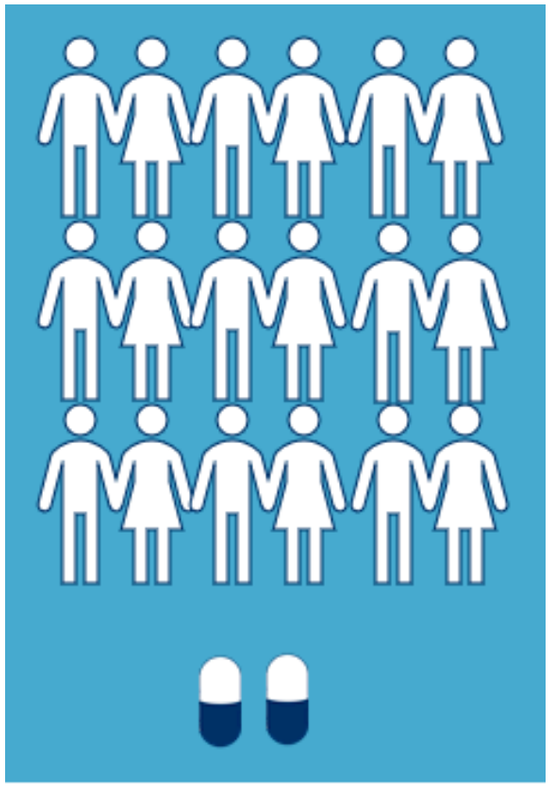
Yanxun Xu
Department of Applied Math and Statistics
Mathematics Institute for Data Science
Division of Biostatistics and Bioinformatics
The Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University

April, BNP 2022

Treatment Effect

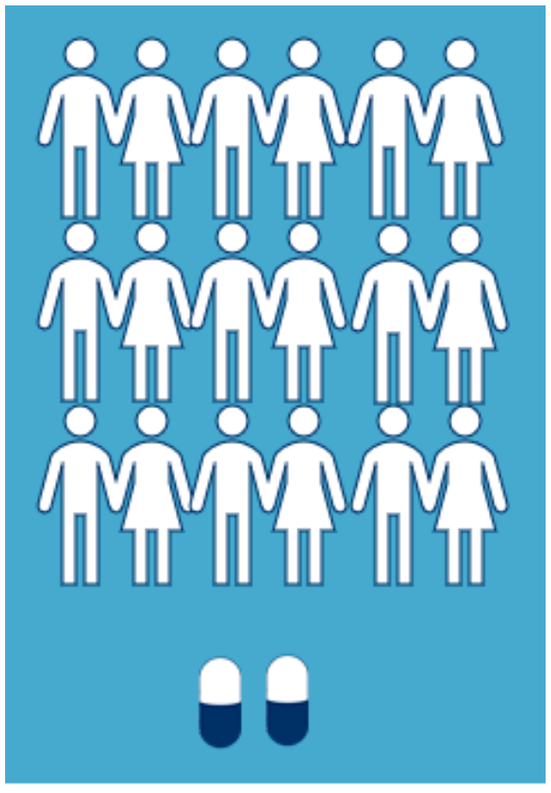
Treatment Effect

Clinical Trials



Treatment Effect

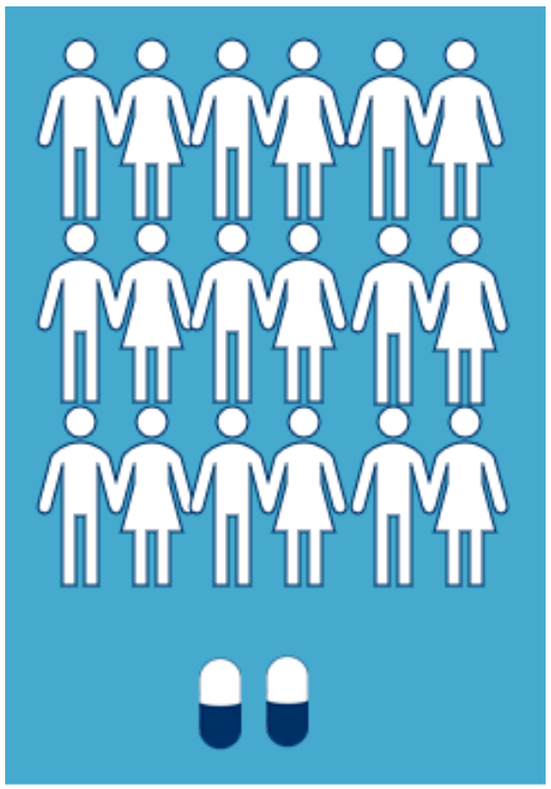
Clinical Trials



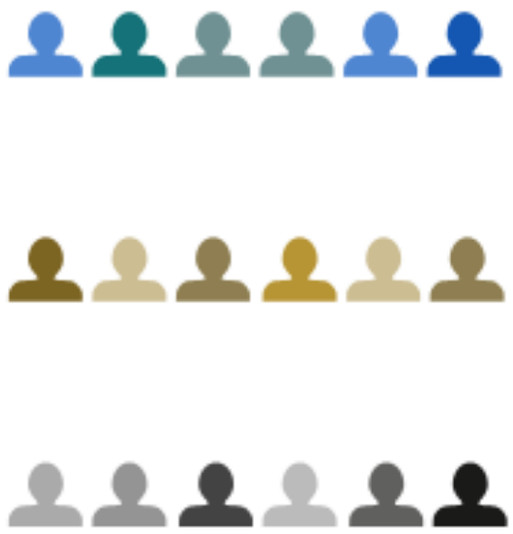
Population

Treatment Effect

Clinical Trials



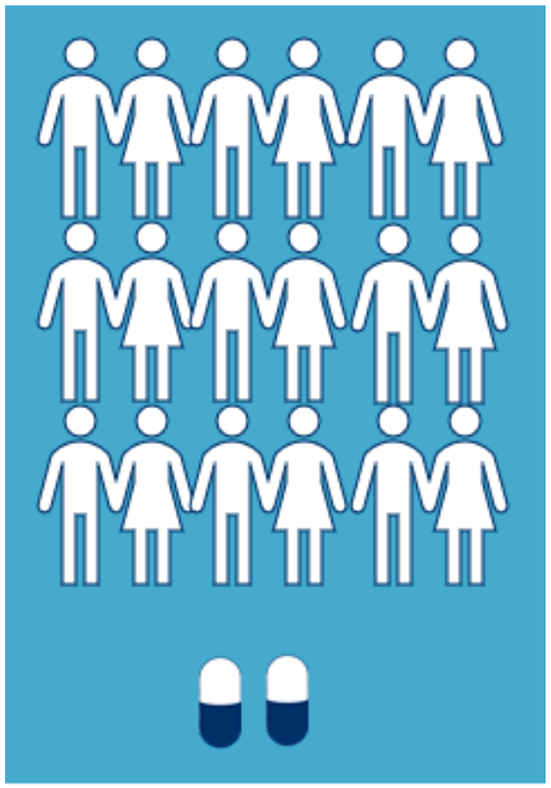
Biomarker-based Trials



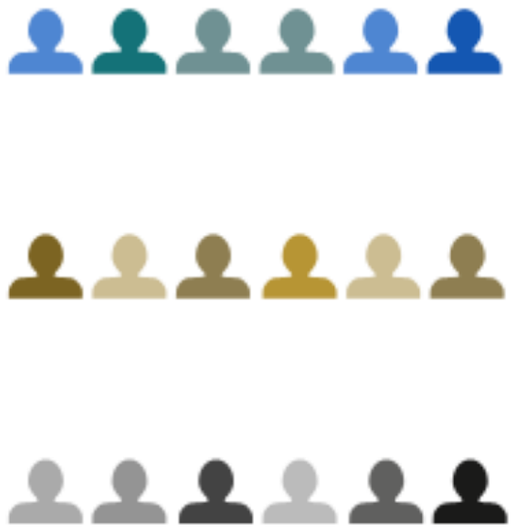
Population

Treatment Effect

Clinical Trials



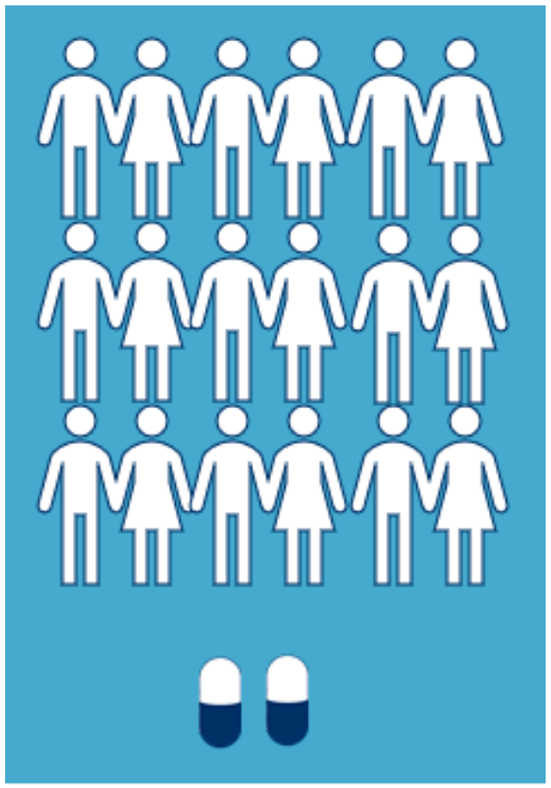
Biomarker-based Trials



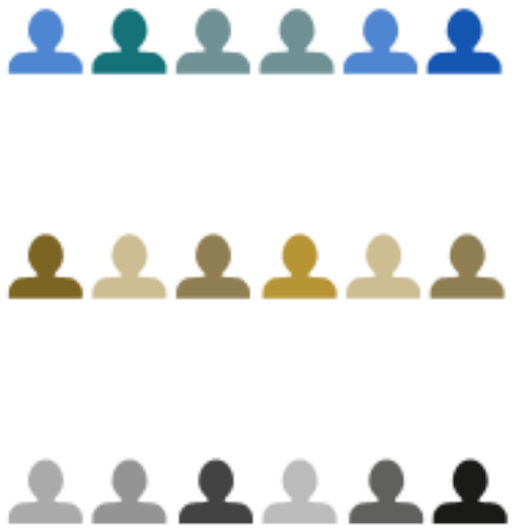
Population

Treatment Effect

Clinical Trials



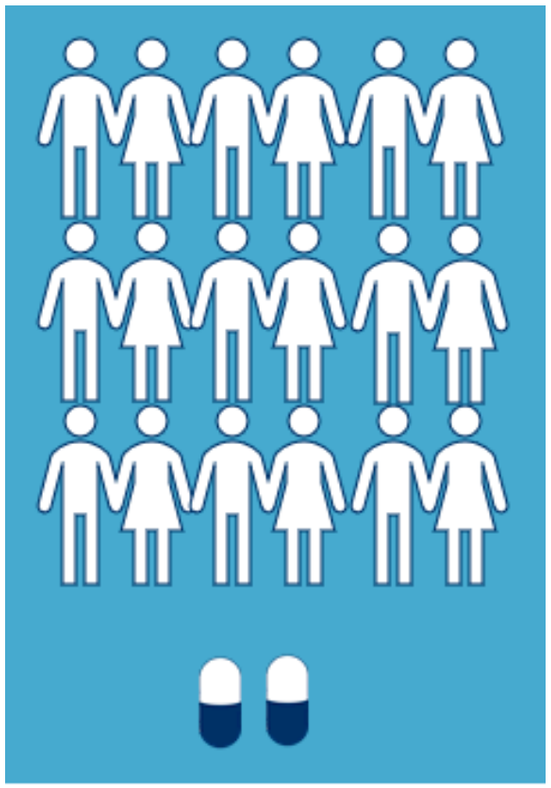
Biomarker-based Trials



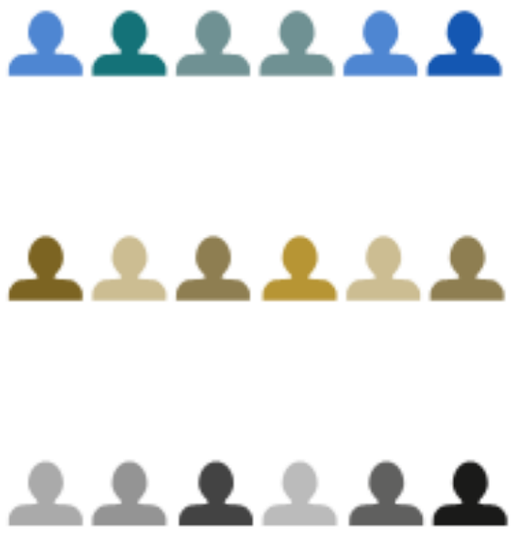
Population → Subgroup

Treatment Effect

Clinical Trials



Biomarker-based Trials

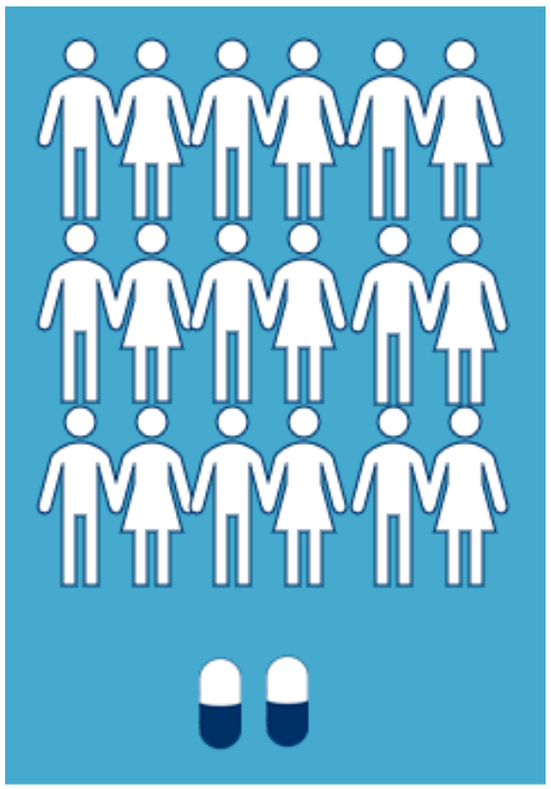


Observational Data

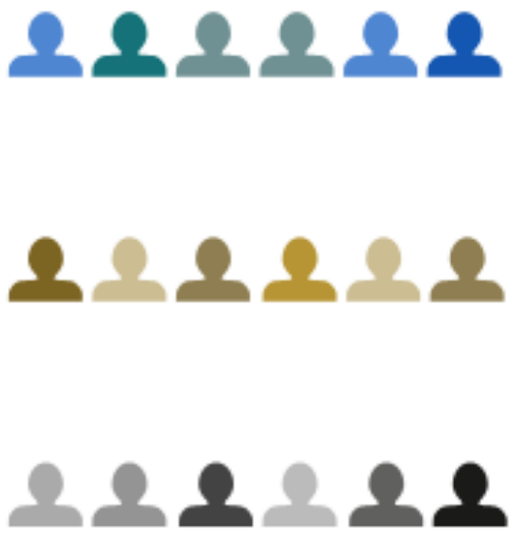
Population → Subgroup

Treatment Effect

Clinical Trials



Biomarker-based Trials

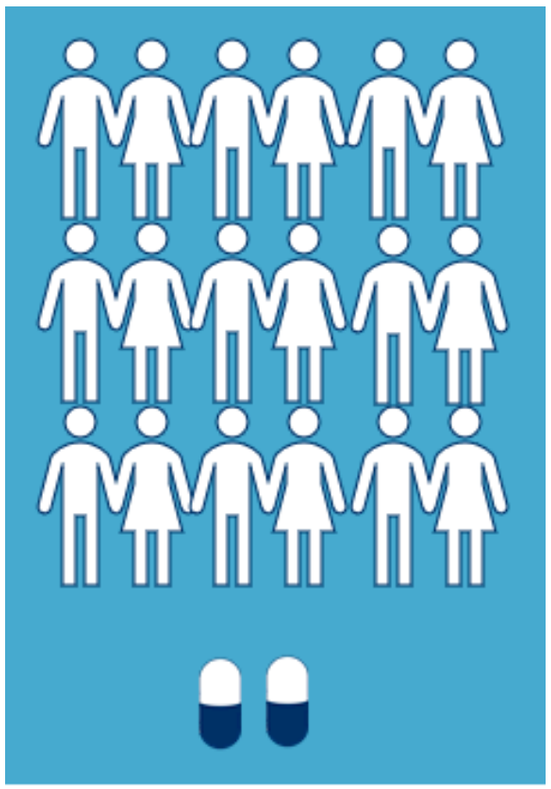


Observational Data

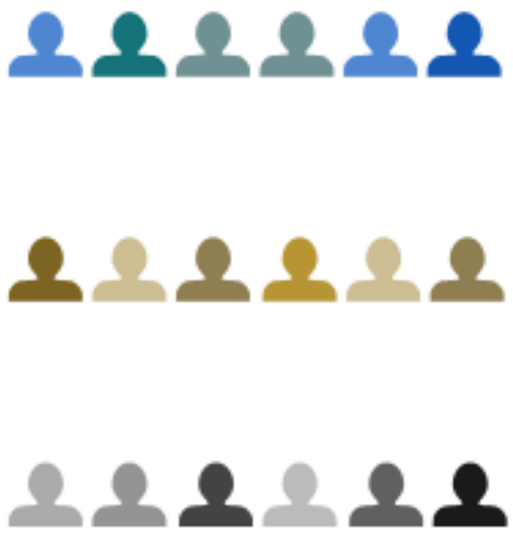
Population → Subgroup

Treatment Effect

Clinical Trials



Biomarker-based Trials

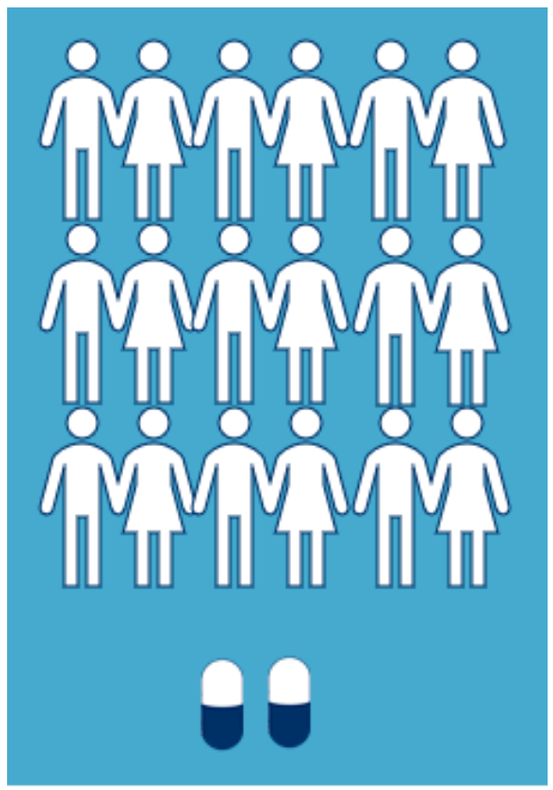


Observational Data

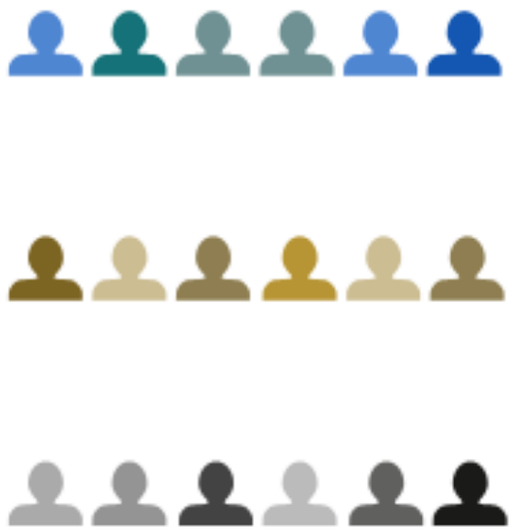
Population → Subgroup

Treatment Effect

Clinical Trials



Biomarker-based Trials



Observational Data

Population → Subgroup → Personalization

Traditional Methods

Traditional Methods

Rely on incorrect assumptions

Traditional Methods

Rely on incorrect assumptions

- Normality (e.g., t test for sample size calculation in clinical trials).

Traditional Methods

Rely on incorrect assumptions

- Normality (e.g., t test for sample size calculation in clinical trials).
- Proportional hazard (e.g., survival data analysis).

Traditional Methods

Rely on incorrect assumptions

- Normality (e.g., t test for sample size calculation in clinical trials).
- Proportional hazard (e.g., survival data analysis).
- Strong parametric assumption (e.g., Poisson for count data).

Traditional Methods

Rely on incorrect assumptions

- Normality (e.g., t test for sample size calculation in clinical trials).
- Proportional hazard (e.g., survival data analysis).
- Strong parametric assumption (e.g., Poisson for count data).

Oversimplify framework

Traditional Methods

Rely on incorrect assumptions

- Normality (e.g., t test for sample size calculation in clinical trials).
- Proportional hazard (e.g., survival data analysis).
- Strong parametric assumption (e.g., Poisson for count data).

Oversimplify framework

- Homogeneous population.

Traditional Methods

Rely on incorrect assumptions

- Normality (e.g., t test for sample size calculation in clinical trials).
- Proportional hazard (e.g., survival data analysis).
- Strong parametric assumption (e.g., Poisson for count data).

Oversimplify framework

- Homogeneous population.
- Compare survival outcomes only based on firstling therapy, ignoring the subsequent savage treatments.

Why Bayesian Nonparametrics (BNP)?

Why Bayesian Nonparametrics (BNP)?

- Bayesian

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- BNP: Bayesian models that are not parametric (unbounded/growing/infinite number of parameters)

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- BNP: Bayesian models that are not parametric (unbounded/growing/infinite number of parameters)

Advantages:

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- BNP: Bayesian models that are not parametric (unbounded/growing/infinite number of parameters)

Advantages:

- Uncertainty quantification (e.g., small sample size, decision making)

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- BNP: Bayesian models that are not parametric (unbounded/growing/infinite number of parameters)

Advantages:

- Uncertainty quantification (e.g., small sample size, decision making)
- Nonparametric

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

- BNP: Bayesian models that are not parametric (unbounded/growing/infinite number of parameters)

Advantages:

- Uncertainty quantification (e.g., small sample size, decision making)
- Nonparametric
- Wide support

Why Bayesian Nonparametrics (BNP)?

- Bayesian

$$\mathbb{P}(\text{parameters}|\text{data}) \propto \mathbb{P}(\text{data}|\text{parameters})\mathbb{P}(\text{parameters})$$

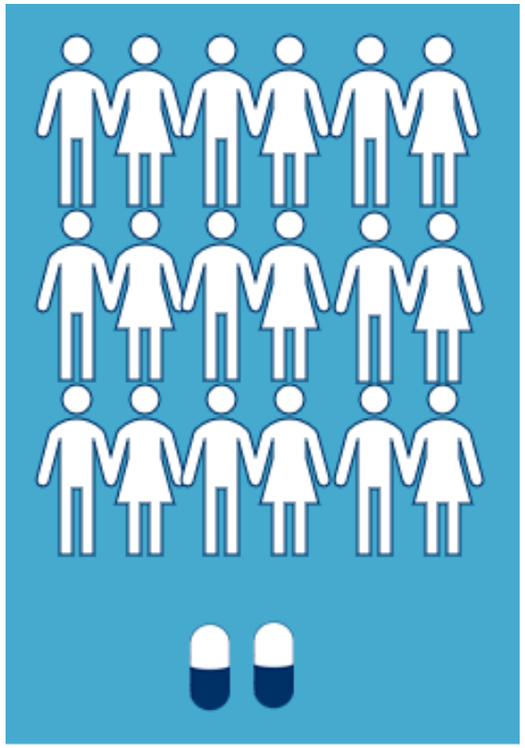
- BNP: Bayesian models that are not parametric (unbounded/growing/infinite number of parameters)

Advantages:

- Uncertainty quantification (e.g., small sample size, decision making)
- Nonparametric
- Wide support
- Flexible

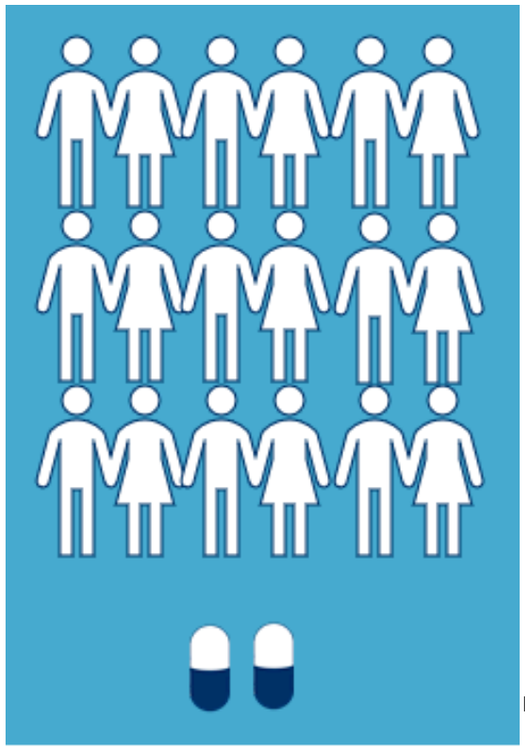
Why Bayesian Nonparametrics (BNP)?

Trial designs

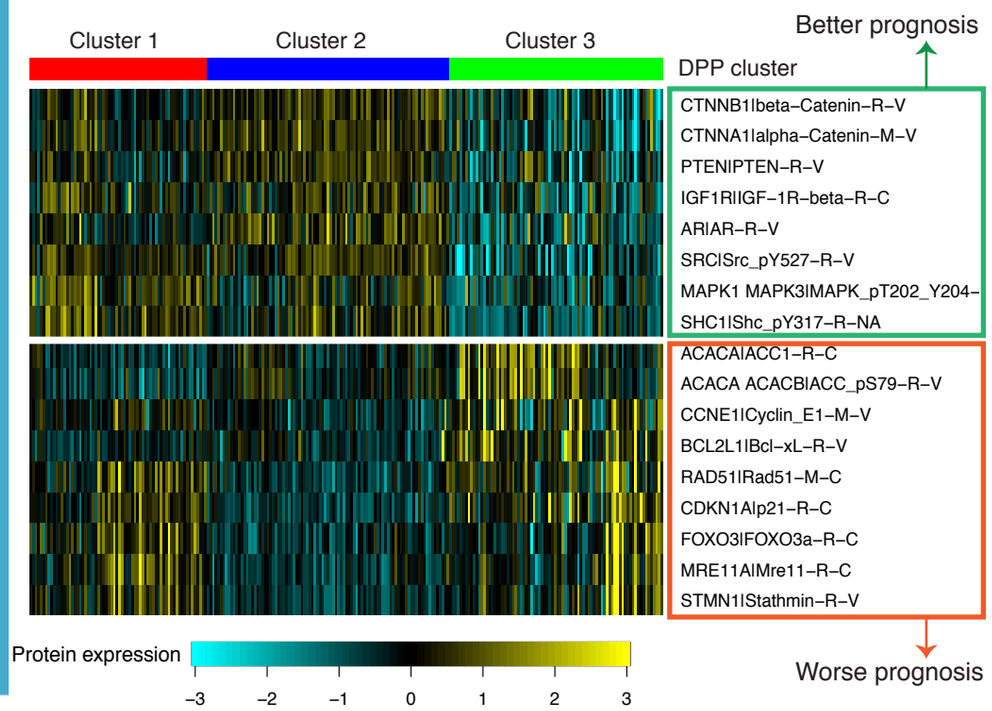


Why Bayesian Nonparametrics (BNP)?

Trial designs

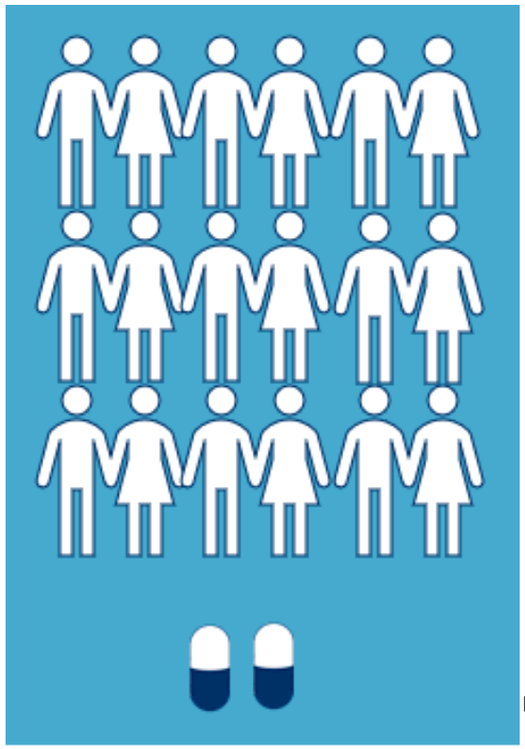


Patients clustering

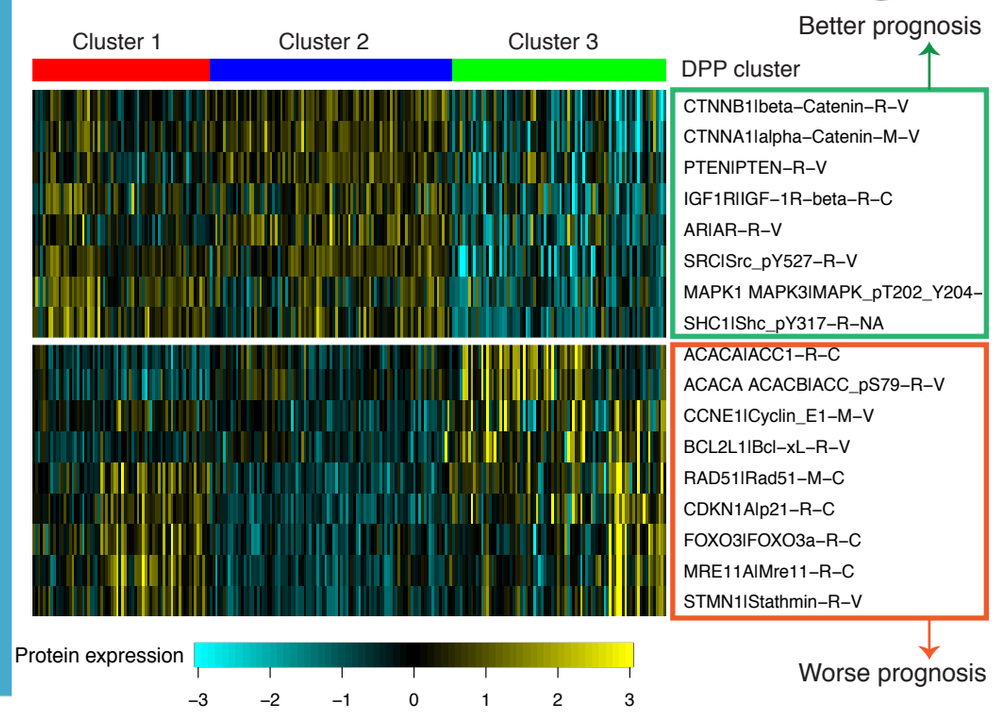


Why Bayesian Nonparametrics (BNP)?

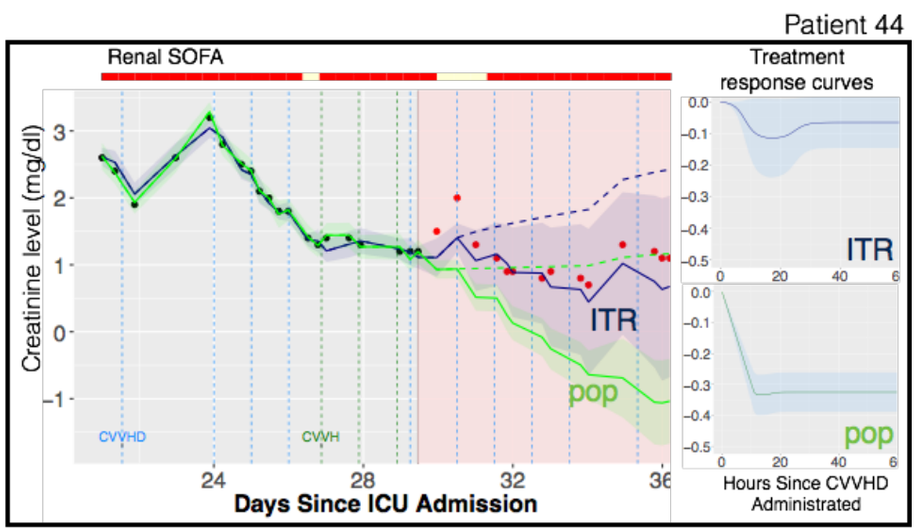
Trial designs



Patients clustering

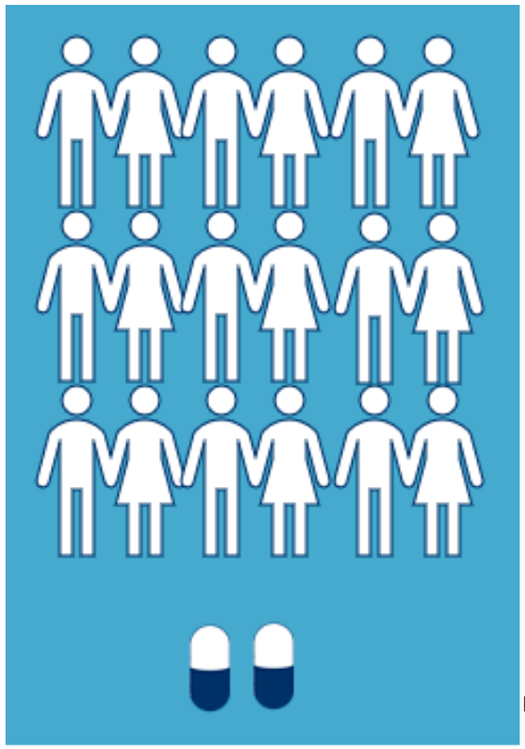


Continuous treatments

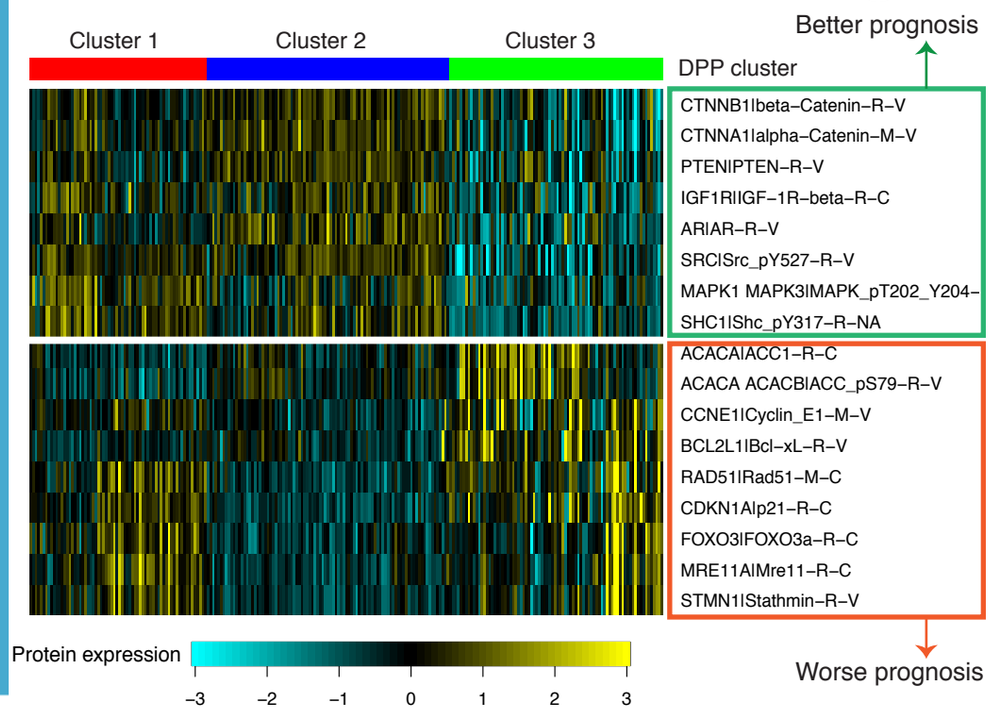


Why Bayesian Nonparametrics (BNP)?

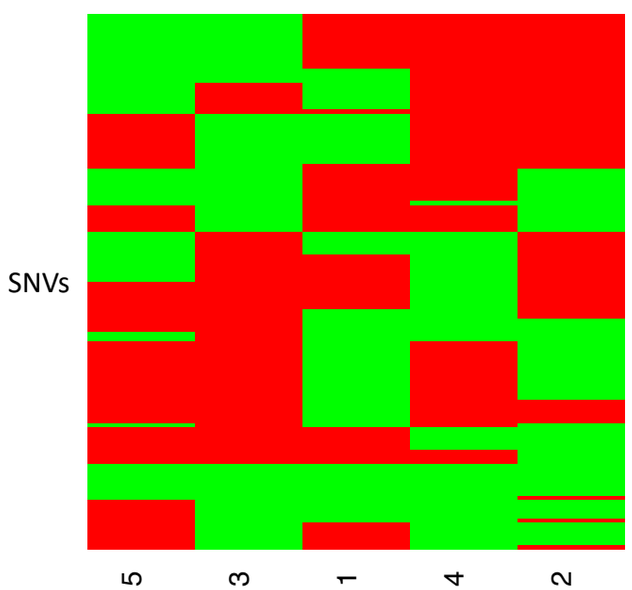
Trial designs



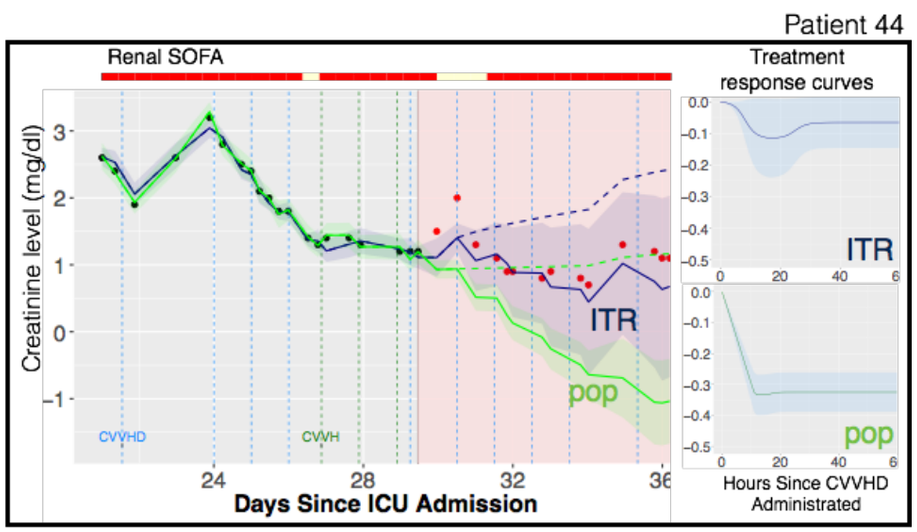
Patients clustering



Genetic signatures

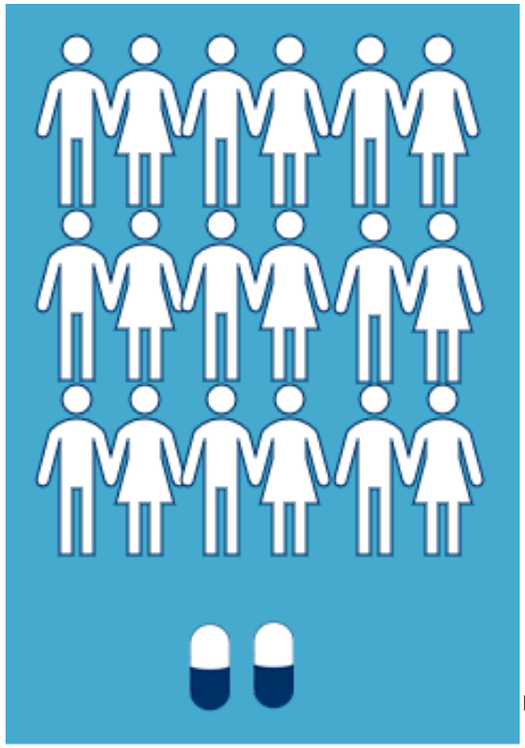


Continuous treatments

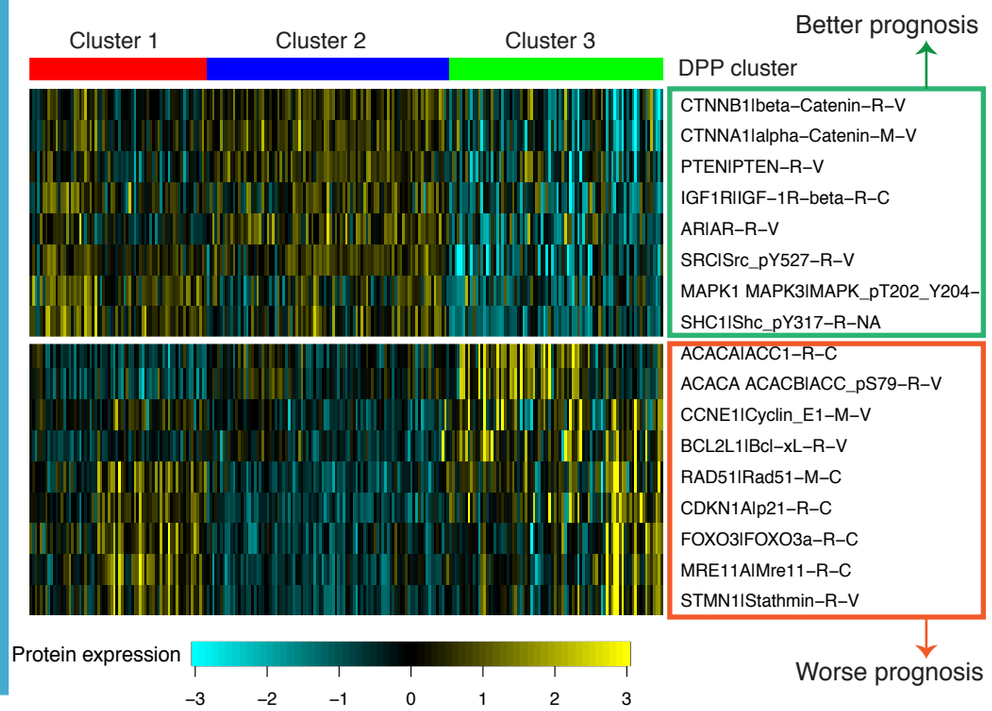


Why Bayesian Nonparametrics (BNP)?

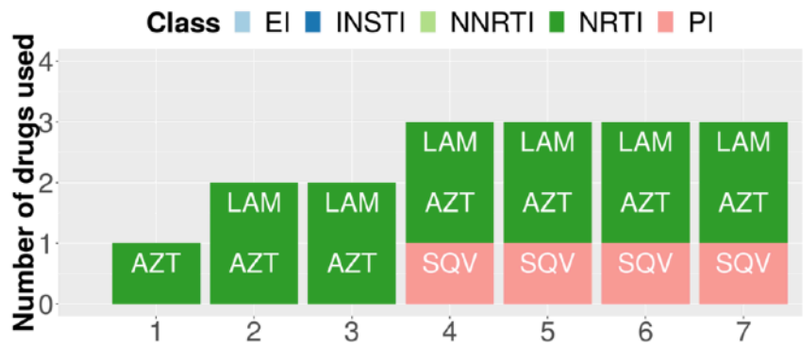
Trial designs



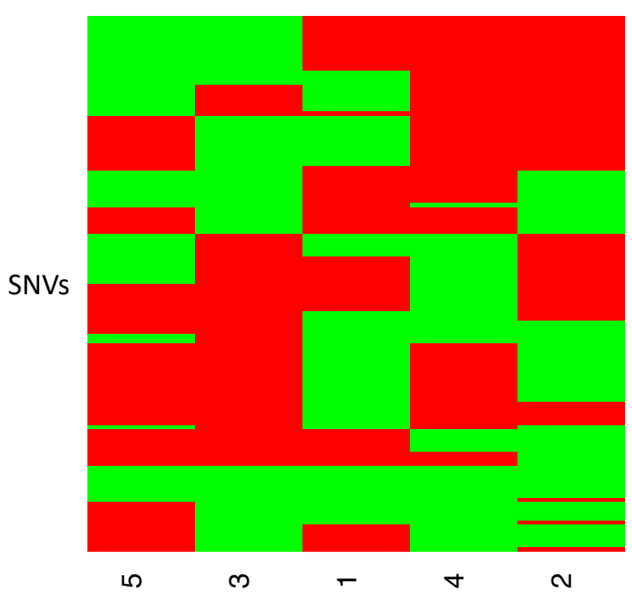
Patients clustering



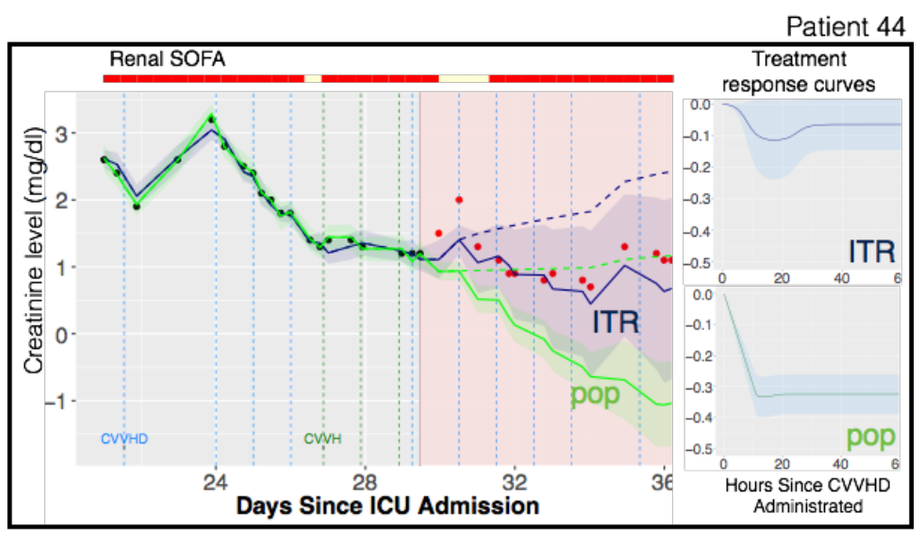
Reinforcement learning



Genetic signatures



Continuous treatments



Outline

- **Part 1:** *Monday*
  - **Density estimation** for efficient clinical trial designs
  - **Regression** for precision dosing

- **Part 2:** *Wednesday*
  - **Clustering** for subgroup finding
  - **Latent feature models** for tumor heterogeneity

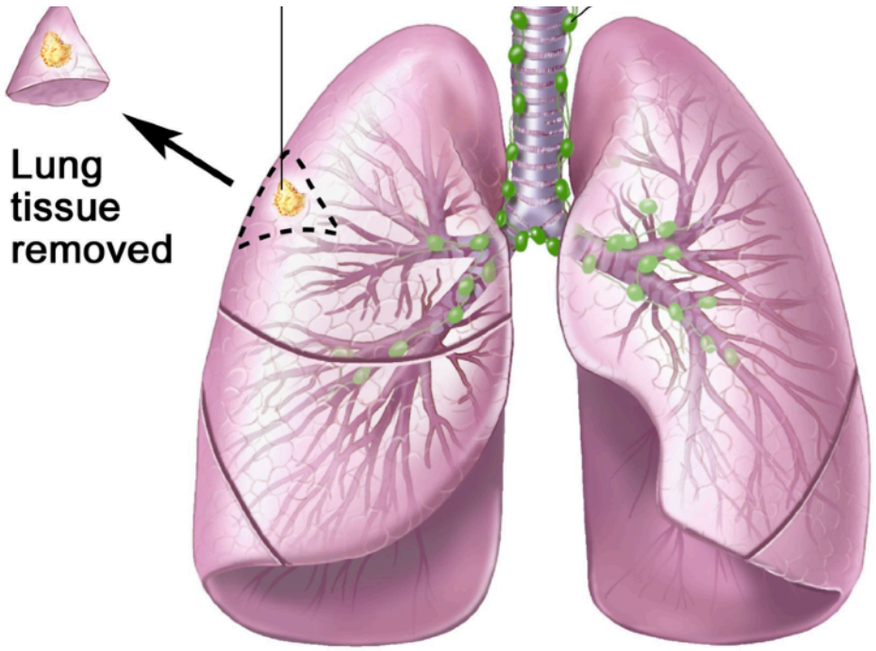
- **Part 3:** *Friday*
  - Estimating treatment effects from observational data

Outline

- **Part 1:** *Monday*
  - **Density estimation** for efficient clinical trial designs
  - **Regression** for precision dosing
- **Part 2:** *Wednesday*
  - **Clustering** for subgroup finding
  - **Latent feature models** for tumor heterogeneity
- **Part 3:** *Friday*
  - Estimating treatment effects from observational data

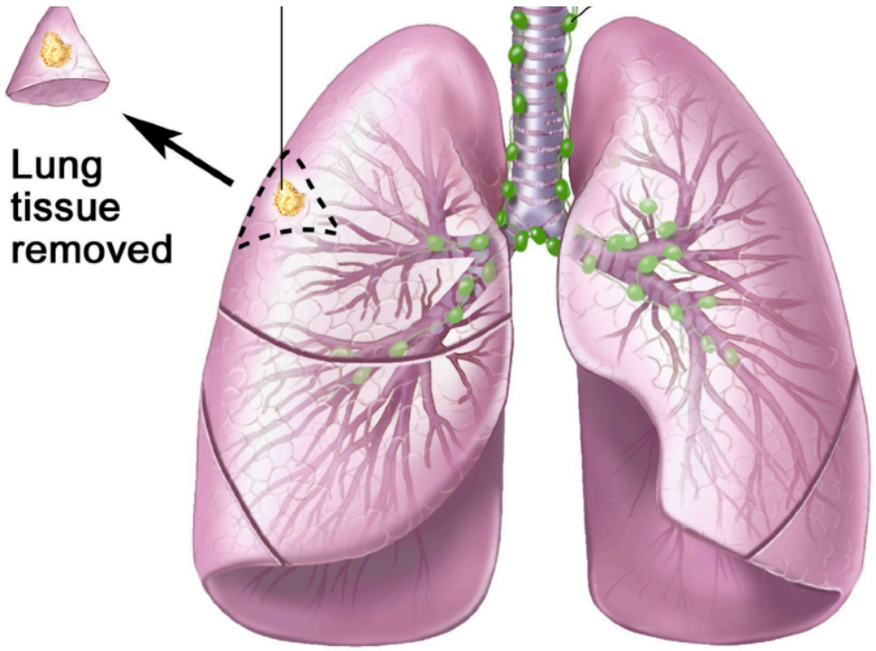
Motivating Trial

Pulmonary resection



Motivating Trial

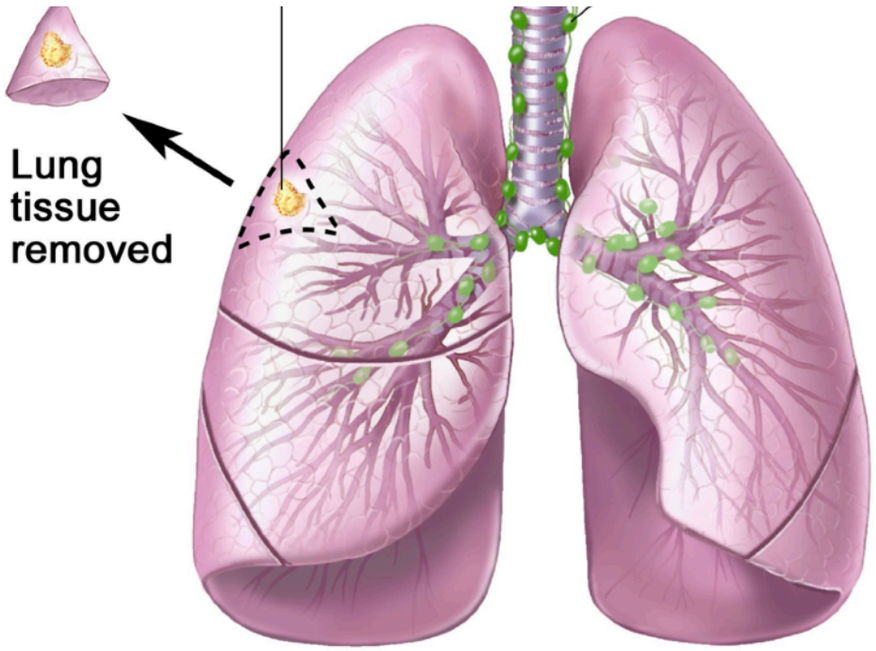
Pulmonary resection



Intraoperative air leaks (IALs) occur in 48 to 75% of patients.

Motivating Trial

Pulmonary resection

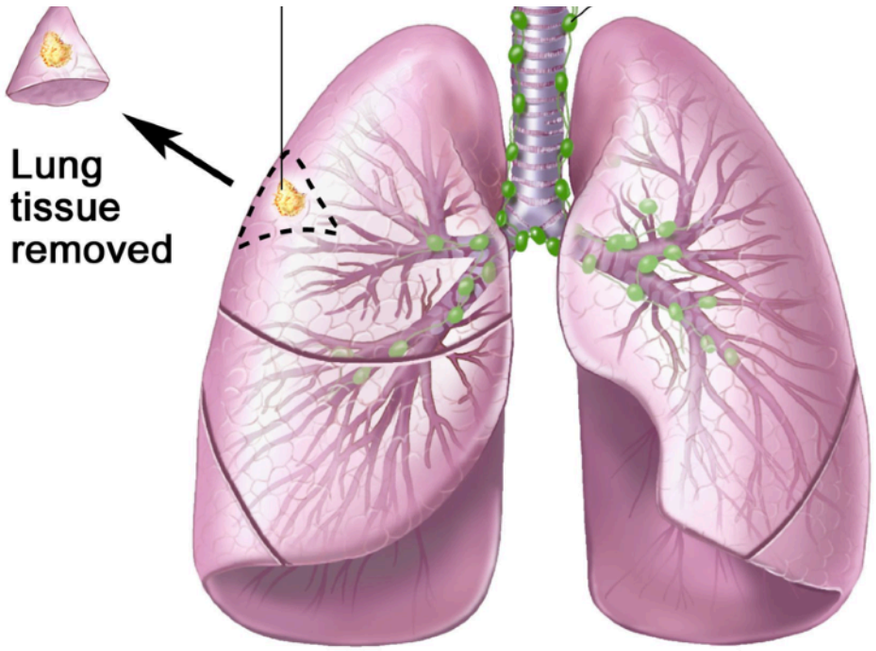


Intraoperative air leaks (IALs) occur in 48 to 75% of patients.

- Routine use: sutures and stapling devices

Motivating Trial

Pulmonary resection

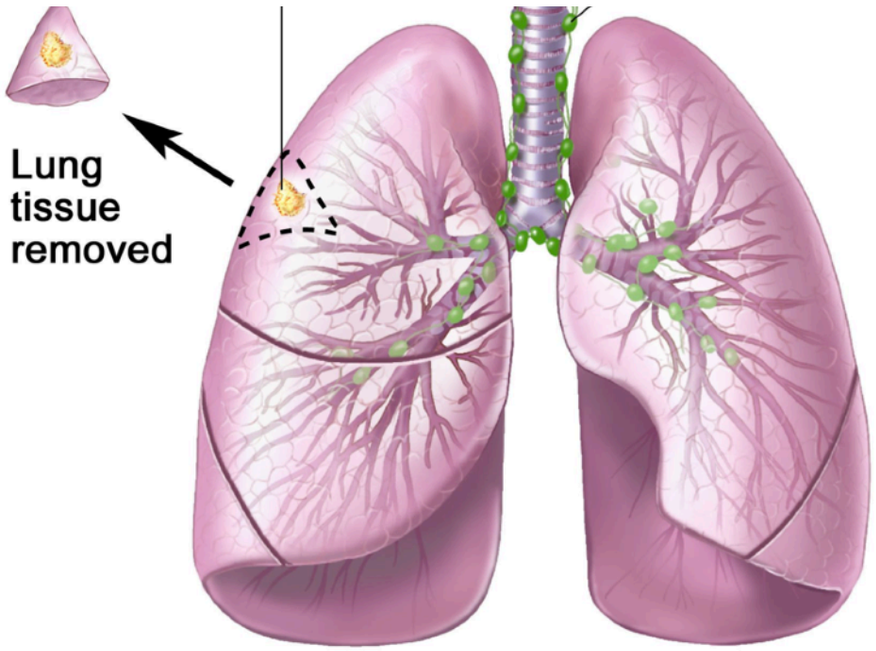


Intraoperative air leaks (IALs) occur in 48 to 75% of patients.

- Routine use: sutures and stapling devices
- Risks: postoperative pain, infections, longer hospitalization, economic, etc

Motivating Trial

Pulmonary resection



Intraoperative air leaks (IALs) occur in 48 to 75% of patients.

- Routine use: sutures and stapling devices
- Risks: postoperative pain, infections, longer hospitalization, economic, etc
- New way: Progel (liquid sealants)

Motivating Trial

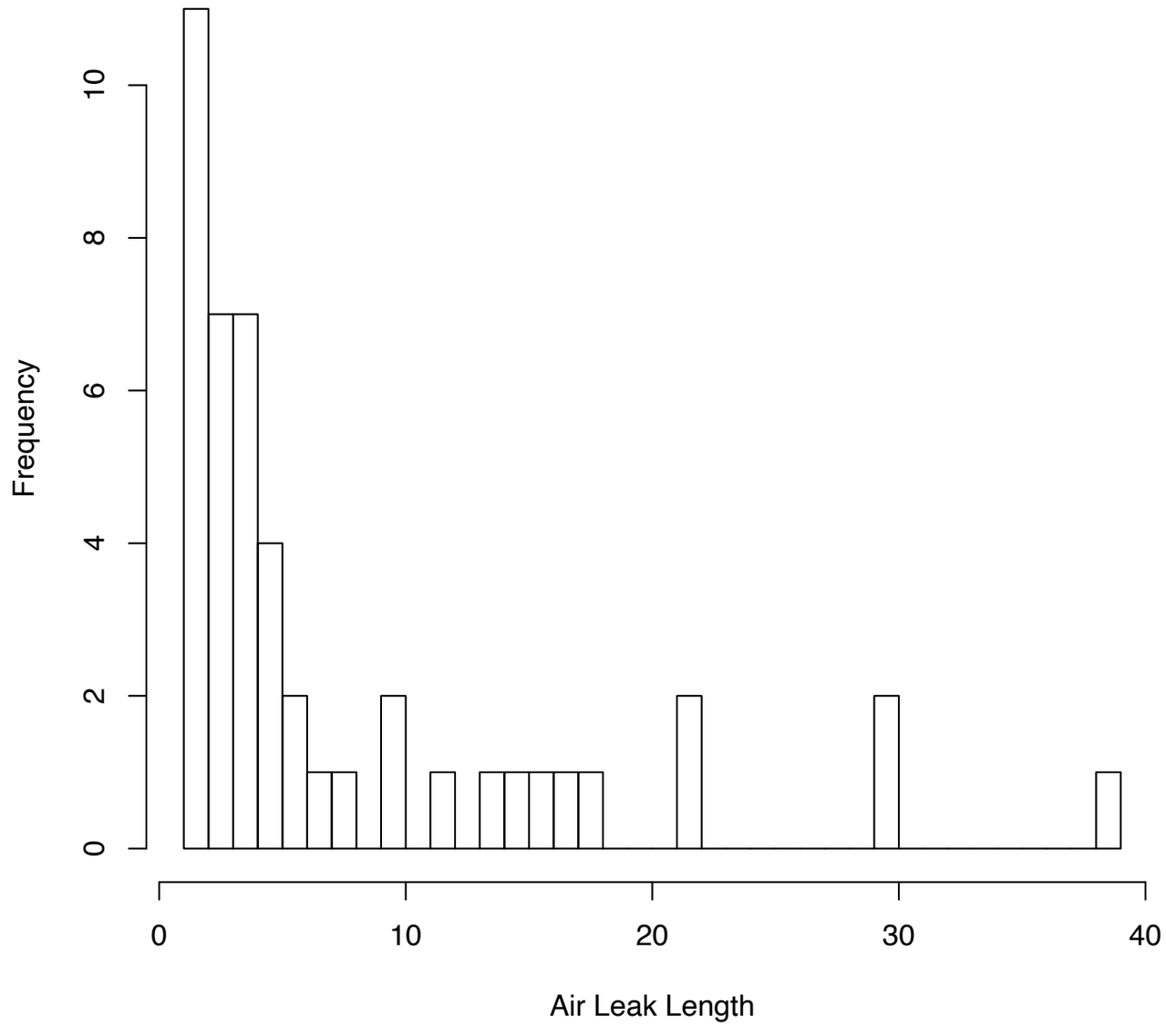
Goal: design a randomized trial comparing

Progel versus standard care!

Motivating Trial

Goal: design a randomized trial comparing

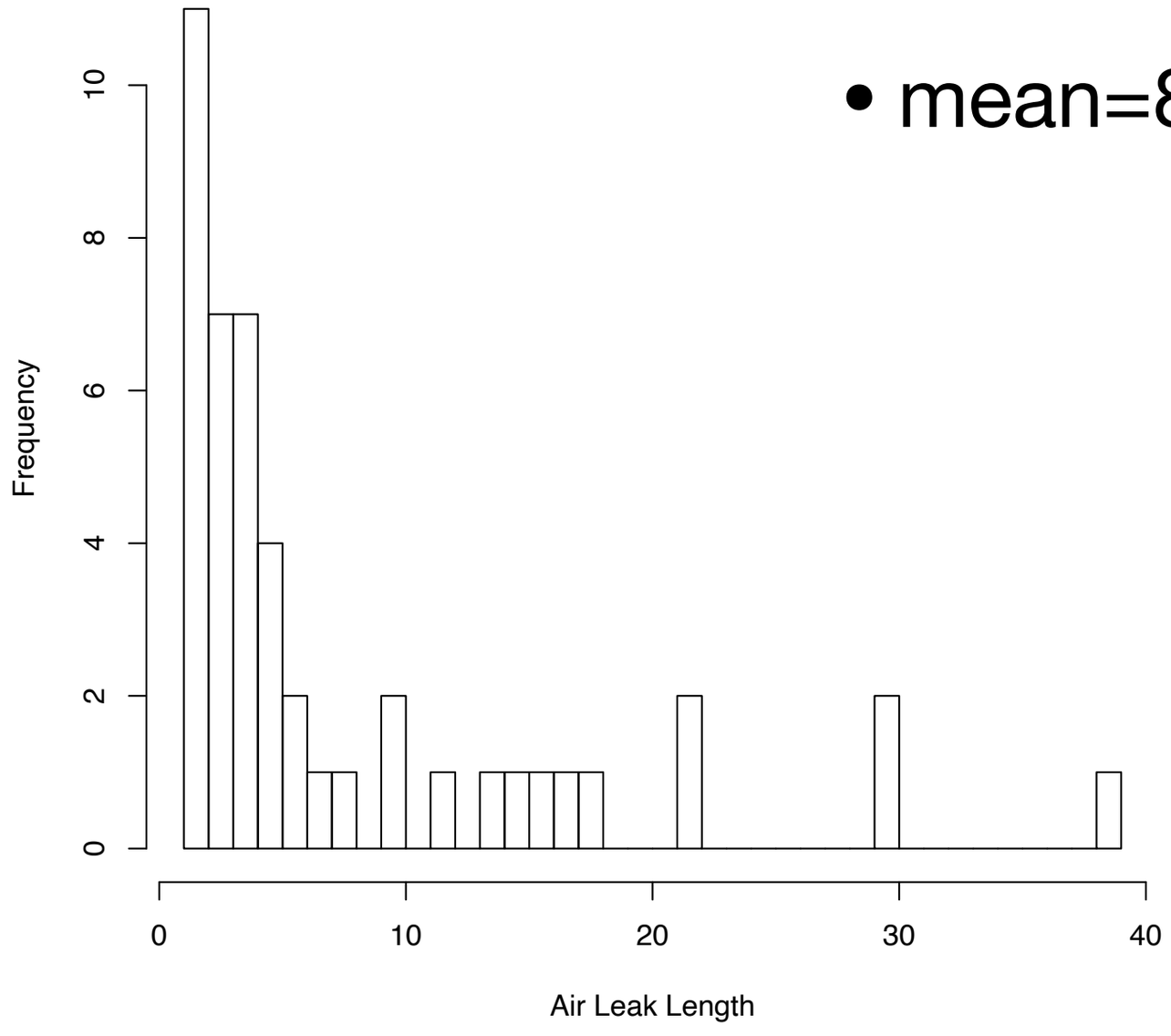
Progel versus standard care!



Motivating Trial

Goal: design a randomized trial comparing

Progel versus standard care!

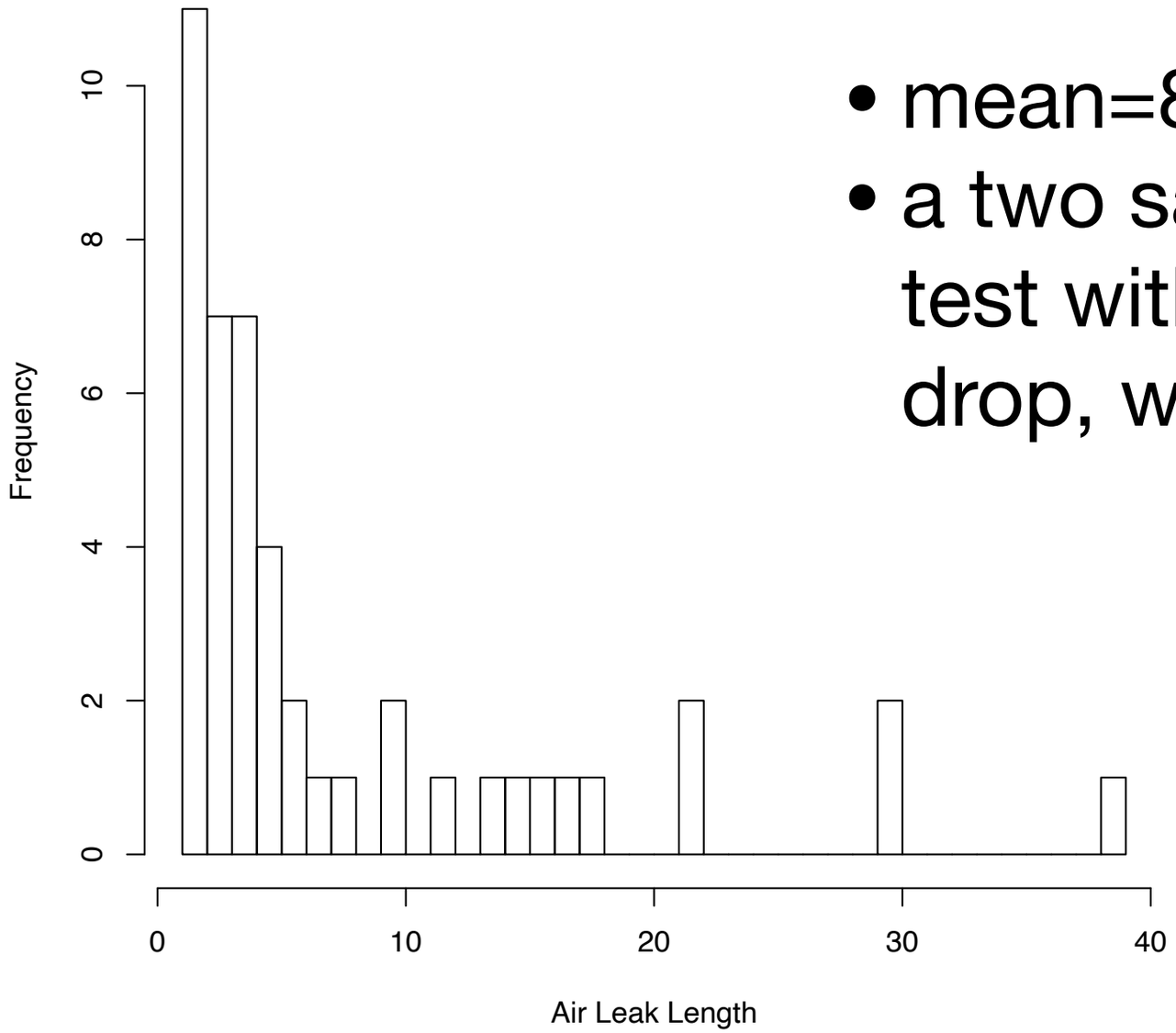


- mean=8, sd=8.76

Motivating Trial

Goal: design a randomized trial comparing

Progel versus standard care!

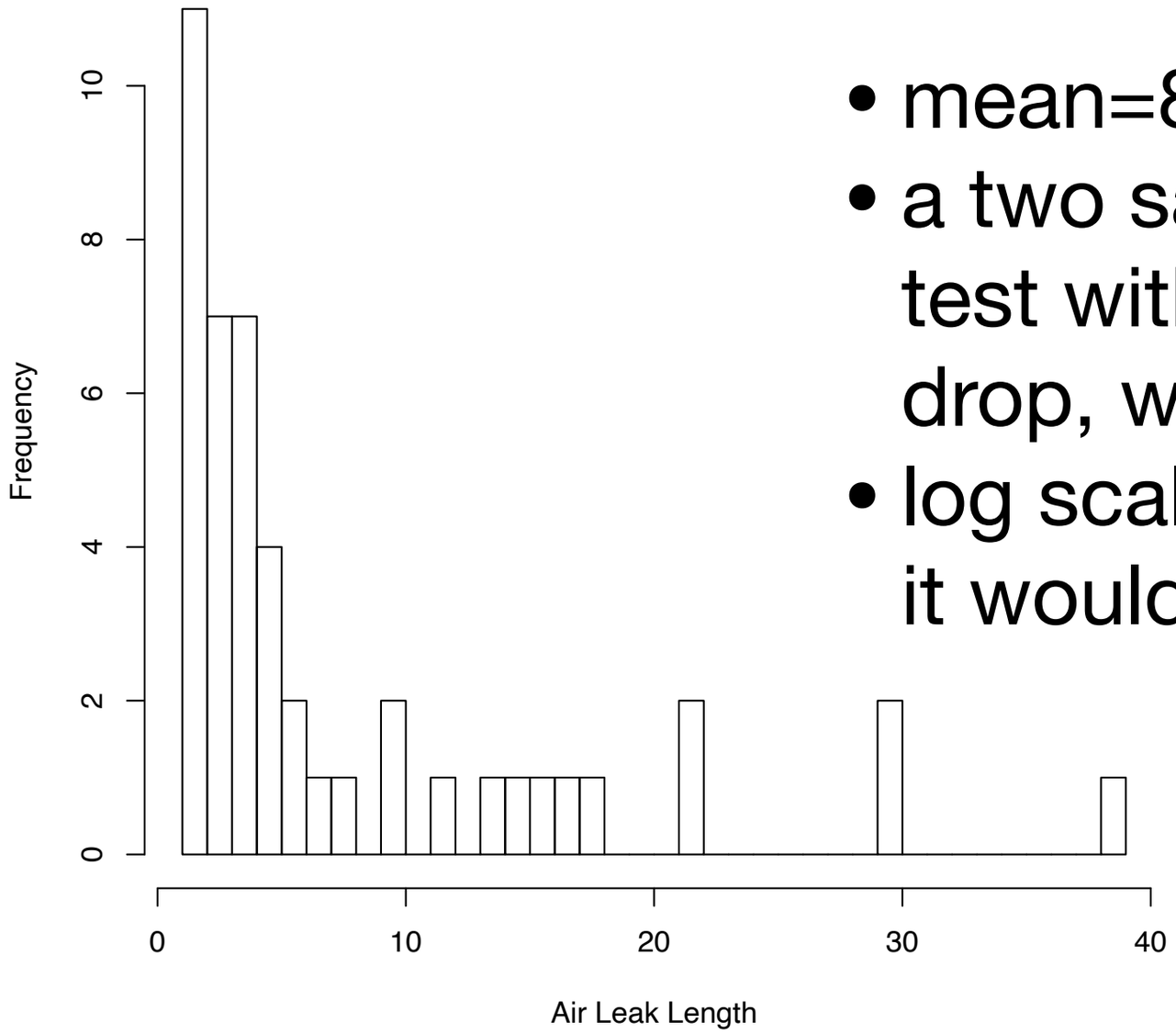


- mean=8, sd=8.76
- a two sample one-sided 0.05-level *t* test with power 0.8 to detect a 25% drop, would require n = 476.

Motivating Trial

Goal: design a randomized trial comparing

Progel versus standard care!

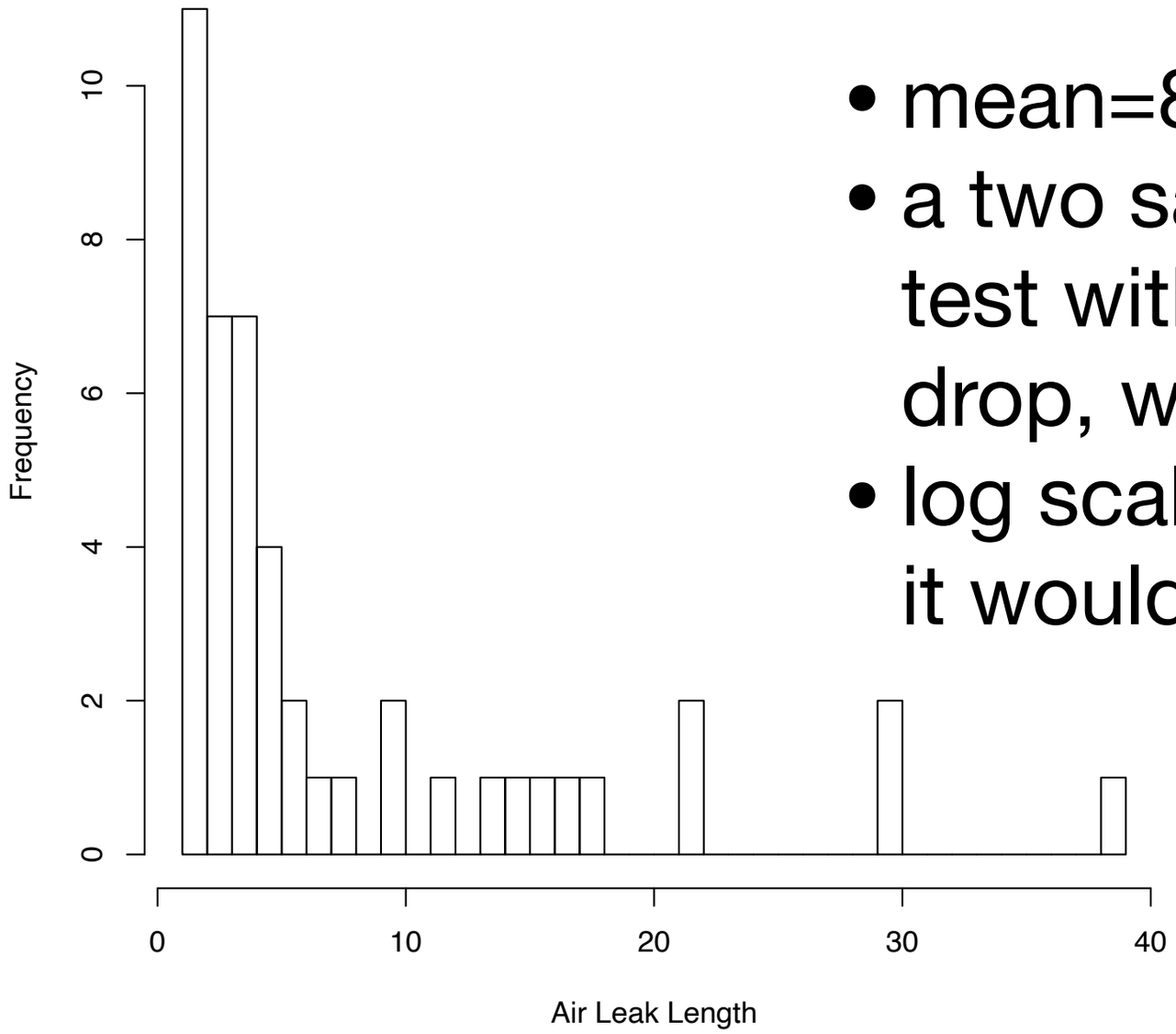


- mean=8, sd=8.76
- a two sample one-sided 0.05-level *t* test with power 0.8 to detect a 25% drop, would require n = 476.
- log scale mean = 1.61 and sd = 0.97, it would require n = 280.

Motivating Trial

Goal: design a randomized trial comparing

Progel versus standard care!



- mean=8, sd=8.76
- a two sample one-sided 0.05-level *t* test with power 0.8 to detect a 25% drop, would require n = 476.
- log scale mean = 1.61 and sd = 0.97, it would require n = 280.

Standard parametric model is INAPPROPRIATE!!

Modeling Considerations

Modeling Considerations

- Free of air leaks immediately following surgery is possible.

Modeling Considerations

- Free of air leaks immediately following surgery is possible.
- Consider the distribution of the time to resolve an air leak G_j as a mixture:

$$G_j = \nu_j \delta_0 + (1 - \nu_j) M_j, j = 0,1.$$

Modeling Considerations

- Free of air leaks immediately following surgery is possible.
- Consider the distribution of the time to resolve an air leak G_j as a mixture:

$$G_j = \nu_j \delta_0 + (1 - \nu_j) M_j, j = 0,1.$$

- M_1 is a left-shifted version of M_0.

Modeling Considerations

- Free of air leaks immediately following surgery is possible.
- Consider the distribution of the time to resolve an air leak G_j as a mixture:

$$G_j = \nu_j \delta_0 + (1 - \nu_j) M_j, j = 0,1.$$

- M_1 is a left-shifted version of M_0.
- Progel is inert, it cannot react chemically with the patient's lung tissue

Modeling Considerations

- Free of air leaks immediately following surgery is possible.
- Consider the distribution of the time to resolve an air leak G_j as a mixture:

$$G_j = \nu_j \delta_0 + (1 - \nu_j) M_j, j = 0,1.$$

- M_1 is a left-shifted version of M_0.
- Progel is inert, it cannot react chemically with the patient's lung tissue
- not a potential source of infection

Modeling Considerations

- Free of air leaks immediately following surgery is possible.
- Consider the distribution of the time to resolve an air leak G_j as a mixture:

$$G_j = \nu_j \delta_0 + (1 - \nu_j) M_j, j = 0,1.$$

- M_1 is a left-shifted version of M_0.
- Progel is inert, it cannot react chemically with the patient's lung tissue
- not a potential source of infection
- does not slow down the healing process

Modeling Considerations

- Free of air leaks immediately following surgery is possible.
- Consider the distribution of the time to resolve an air leak G_j as a mixture:

$$G_j = \nu_j \delta_0 + (1 - \nu_j) M_j, j = 0,1.$$

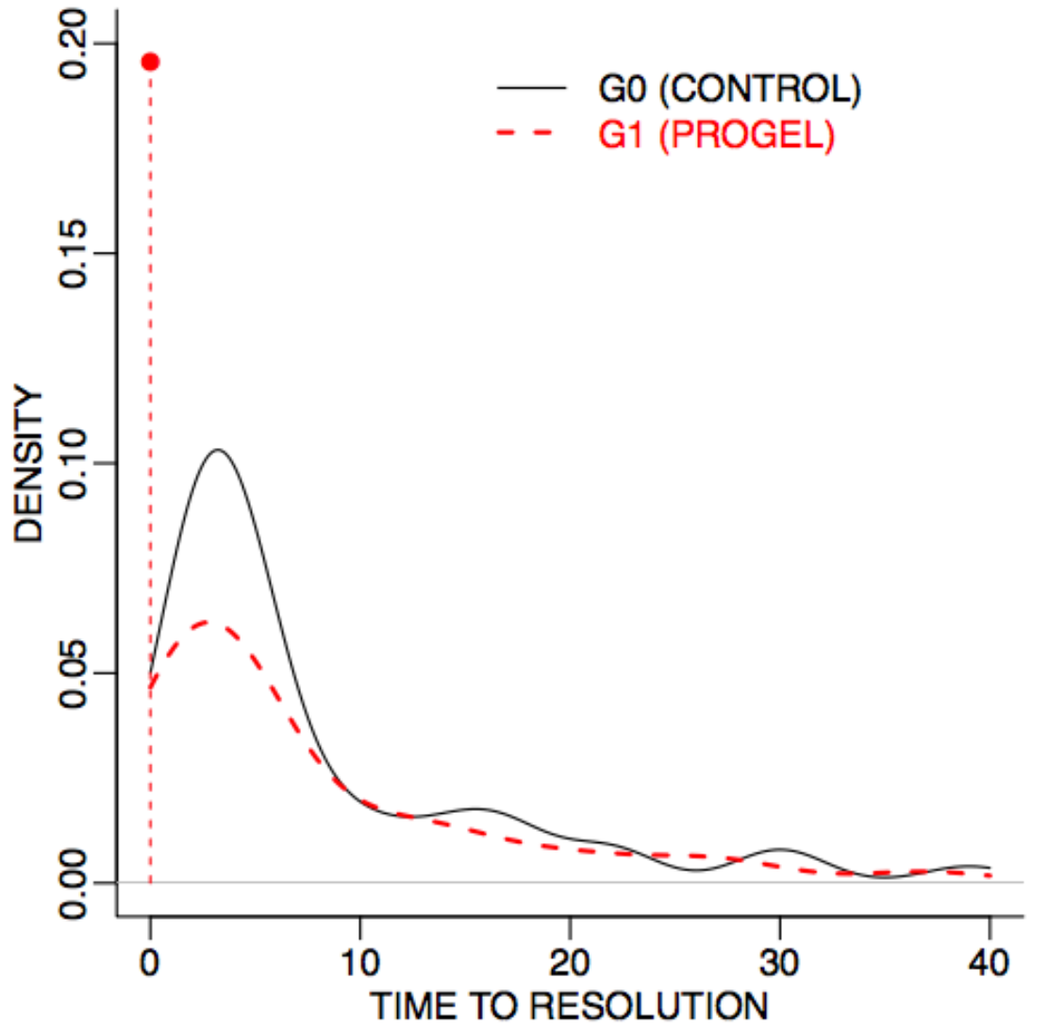
- M_1 is a left-shifted version of M_0.
- Progel is inert, it cannot react chemically with the patient's lung tissue
- not a potential source of infection
- does not slow down the healing process
- does not contribute to air leak formation

Modeling Considerations

Stochastic ordering on G_1 and G_0: Progel may be better, but not worse.

Modeling Considerations

Stochastic ordering on G_1 and G_0: Progel may be better, but not worse.



(a)

G_0 and G_1

Modeling Considerations

T: the time (in days) to resolve an air leaks. Define $Y = \log(T + 1)$.

Modeling Considerations

T: the time (in days) to resolve an air leaks. Define $Y = \log(T + 1)$.

$$\bar{U}_j = \int u(Y) G_j(dY), \ j = 0, 1$$

Modeling Considerations

T: the time (in days) to resolve an air leaks. Define $Y = \log(T + 1)$.

$$\bar{U}_j = \int u(Y) G_j(dY), \ j = 0, 1$$

Utility

Modeling Considerations

T: the time (in days) to resolve an air leaks. Define $Y = \log(T + 1)$.

$$\bar{U}_j = \int u(Y) G_j(dY), \ j = 0, 1$$

Utility

Trial Design: $Pr(\bar{U}_1 > \bar{U}_0 + \epsilon \mid \text{Data})$

Utility Considerations

A formal utility elicitation with our clinical collaborator. Both medical and economic:

Utility Considerations

A formal utility elicitation with our clinical collaborator. Both medical and economic:

- the most desirable resolution time is $T = 0$ (free of air leaks immediately after surgery, although this ideal outcome is almost never seen with standard care).

Utility Considerations

A formal utility elicitation with our clinical collaborator. Both medical and economic:

- the most desirable resolution time is $T = 0$ (free of air leaks immediately after surgery, although this ideal outcome is almost never seen with standard care).
- early ($1 \leq T \leq 5$) resolution of air leaks is very desirable and therefore the interval [1, 5] received a relatively high utility.

Utility Considerations

A formal utility elicitation with our clinical collaborator. Both medical and economic:

- the most desirable resolution time is $T = 0$ (free of air leaks immediately after surgery, although this ideal outcome is almost never seen with standard care).
- early ($1 \leq T \leq 5$) resolution of air leaks is very desirable and therefore the interval [1, 5] received a relatively high utility.
- the utilities drop off steeply for later resolution times ($T > 5$).

Utility Considerations

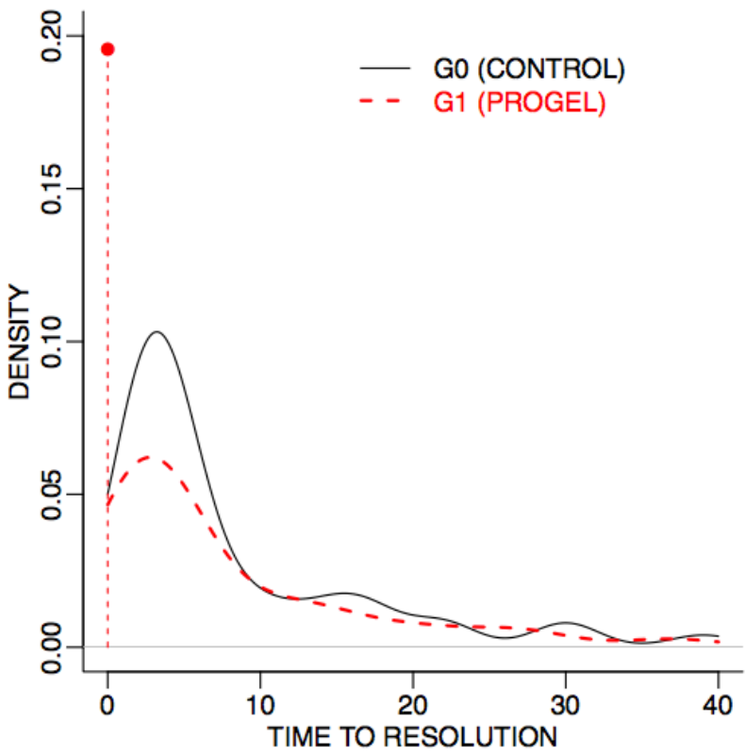
A formal utility elicitation with our clinical collaborator. Both medical and economic:

- the most desirable resolution time is $T = 0$ (free of air leaks immediately after surgery, although this ideal outcome is almost never seen with standard care).
- early ($1 \leq T \leq 5$) resolution of air leaks is very desirable and therefore the interval [1, 5] received a relatively high utility.
- the utilities drop off steeply for later resolution times ($T > 5$).

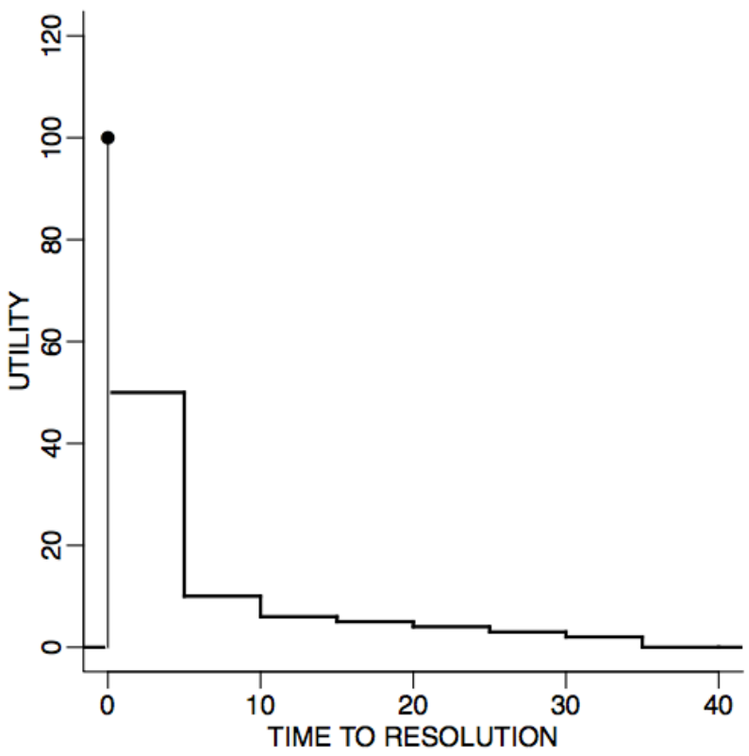
T (days)	0	5	10	15	20	25	30	35	≥ 40
Utility	100	50	10	6	5	4	3	2	0

Main Idea

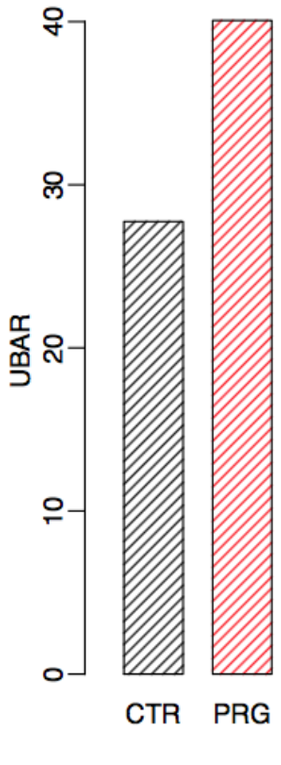
$$\bar{U}_j = \int u(Y) G_j(dY), \ j = 0, 1$$



(a)
G_0 and G_1



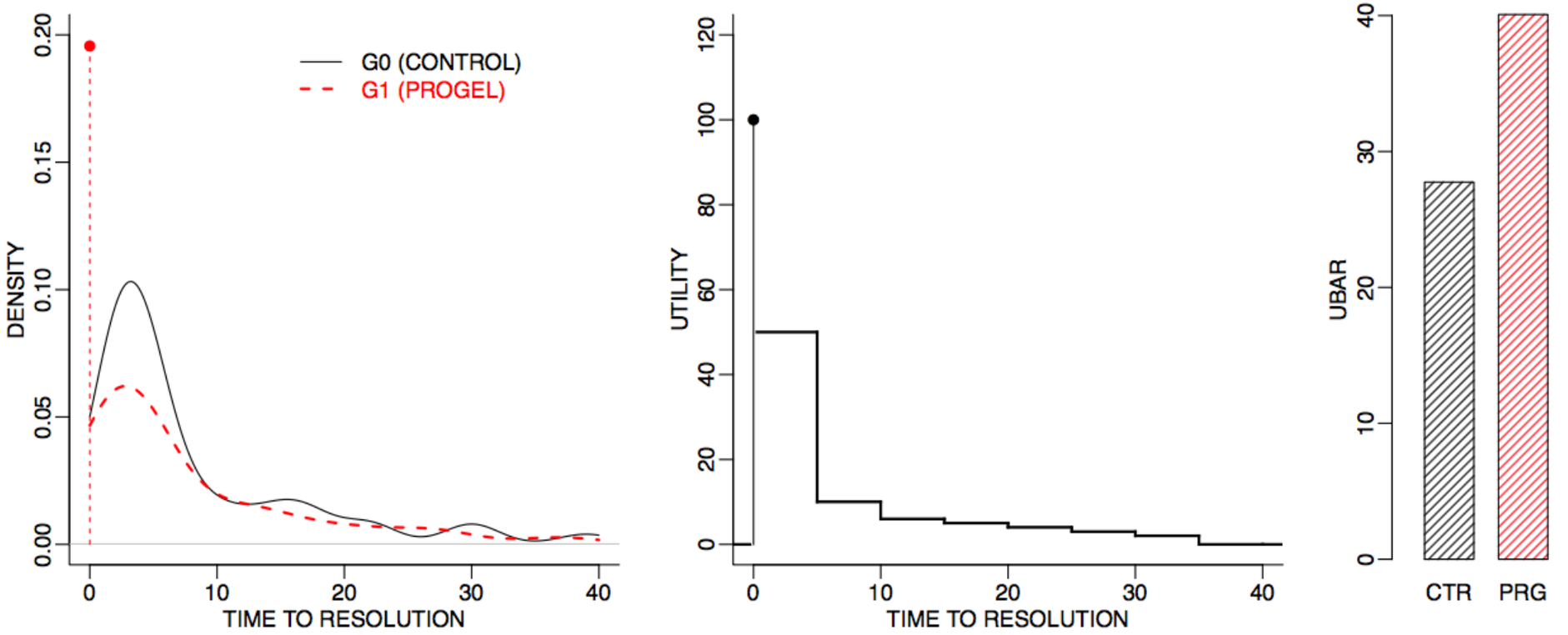
(b)
utilities u



(c)
$\overline{U}$.

Main Idea

$$\bar{U}_j = \int u(Y) G_j(dY), \quad j = 0, 1$$

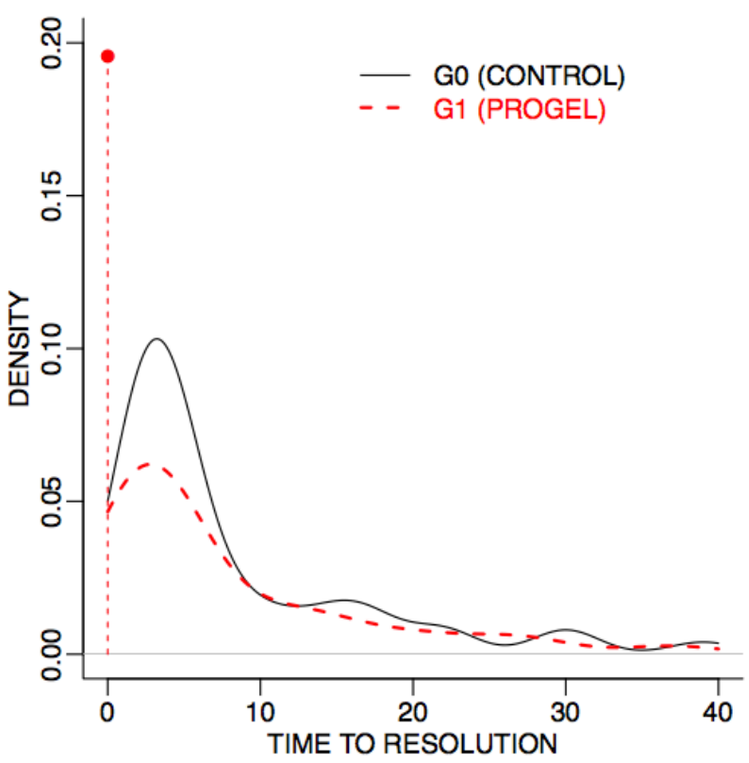


(a) G_0 and G_1 $\times$ (b) utilities u $=$ (c) $\overline{U}$.

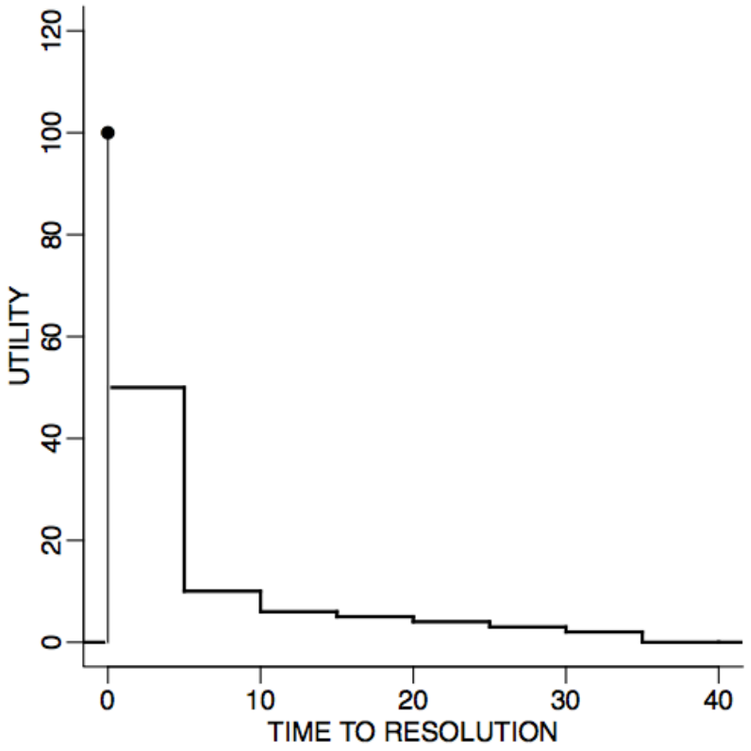
Main Idea

BNP

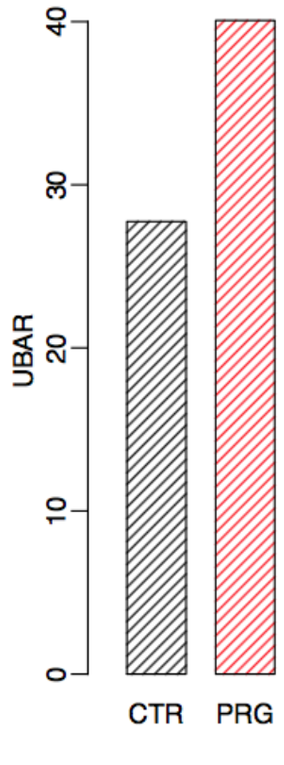
$$\bar{U}_j = \int u(Y) G_j(dY), \quad j = 0, 1$$



(a)
G_0 and G_1



(b)
utilities u



(c)
$\overline{U}$.

Dirichlet Process (DP)

The most commonly used BNP prior $p(F)$ for a random probability measure (Ferguson, 1973):

$$F \sim \mathrm{DP}(\alpha, F_0) \text{ with } F(y) = \sum_{h=1}^{\infty} w_h \delta_{m_h}(y) \,.$$

Dirichlet Process (DP)

The most commonly used BNP prior $p(F)$ for a random probability measure (Ferguson, 1973):

$$F \sim \mathrm{DP}(\alpha, F_0) \text{ with } F(y) = \sum_{h=1}^{\infty} w_h \delta_{m_h}(y) \,.$$

- Locations: $m_h \sim F_0$, i.i.d.

Dirichlet Process (DP)

The most commonly used BNP prior $p(F)$ for a random probability measure (Ferguson, 1973):

$$F \sim \mathrm{DP}(\alpha, F_0) \text{ with } F(y) = \sum_{h=1}^{\infty} w_h \delta_{m_h}(y)\,.$$

- Locations: $m_h \sim F_0$, i.i.d.
- Weights: $w_h = v_h \prod_{l<h} (1 - v_l)$ with $v_h \sim \mathrm{beta}(1,\alpha)$.

Dirichlet Process (DP)

The most commonly used BNP prior $p(F)$ for a random probability measure (Ferguson, 1973):

$$F \sim \mathrm{DP}(\alpha, F_0) \text{ with } F(y) = \sum_{h=1}^{\infty} w_h \delta_{m_h}(y)\,.$$

- Locations: $m_h \sim F_0$, i.i.d.
- Weights: $w_h = v_h \prod_{l<h} (1 - v_l)$ with $v_h \sim \mathrm{beta}(1,\alpha)$.

Stick-breaking

Dirichlet Process (DP)

Parameters:

- base measure: $F_0(A) = E\{F(A)\}$
- total mass α: $F(A) \sim \text{beta}(\alpha F_0(A), \alpha F_0(A^c))$.

Dirichlet Process (DP)

Parameters:

- base measure: $F_0(A) = E\{F(A)\}$
- total mass α: $F(A) \sim \text{beta}(\alpha F_0(A), \alpha F_0(A^c))$.

Posterior Inference:

- Likelihood: $p(y \mid F) = F$
- Prior: $F \sim \text{DP}(\alpha, F_0)$
- Posterior: $p(F \mid y) = \text{DP}(\alpha + 1, F^*)$ with

$$F^* = \frac{\alpha F_o + \delta_y}{\alpha + 1}.$$

DP Mixtures

- DP random measure: discrete F is not appropriate for many problems
- **DP mixture (DPM):** convolution of discrete F with (continuous) kernel, e.g., normal (Escobar & West, 1995)

DP Mixtures

- DP random measure: discrete F is not appropriate for many problems
- **DP mixture (DPM):** convolution of discrete F with (continuous) kernel, e.g., normal (Escobar & West, 1995)

$$G(y) = \int N(y \mid \theta, \sigma^2) dF(\theta), F \sim \text{DP}$$

$$= \sum_{h=1}^{\infty} w_h N(m_h, \sigma^2)$$

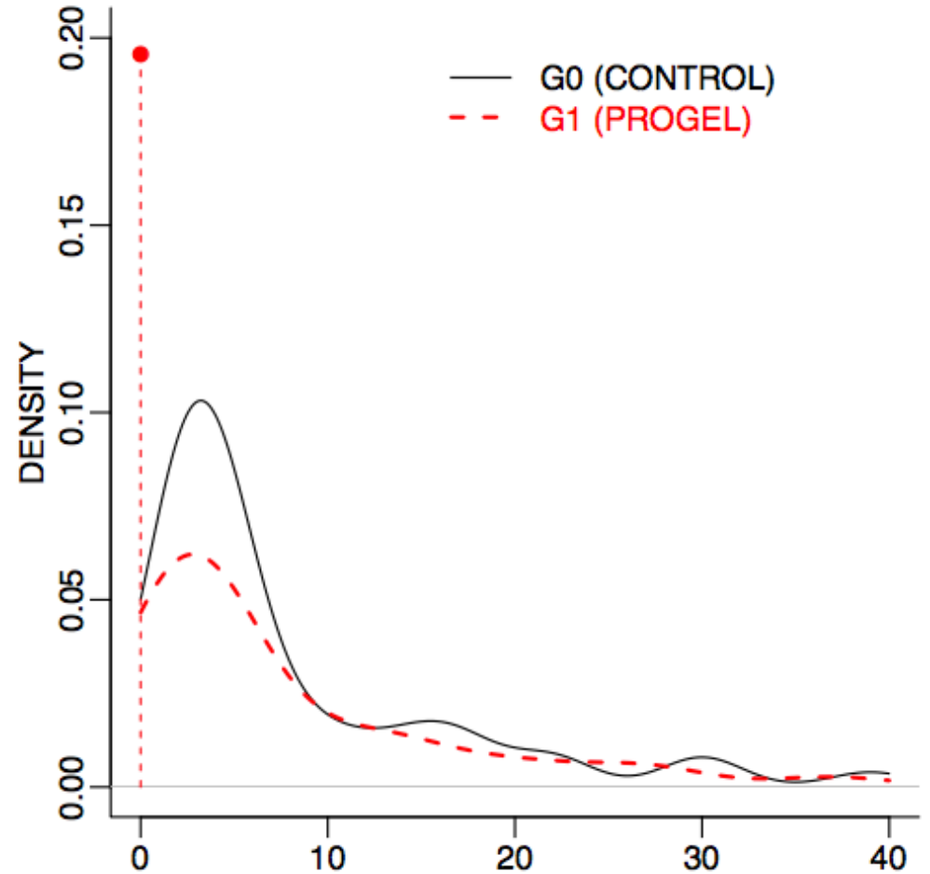
DP Mixtures

- DP random measure: discrete F is not appropriate for many problems
- **DP mixture (DPM):** convolution of discrete F with (continuous) kernel, e.g., normal (Escobar & West, 1995)

$$G(y) = \int N(y \mid \theta, \sigma^2) dF(\theta), F \sim \text{DP}$$

$$= \sum_{h=1}^{\infty} w_h N(m_h, \sigma^2)$$

Probability Model

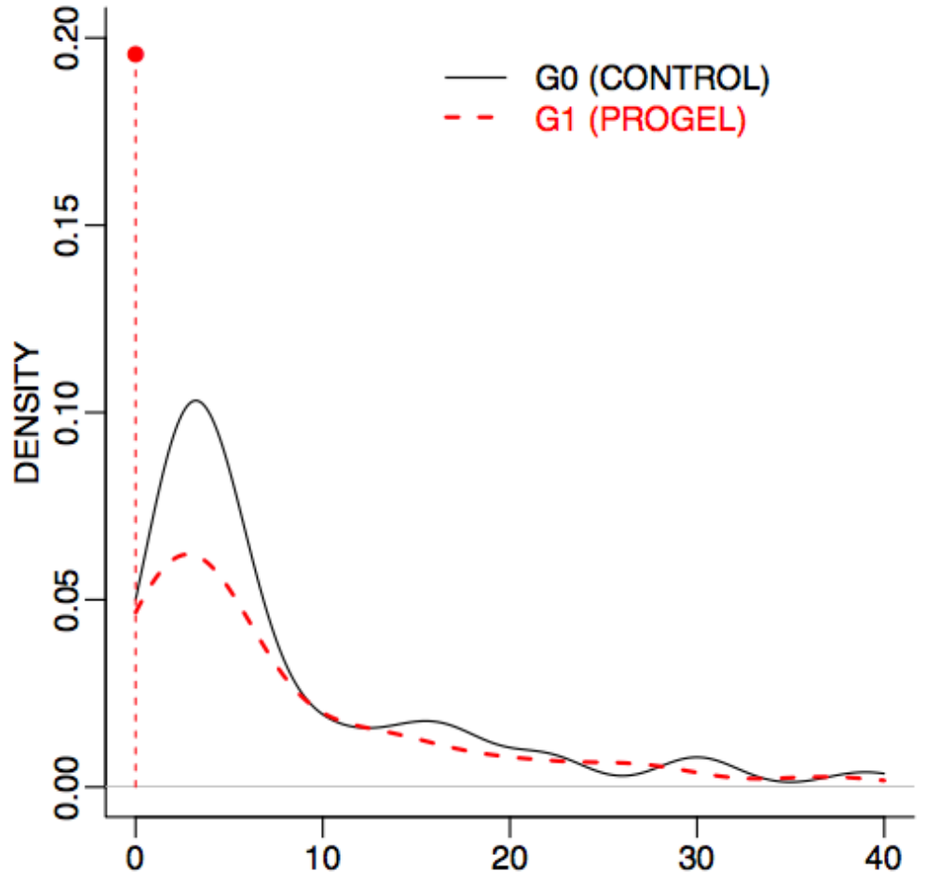


(a)

G_0 and G_1

Probability Model

$$G_j = \nu_{j0}\delta_0 + \sum_{h=1}^{\infty} \nu_{jh} N(\theta_{jh}, \sigma^2)$$

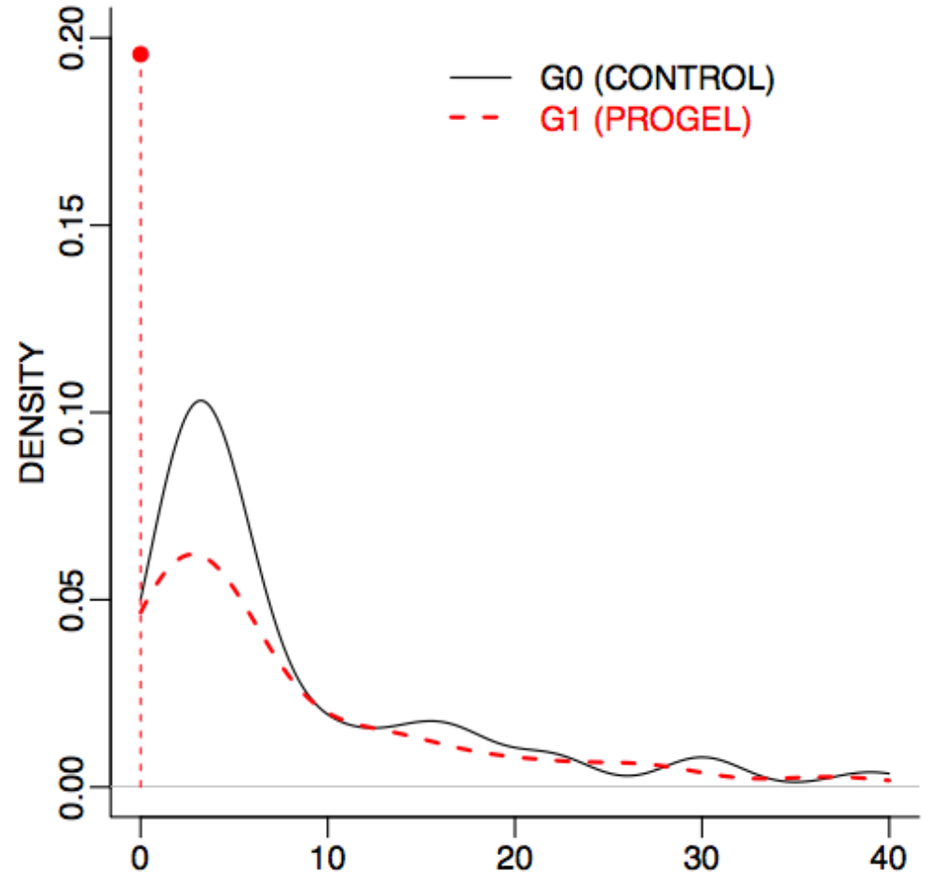


(a)
G_0 and G_1

Probability Model

$$G_j = \nu_{j0}\delta_0 + \sum_{h=1}^{\infty} \nu_{jh} N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) \sum_{h=1}^{\infty} w_h N(\theta_{jh}, \sigma^2)$$

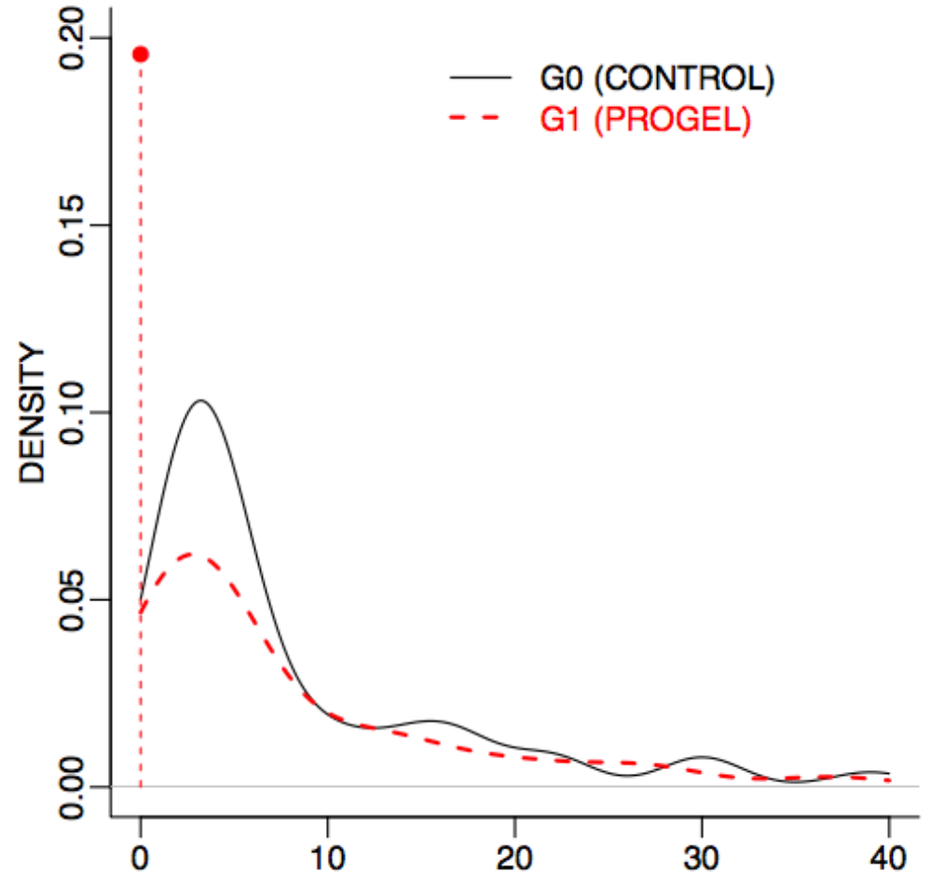


(a)
G_0 and G_1

Probability Model

$$G_j = \nu_{j0}\delta_0 + \sum_{h=1}^{\infty} \nu_{jh} N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) \sum_{h=1}^{\infty} w_h N(\theta_{jh}, \sigma^2)$$



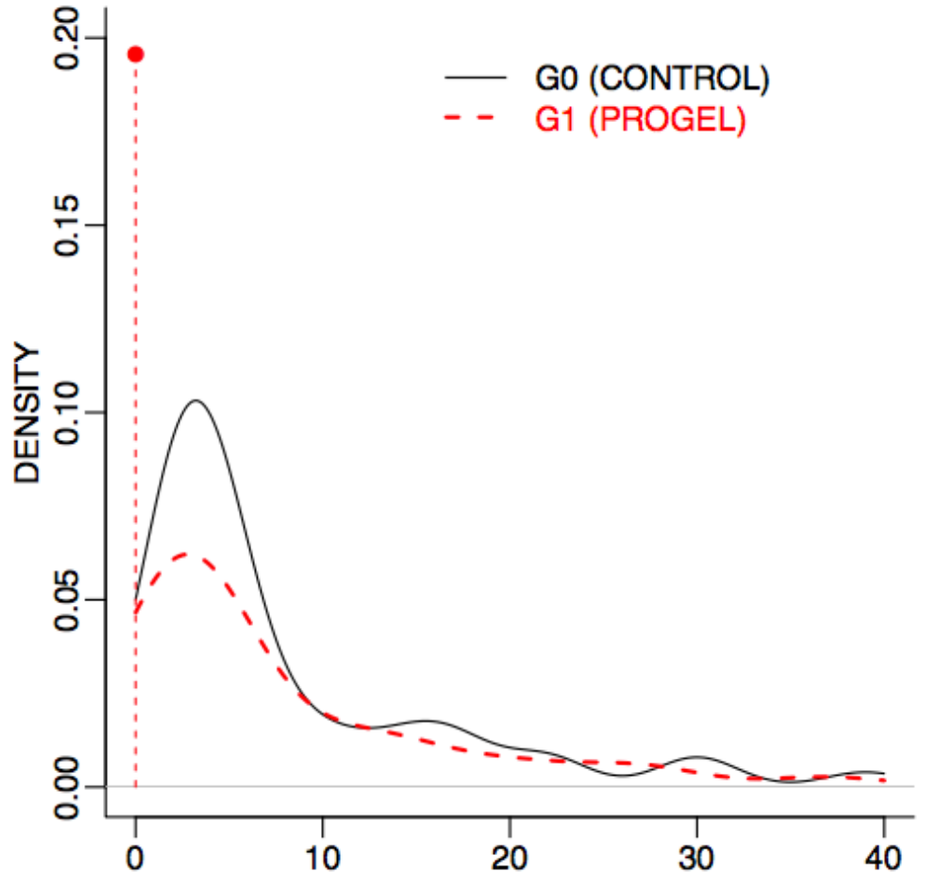
(a)
G_0 and G_1

Probability Model

$$G_j = \nu_{j0}\delta_0 + \sum_{h=1}^{\infty} \nu_{jh} N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) \sum_{h=1}^{\infty} w_h N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) M_j$$



(a)
G_0 and G_1

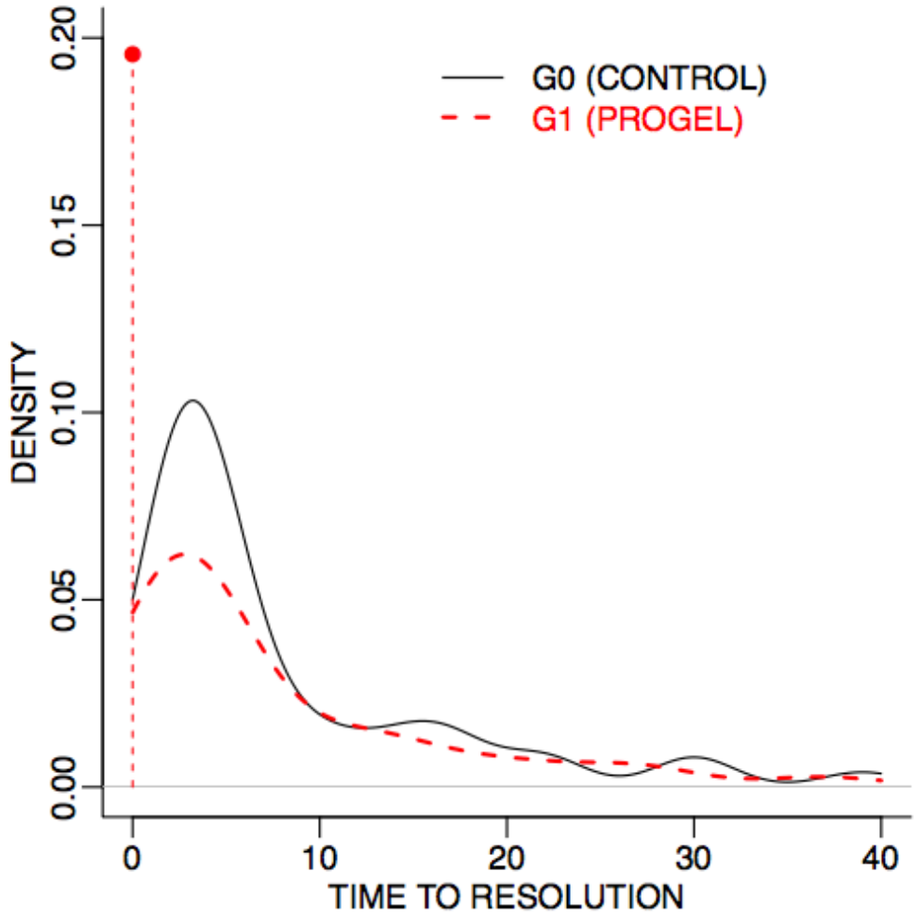
Probability Model

$$G_j = \nu_{j0}\delta_0 + \sum_{h=1}^{\infty} \nu_{jh} N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) \sum_{h=1}^{\infty} w_h N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) M_j$$

DP Mixtures



(a)
G_0 and G_1

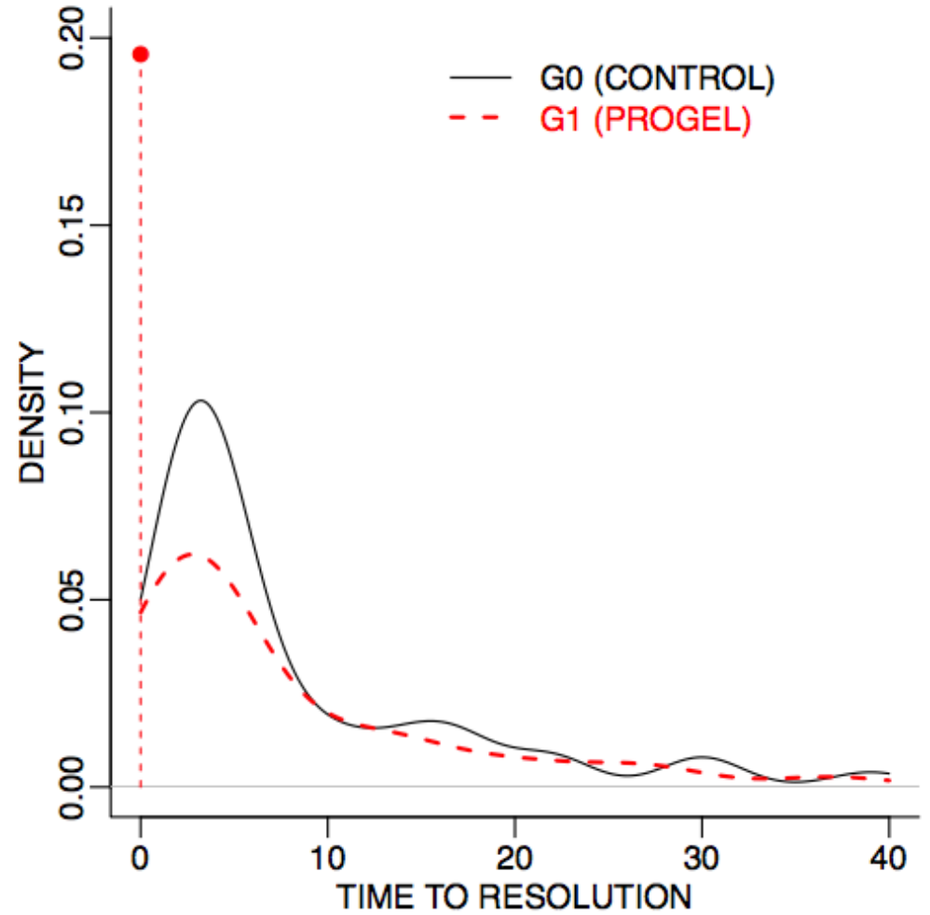
Probability Model

$$G_j = \nu_{j0}\delta_0 + \sum_{h=1}^{\infty} \nu_{jh} N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) \sum_{h=1}^{\infty} w_h N(\theta_{jh}, \sigma^2)$$

$$= \nu_{j0}\delta_0 + (1 - \nu_{j0}) M_j$$

DP Mixtures



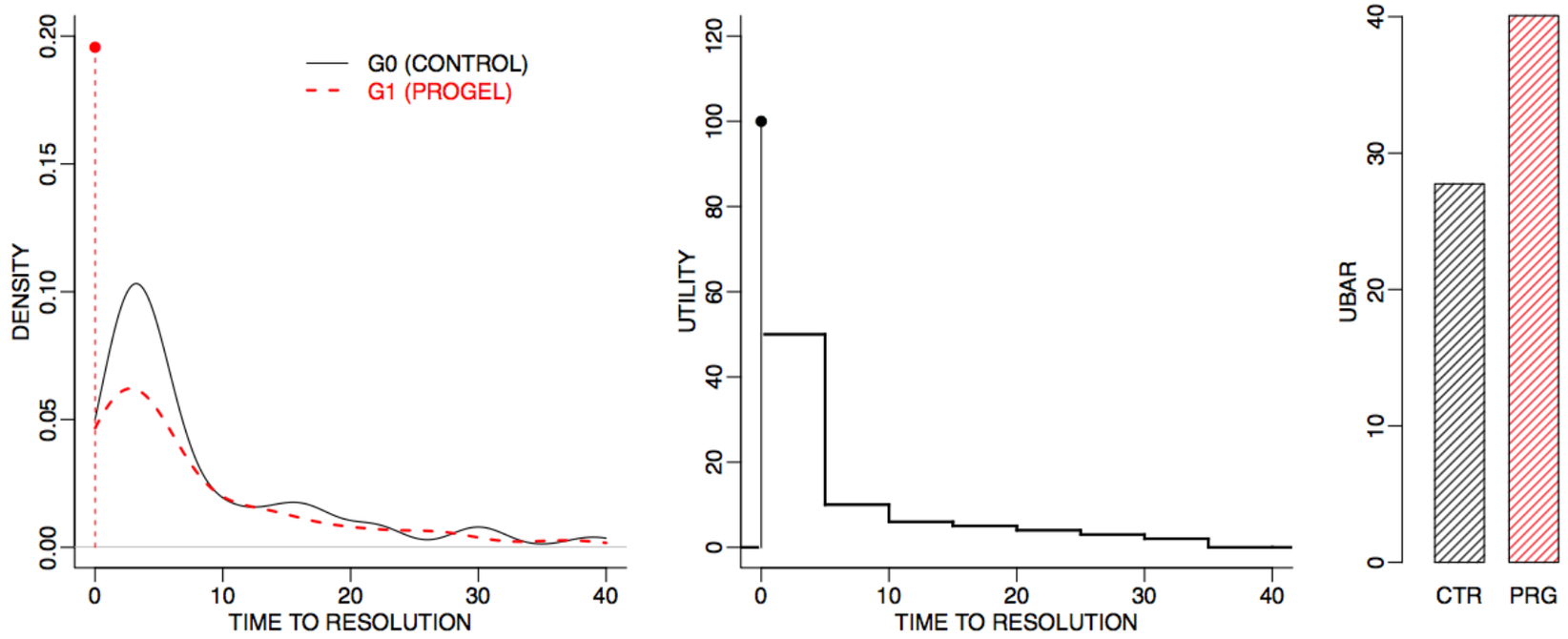
(a)
G_0 and G_1

Further prior beliefs can be added.

Overview

Utility **BNP**

$$\bar{U}_j = \int u(Y) G_j(dY), \quad j = 0, 1$$



(a) G_0 and G_1 × (b) utilities u = (c) $\overline{U}$.

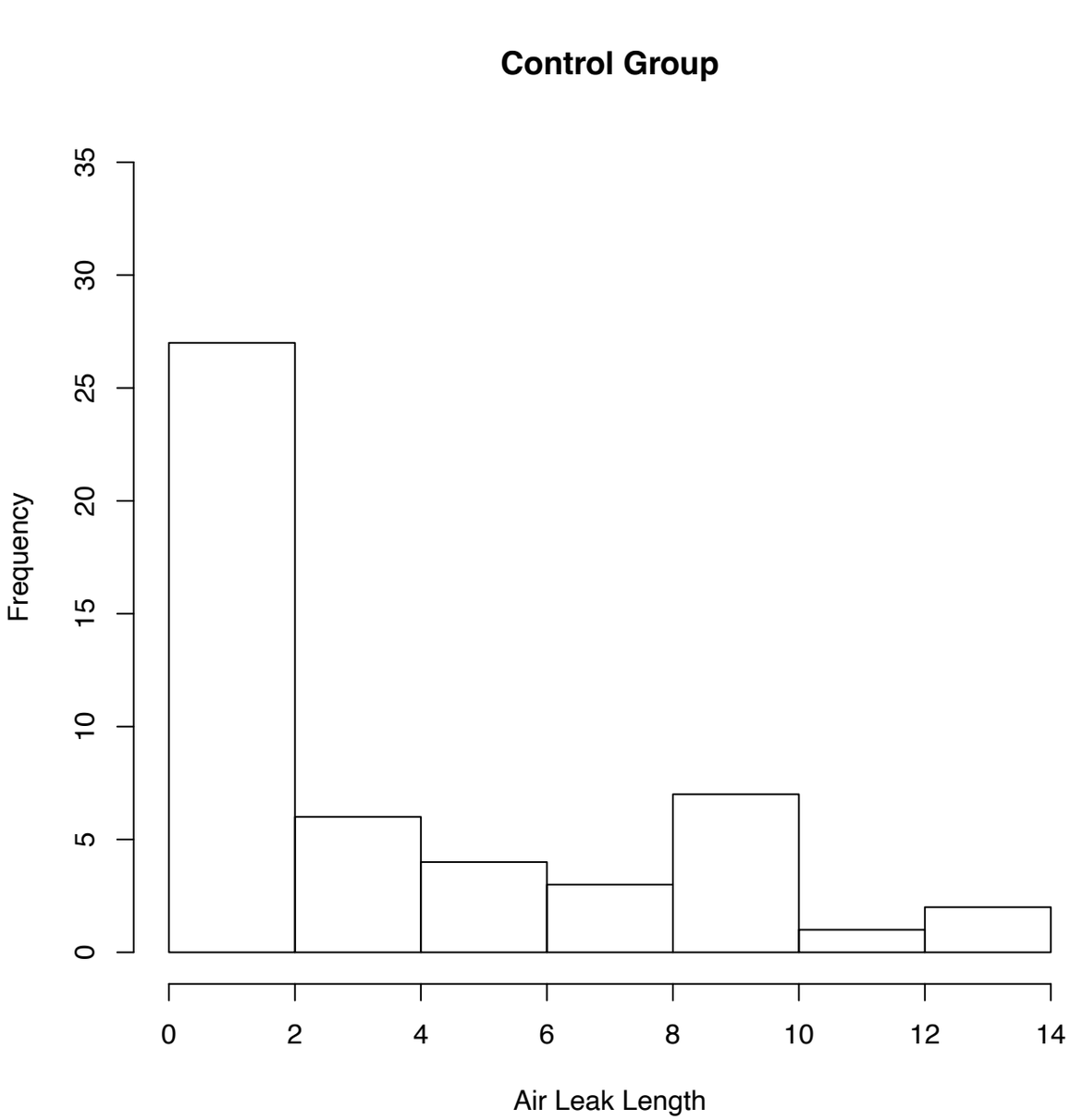
Trial Design: $Pr(\bar{U}_1 > \bar{U}_0 + \epsilon \mid \text{Data})$

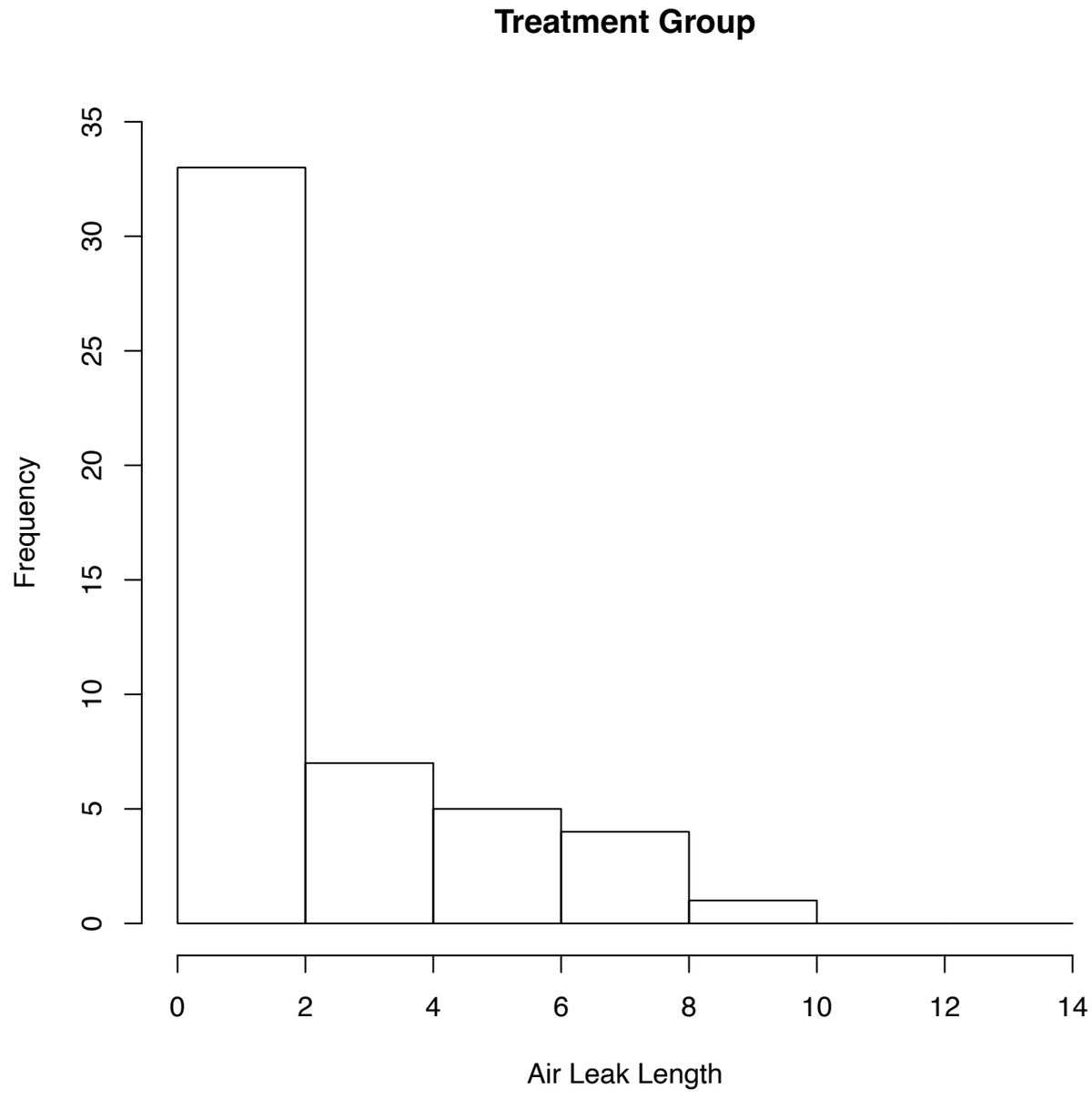
Xu et al., 2017

Simulation Study: Setup

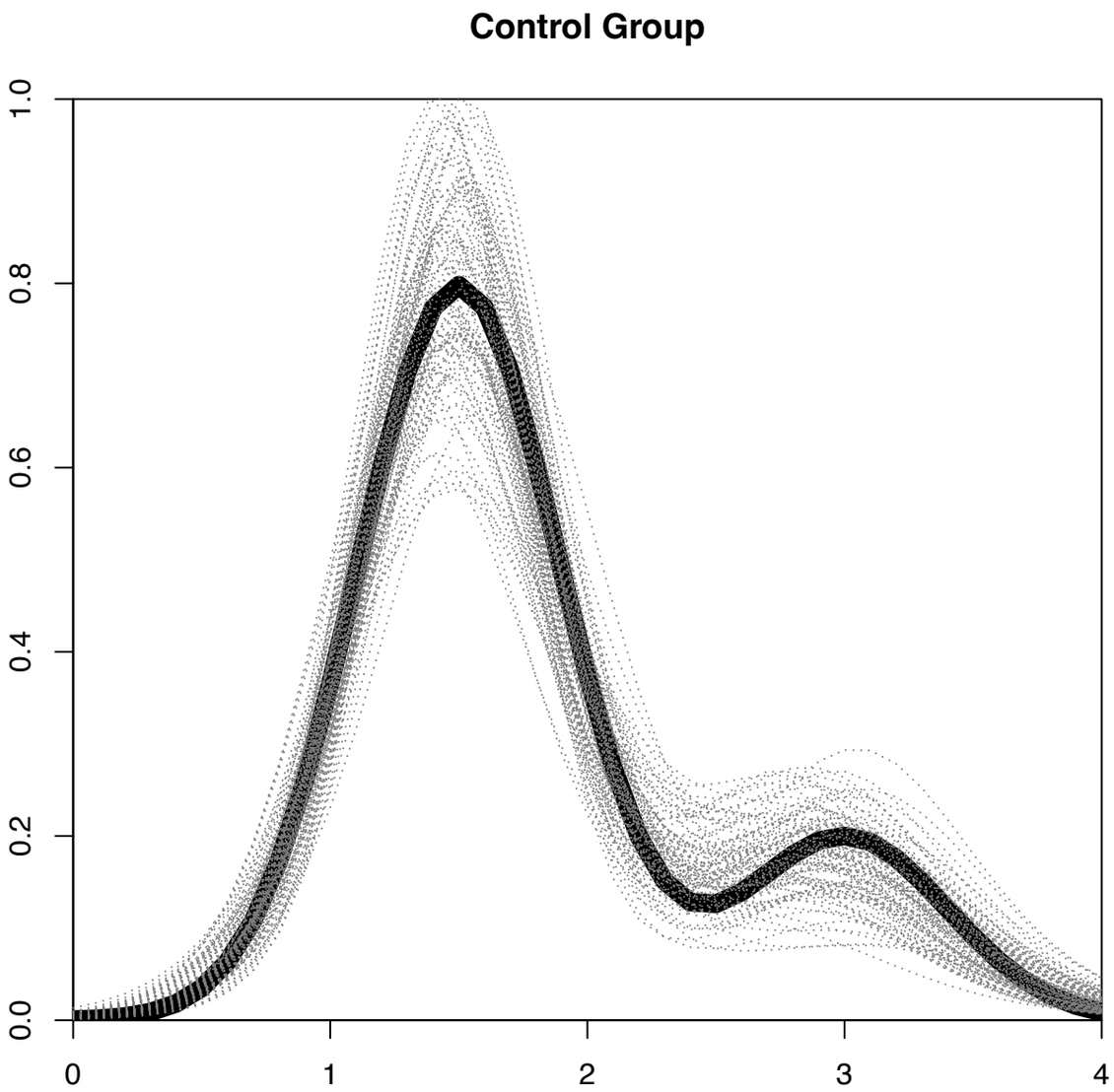
$$Y_{ji} \sim 0.8N(1.5, 0.3^2) + 0.2N(3, 0.3^2), j = 0$$

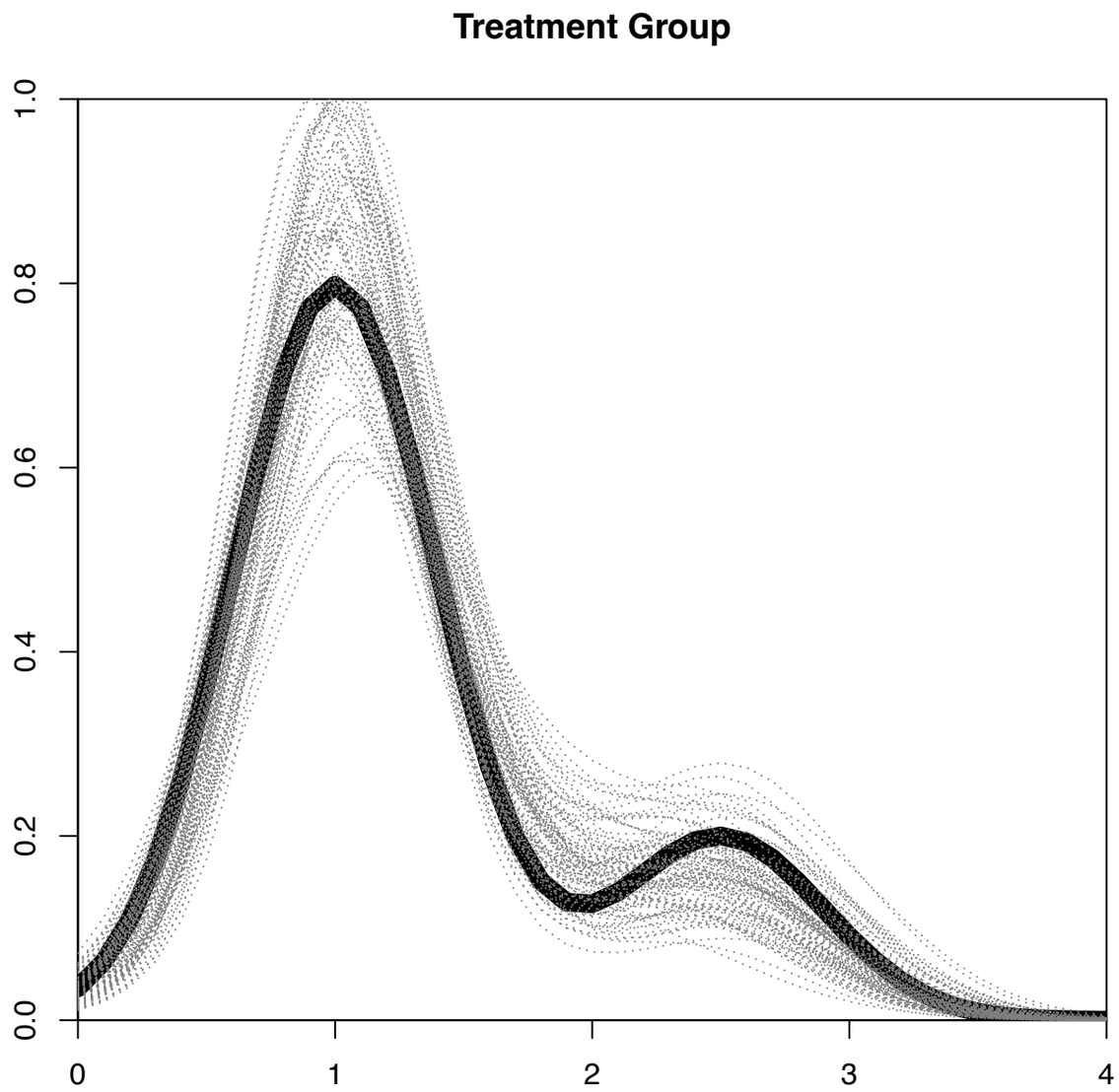
$$Y_{ji} \sim 0.8N(1, 0.3^2) + 0.2N(2.5, 0.3^2), j = 1$$





Simulation Study: Results





Take-home Messages

Take-home Messages

- BNP utility-based trial designs.

Take-home Messages

- BNP utility-based trial designs.
- The utility function is only meaningful if the probability model allows learning about detailed features of the event time distribution, and the nonparametric model is only needed when the decision hinges on such details.

Take-home Messages

- BNP utility-based trial designs.
- The utility function is only meaningful if the probability model allows learning about detailed features of the event time distribution, and the nonparametric model is only needed when the decision hinges on such details.
- Novelties

Take-home Messages

- BNP utility-based trial designs.
- The utility function is only meaningful if the probability model allows learning about detailed features of the event time distribution, and the nonparametric model is only needed when the decision hinges on such details.
- Novelties
  - a small-scale trial design (n=48).

Take-home Messages

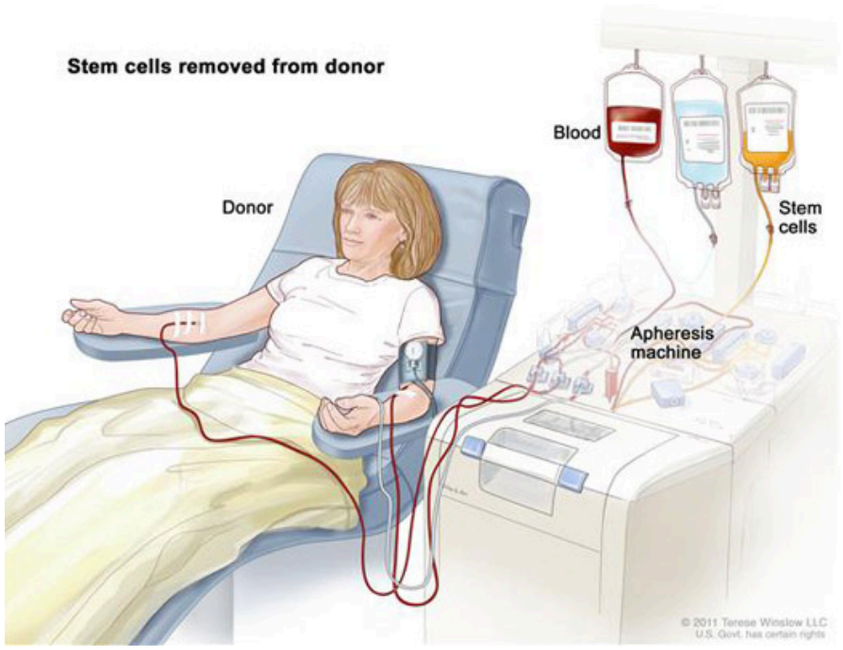
- BNP utility-based trial designs.
- The utility function is only meaningful if the probability model allows learning about detailed features of the event time distribution, and the nonparametric model is only needed when the decision hinges on such details.
- Novelties
  - a small-scale trial design (n=48).
  - a convincing case for the need of a full probabilistic description of uncertainties on random probability measures

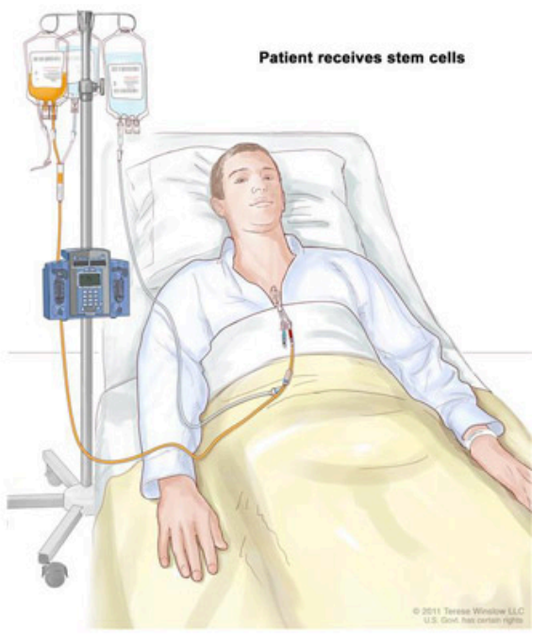
Outline

- **Part 1:** *Monday*
  - **Density estimation** for efficient clinical trial designs
  - **Regression** for precision dosing
- **Part 2:** *Wednesday*
  - **Clustering** for subgroup finding
  - **Latent feature models** for tumor heterogeneity
- **Part 3:** *Friday*
  - Estimating treatment effects from observational data

Background

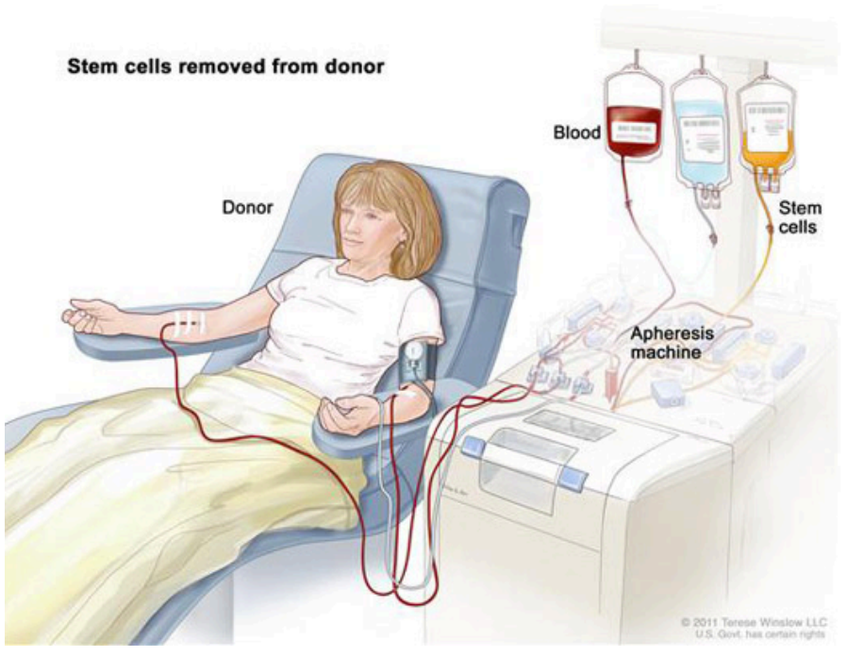
Allogeneic Stem Cell Transplantation (alloSCT) is an aggressive therapy for various hematologic diseases, such as lymphocytic leukemia and non-Hodgkins lymphoma.

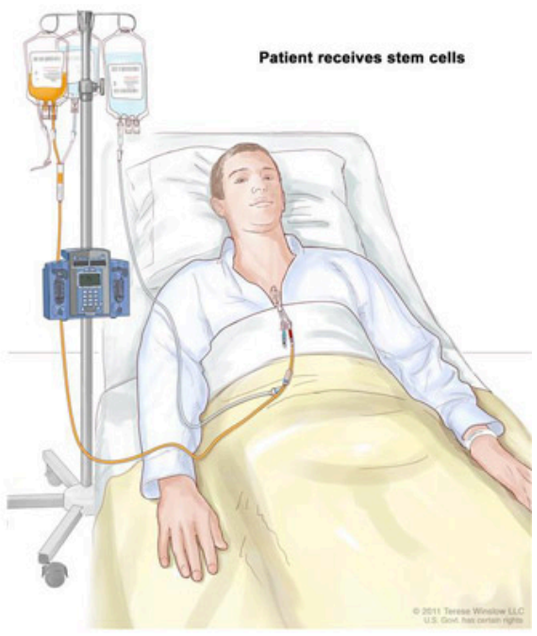




Background

Allogeneic Stem Cell Transplantation (alloSCT) is an aggressive therapy for various hematologic diseases, such as lymphocytic leukemia and non-Hodgkins lymphoma.





Busulfan

Background

Systemic busulfan exposure, characterized by area under the plasma concentration curve, AUC = the delivered dose, is strongly associated with clinical outcome.

- AUC too high $\Rightarrow$ High risks of severe toxicity.
- AUC too low $\Rightarrow$ High risks of graft failure and disease recurrence.

Background

Systemic busulfan exposure, characterized by area under the plasma concentration curve, AUC = the delivered dose, is strongly associated with clinical outcome.

- AUC too high ⇒ High risks of severe toxicity.
- AUC too low ⇒ High risks of graft failure and disease recurrence.

- Earlier practice: busulfan orally, resulting in 10 to 20 times the variability in AUC. Hard to control

Background

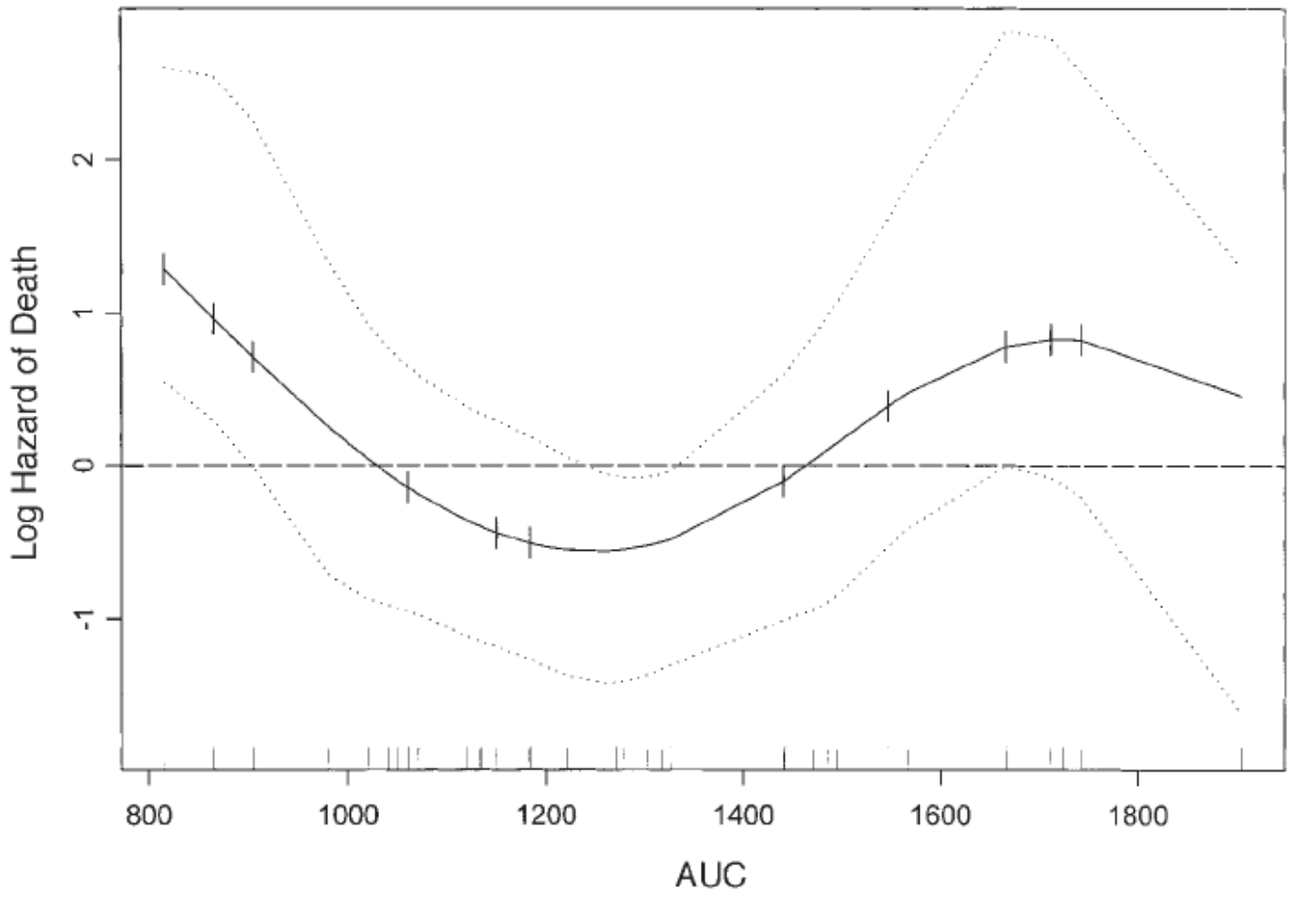
Systemic busulfan exposure, characterized by area under the plasma concentration curve, AUC = the delivered dose, is strongly associated with clinical outcome.

- AUC too high ⇒ High risks of severe toxicity.
- AUC too low ⇒ High risks of graft failure and disease recurrence.

- Earlier practice: busulfan orally, resulting in 10 to 20 times the variability in AUC. Hard to control
- Current practice: busulfan intravenous (IV), improving its bioavailability and delivered dosing accuracy

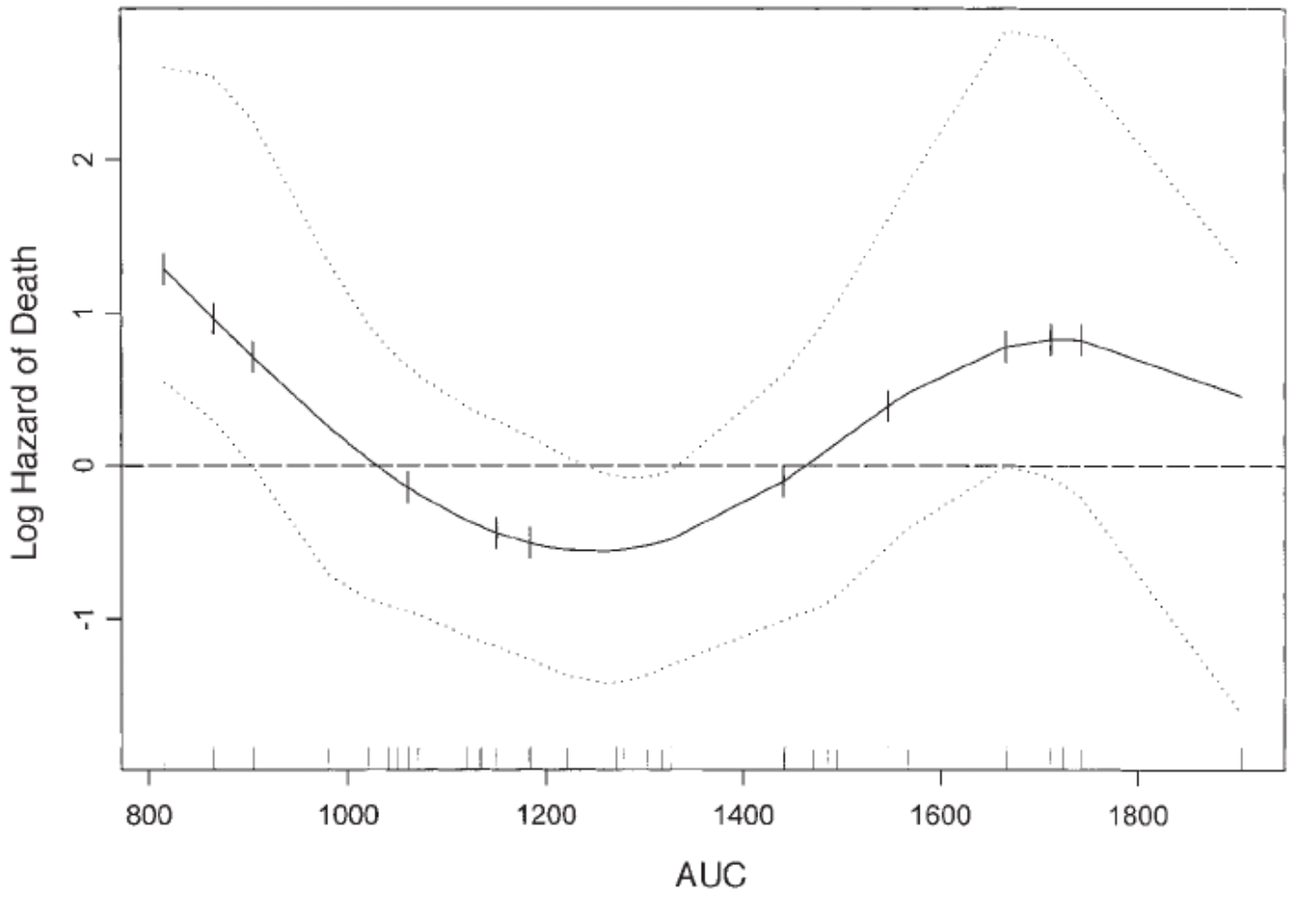
Background

Cox proportional hazard model (details later):



Background

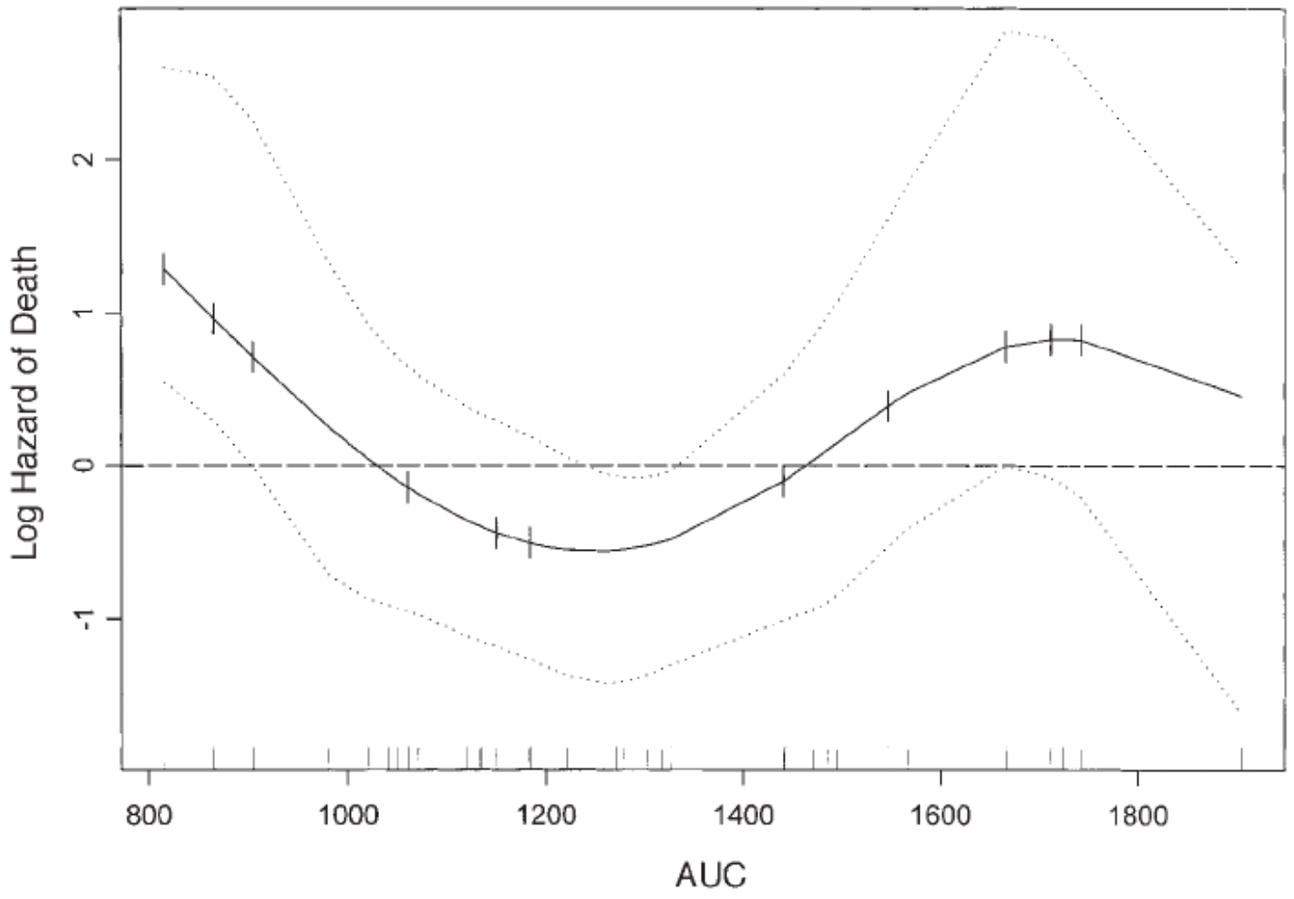
Cox proportional hazard model (details later):



Andersson et al. (2002) estimated an optimal AUC range (950 to 1520 μMol-min)

Background

Cox proportional hazard model (details later):



Andersson et al. (2002) estimated an optimal AUC range (950 to 1520 μMol-min)

Bartelink et al. (2016) reported an optimal AUC range (19100 to 21200 μMol-min) when treating children and young adults.

Personalizing Targeted IV Busulfan Dose

Age and **disease status at alloSCT** (CR = Complete Remission, No CR = Active Disease) are strongly predictive of T = survival time.

Personalizing Targeted IV Busulfan Dose

Age and **disease status at alloSCT** (CR = Complete Remission, No CR = Active Disease) are strongly predictive of T = survival time.

Can the optimal AUC interval be personalized using (Age, CR) to maximize $E(T \mid \text{Age}, \text{CR}, \text{AUC})$?

Personalizing Targeted IV Busulfan Dose

Age and **disease status at alloSCT** (CR = Complete Remission, No CR = Active Disease) are strongly predictive of T = survival time.

Can the optimal AUC interval be personalized using (Age, CR) to maximize $E(T \mid \text{Age}, \text{CR}, \text{AUC})$?

Historical Data: 151 alloSCT patients who received a standard 4-day preparative regimen of IV busulfan.

Challenges and Potential Benefits

Challenges in this Statistical Analysis

Challenges and Potential Benefits

Challenges in this Statistical Analysis

- Modeling $p(T \mid \text{Age}, \text{CR}, \text{AUC})$ robustly.

Challenges and Potential Benefits

Challenges in this Statistical Analysis

- Modeling $p(T \mid \text{Age}, \text{CR}, \text{AUC})$ robustly.
- Identifying and characterizing possibly nonlinear [Age × AUC] or [CR ×AUC] or [Age ×CR ×AUC] interactive effects on T.

Challenges and Potential Benefits

Challenges in this Statistical Analysis

- Modeling $p(T \mid \text{Age}, \text{CR}, \text{AUC})$ robustly.
- Identifying and characterizing possibly nonlinear [Age × AUC] or [CR ×AUC] or [Age ×CR ×AUC] interactive effects on T.
- The historical dataset has only 151 patients.

Challenges and Potential Benefits

Challenges in this Statistical Analysis

- Modeling $p(T \mid \text{Age}, \text{CR}, \text{AUC})$ robustly.
- Identifying and characterizing possibly nonlinear [Age × AUC] or [CR ×AUC] or [Age ×CR ×AUC] interactive effects on T.
- The historical dataset has only 151 patients.

Potential Payoff of this Statistical Analysis

Challenges and Potential Benefits

Challenges in this Statistical Analysis

- Modeling $p(T \mid \text{Age}, \text{CR}, \text{AUC})$ robustly.
- Identifying and characterizing possibly nonlinear [Age × AUC] or [CR ×AUC] or [Age ×CR ×AUC] interactive effects on T.
- The historical dataset has only 151 patients.

Potential Payoff of this Statistical Analysis

If optimal AUC intervals based on (Age, CR) can be estimated, $E(T \mid \text{Age}, \text{CR}, \text{AUC})$ can be increased for future patients by making this personalized IV busulfan dosing in alloSCT standard clinical practice.

Traditional Survival Models

Event time data, usually involving censoring.

- Survival time: T
- Censoring time: C
- Censoring indicator: $\delta = I(T \leq C)$
- Observed data: (Y, C, δ), where $Y = min(T, C)$

Traditional Survival Models

- **Survival function**: $S(t) = Pr(T > t)$.

Traditional Survival Models

- **Survival function**: $S(t) = Pr(T > t)$.
- **Density**: $f(t) = -\frac{d}{dt}[1 - F(t)] = -\frac{d}{dt}S(t)$.

Traditional Survival Models

- **Survival function**: $S(t) = Pr(T > t)$.
- **Density**: $f(t) = -\frac{d}{dt}[1 - F(t)] = -\frac{d}{dt}S(t)$.
- **Hazard function**:

$$\lambda(t) = lim_{dt \to 0} \frac{Pr(t \leq T < t + dt)}{dtS(t)} = \frac{f(t)}{S(t)} .$$

Traditional Survival Models

- **Survival function**: $S(t) = Pr(T > t)$.
- **Density**: $f(t) = -\frac{d}{dt}[1 - F(t)] = -\frac{d}{dt}S(t)$.
- **Hazard function**:

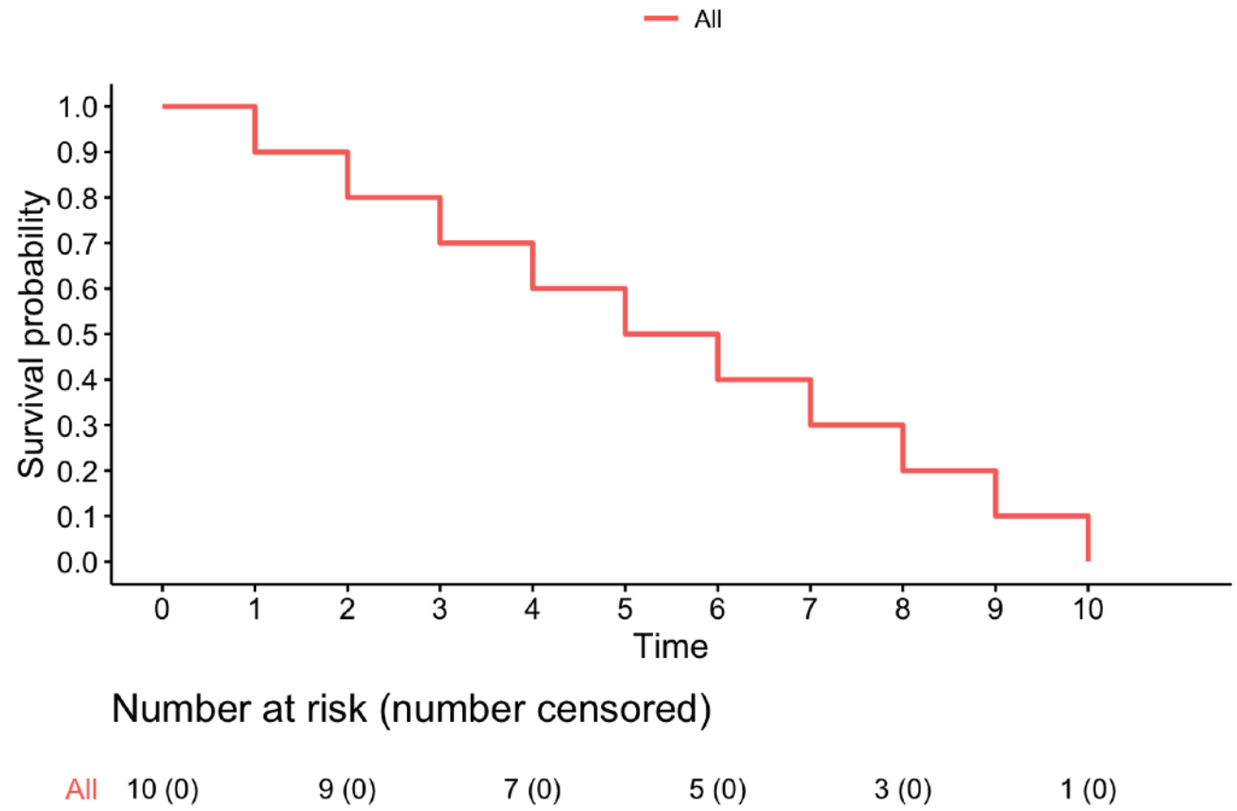
$$\lambda(t) = lim_{dt \to 0} \frac{Pr(t \leq T < t + dt)}{dtS(t)} = \frac{f(t)}{S(t)} .$$

- **Cumulative hazard function**:

$$\Lambda(t) = \int_0^t \lambda(u)du = -\log S(t) .$$

Traditional Survival Models

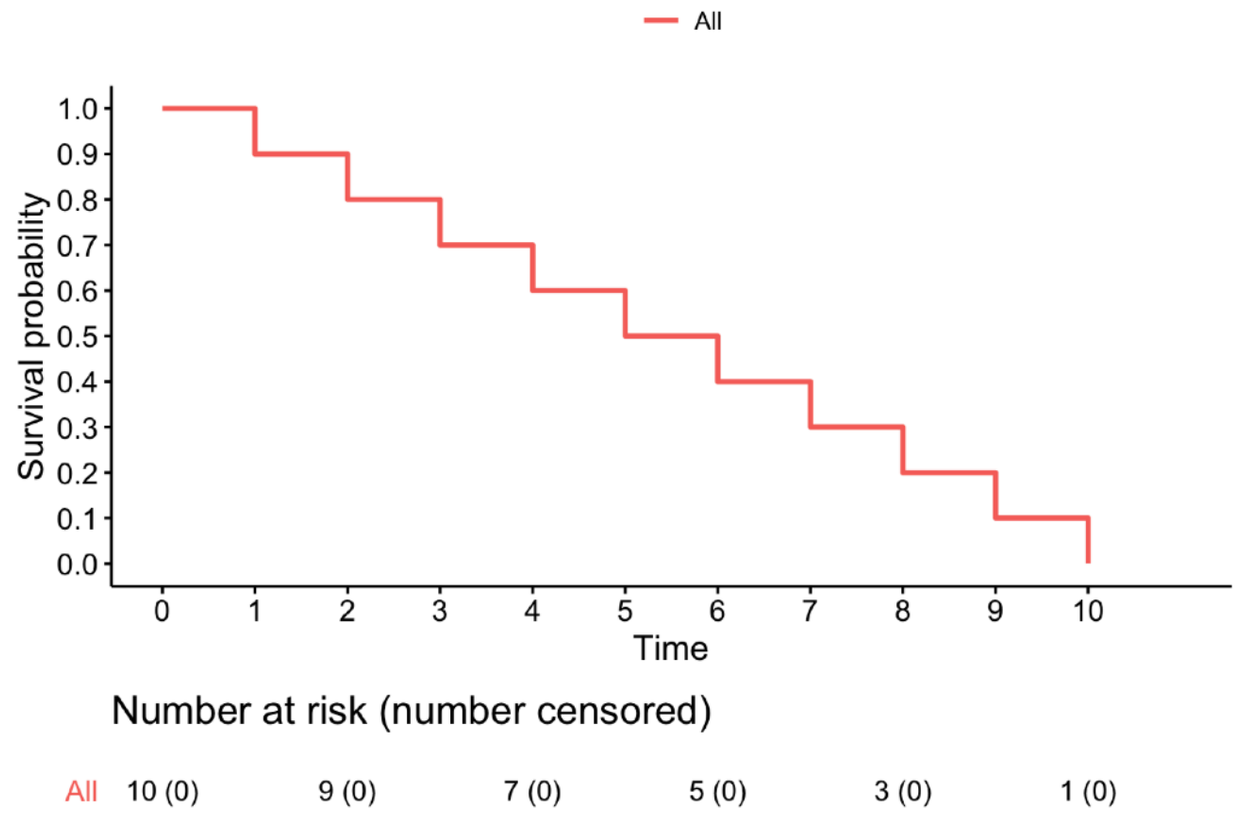
Kaplan–Meier estimator



Traditional Survival Models

Kaplan–Meier estimator

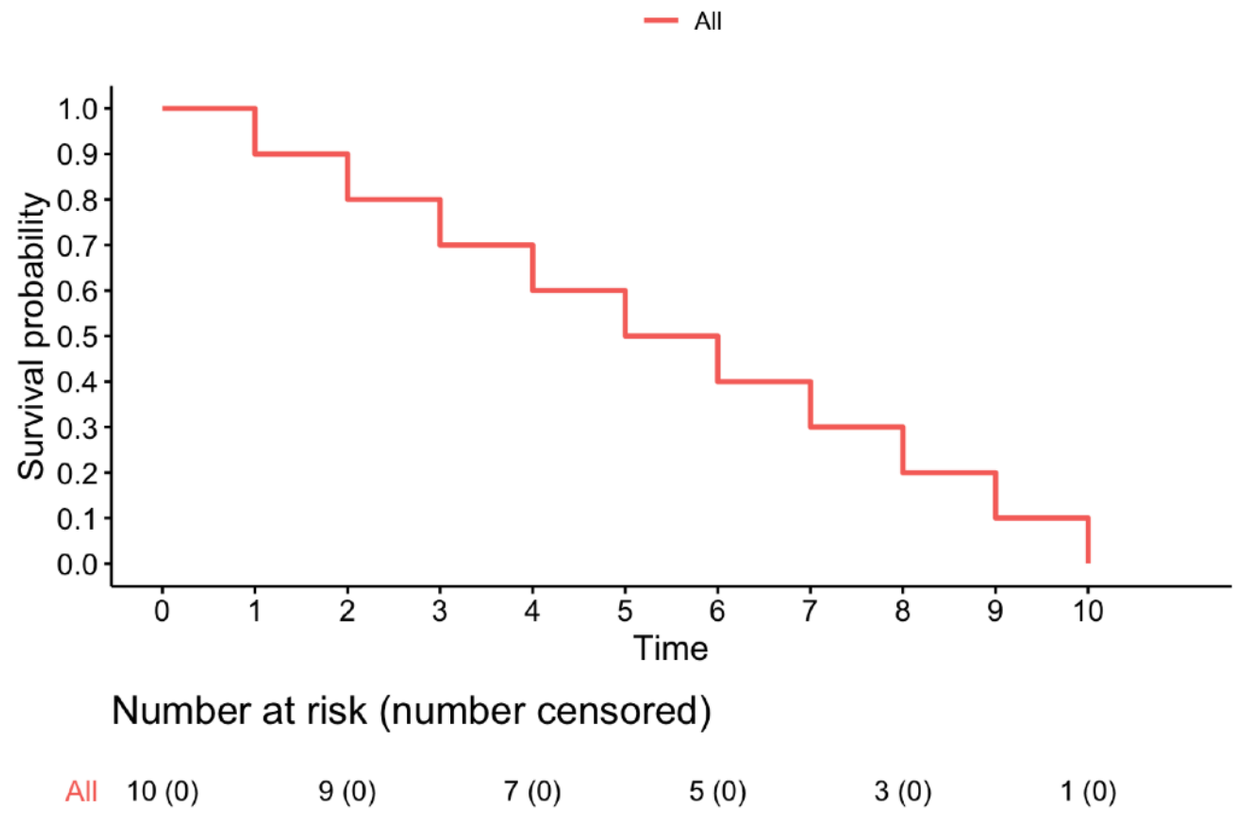
$$\widehat{S}(t) = \prod_{i:\, t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$



Traditional Survival Models

Kaplan–Meier estimator

$$\widehat{S}(t) = \prod_{i:\, t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$



t_i: time when at least one event happened

d_i: the number of events (e.g., death) that happened at time t_i

n_i: the *individuals known to have survived* (have not yet had an event or been censored) up to time t_i

Traditional Survival Models

- Cox Proportional Hazards (PH) model:

$$S_x(t) = S_0(t)^{\exp(x'\beta)}$$

Traditional Survival Models

- Cox Proportional Hazards (PH) model:

$$S_x(t) = S_0(t)^{\exp(x'\beta)}$$

- In terms of hazards, this model reduces to

$$h_x(t) = h_0(t)\exp(x'\beta)$$

Traditional Survival Models

- Cox Proportional Hazards (PH) model:

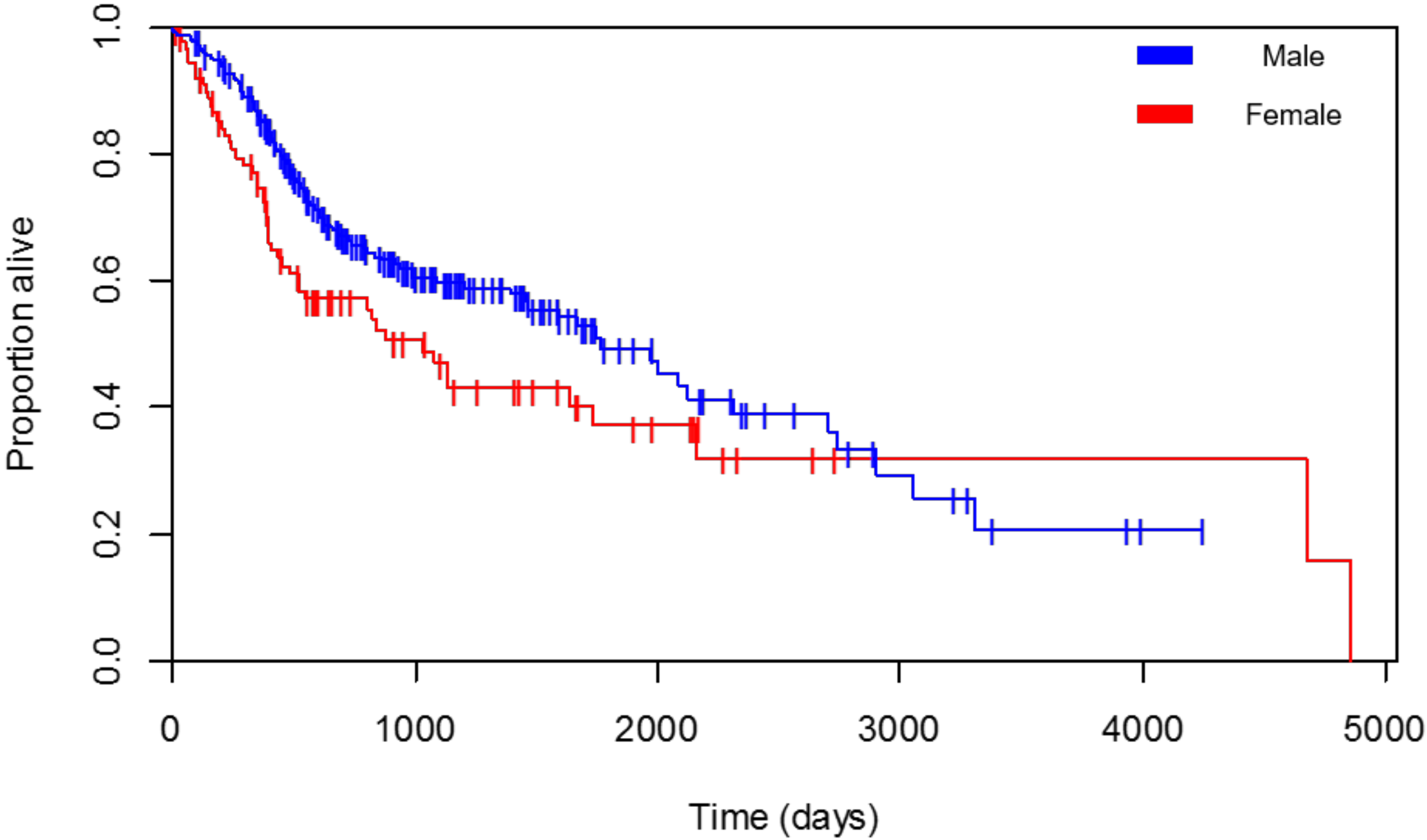
$$S_x(t) = S_0(t)^{\exp(x'\beta)}$$

- In terms of hazards, this model reduces to

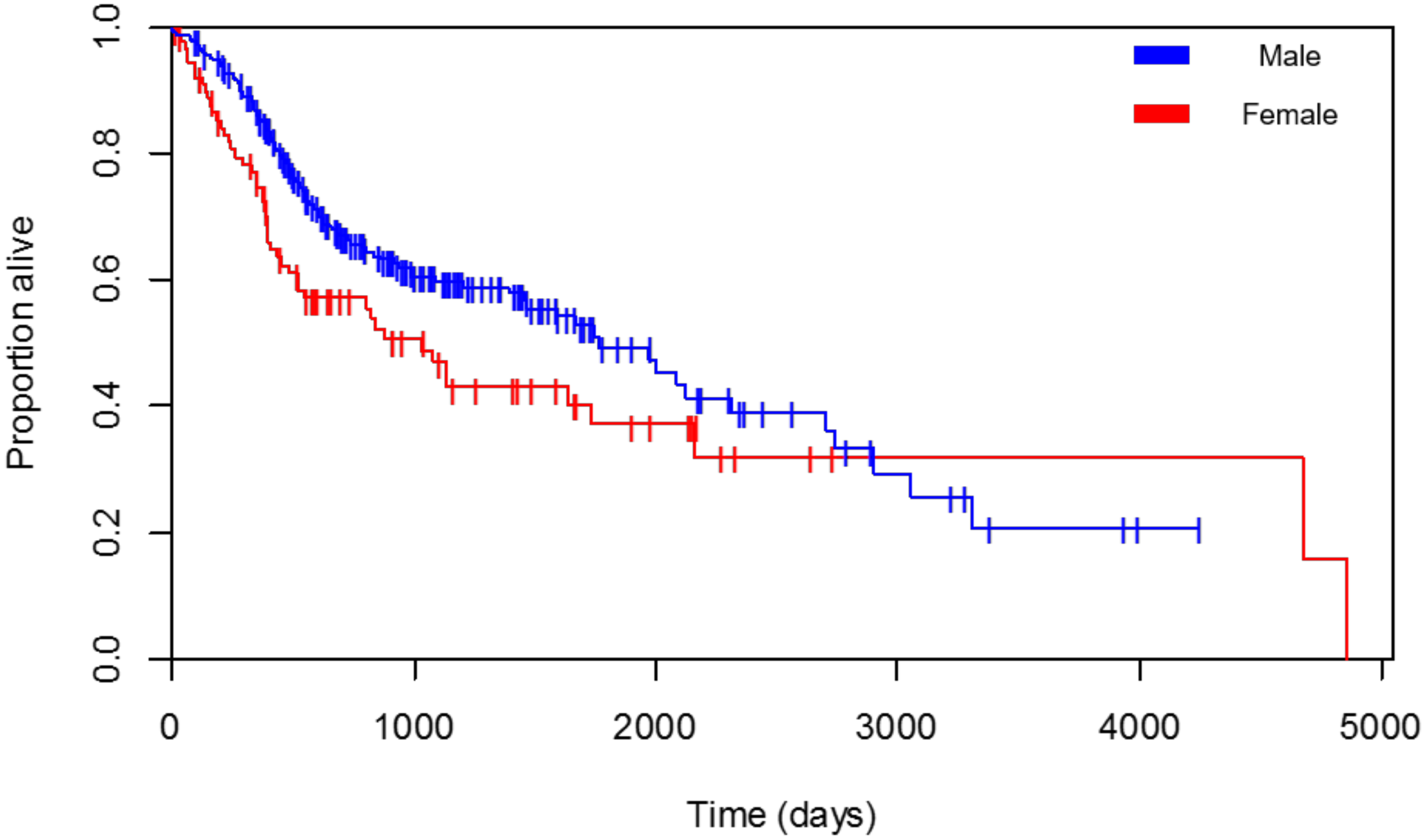
$$h_x(t) = h_0(t)\exp(x'\beta)$$

Note then that for two individuals with covariates x_1 and x_2, the ratio of hazard curves is constant, equal to $\exp((x_1 - x_2)'\beta)$, hence the name "proportional hazards."

Traditional Survival Models



Traditional Survival Models



Cox Proportional Hazards (PH) model is not appropriate!

Traditional Survival Models

- Accelerated Failure Time (AFT) model:

$$S_x(t) = S_0\{\exp(-x'\beta)t\}$$

Traditional Survival Models

- Accelerated Failure Time (AFT) model:

$$S_x(t) = S_0\{\exp(-x'\beta)t\}$$

- This is equivalent to a linear model for the log time-to-event T,

$$\log(T) = x'\beta + \epsilon, \text{where } p(\epsilon > \log t) = S_0(t)\,.$$

Dependent DP (DDP)

$$F(y \mid X) = \sum_{h=1}^{\infty} w_h N(y; \theta_h(X), \sigma^2)$$

Dependent DP (DDP)

$$F(y \mid X) = \sum_{h=1}^{\infty} w_h N(y; \theta_h(X), \sigma^2)$$

Dependent DP (DDP)

$$F(y \mid X) = \sum_{h=1}^{\infty} w_h N(y; \theta_h(X), \sigma^2)$$

- Include regression on covariates by assuming $\theta_h(X) = X\beta_h$.

Dependent DP (DDP)

$$F(y \mid X) = \sum_{h=1}^{\infty} w_h N(y; \theta_h(X), \sigma^2)$$

- Include regression on covariates by assuming $\theta_h(X) = X\beta_h$.
- A Gaussian process (GP) prior for $\theta_h(X)$ gives the DDP-GP. $\theta_h(X) \sim GP(\mu_h, C)$ with $\mu_h(X_i; \beta_h) = X_i\beta_h$ for $i = 1,\ldots,n$ and $h = 1,2,\ldots,$ with variance-covariance matrix

Dependent DP (DDP)

$$F(y \mid X) = \sum_{h=1}^{\infty} w_h N(y; \theta_h(X), \sigma^2)$$

- Include regression on covariates by assuming $\theta_h(X) = X\beta_h$.
- A Gaussian process (GP) prior for $\theta_h(X)$ gives the DDP-GP. $\theta_h(X) \sim GP(\mu_h, C)$ with $\mu_h(X_i; \beta_h) = X_i\beta_h$ for $i = 1,\ldots,n$ and $h = 1,2,\ldots,$ with variance-covariance matrix

$$C(\boldsymbol{Z}_i, \boldsymbol{Z}_\ell) = \sigma_0^2 \ \exp\left\{ -\sum_{d=1}^{D} \frac{(Z_{id} - Z_{\ell d})^2}{\lambda_d^2} \right\} + \delta_{i\ell} J^2.$$

Model Summary

Model Summary

Data: Y_i = log survival time, $X_i = (Age_i, CR_i, AUC_i)$ of patient $i = 1,\ldots,151$

Model Summary

Data: $Y_i = \log$ survival time, $X_i = (Age_i, CR_i, AUC_i)$ of patient $i = 1,\ldots,151$

Distribution: $p(y_i \mid X_i, F) = F_{X_i}(y_i)$

Model Summary

Data: $Y_i = \log$ survival time, $X_i = (Age_i, CR_i, AUC_i)$ of patient $i = 1,\ldots,151$

Distribution: $p(y_i \mid X_i, F) = F_{X_i}(y_i)$

Prior: $F_X \sim \text{DDP} - \text{GP}(\{\mu_h\}, C, \alpha, \{\beta_h\}, \{\lambda_d\}, \sigma_0^2, \sigma^2)$

Model Summary

Data: $Y_i = \log$ survival time, $X_i = (Age_i, CR_i, AUC_i)$ of patient $i = 1,\ldots,151$

Distribution: $p(y_i \mid X_i, F) = F_{X_i}(y_i)$

Prior: $F_X \sim \text{DDP} - \text{GP}(\{\mu_h\}, C, \alpha, \{\beta_h\}, \{\lambda_d\}, \sigma_0^2, \sigma^2)$

For right-censored survival data $D_n = \{Y_i, \delta_i, X_i\}_{i=1}^n$, the likelihood function has the usual form

$$L(\theta \mid D_n) = \prod_{i=1}^{n} \{f_{X_i}(Y_i \mid \theta)\}^{\delta_i} \{1 - F_{X_i}(Y_i \mid \theta)\}^{1-\delta_i}$$

Estimated **Optimal Targeted Intervals** of IV Busulfan AUC
Personalized For Given (CR Status, Age)

We define the predicted optimal IV busulfan targeted AUC for future patient $n+1$ with covariates $\boldsymbol{X}$ = (CR Status, Age, AUC) as

$$\widehat{AUC}_{n+1} = \text{argmax}_{AUC}\ E(Y_{n+1} \mid \boldsymbol{X}, \mathcal{D}_n)$$

Since the laboratory error in evaluation of AUC is up to about 6%, the optimal AUC interval for future patient $n+1$ is defined as

$$[\ 0.9\ \widehat{AUC}_{n+1},\quad 1.1\ \widehat{AUC}_{n+1}\ ]$$

Simulation Study: Setup

- Age: x_1 and AUC: x_2 were sampled with replacement from the actual ages and AUC values.
- CR: $x_3 \sim$ Bernoulli(0.5).
- Survival: $T \sim LN(\mu(\boldsymbol{x}_i), \sigma_0^2)$, where $\sigma_0 = 0.4$, and

$$\begin{aligned}\mu(\boldsymbol{x}_i) &= 4 - 0.1x_{i,1} + 0.7x_{i,2} + 0.3x_{i,3} - 0.07x_{i,2}^2 - 0.1x_{i,1}x_{i,2} \\ &\quad +0.2x_{i,2}x_{i,3} - 0.18x_{i,1}x_{i,2}x_{i,3}\end{aligned}$$

- Two scenarios: $n = 200$ observations without censoring and $n = 200$ with 25% censoring.

Simulation Study: Setup

- Age: x_1 and AUC: x_2 were sampled with replacement from the actual ages and AUC values.
- CR: $x_3 \sim$ Bernoulli(0.5).
- Survival: $T \sim LN(\mu(\boldsymbol{x}_i), \sigma_0^2)$, where $\sigma_0 = 0.4$, and

$$\begin{aligned}\mu(\boldsymbol{x}_i) &= 4 - 0.1x_{i,1} + 0.7x_{i,2} + 0.3x_{i,3} - 0.07x_{i,2}^2 + 0.1x_{i,1}x_{i,2} \\ &\quad +0.2x_{i,2}x_{i,3} - 0.18x_{i,1}x_{i,2}x_{i,3}\end{aligned}$$

- Two scenarios: $n = 200$ observations without censoring and $n = 200$ with 25% censoring.

Simulation Study: Setup

- Age: x_1 and AUC: x_2 were sampled with replacement from the actual ages and AUC values.
- CR: $x_3 \sim$ Bernoulli(0.5).
- Survival: $T \sim LN(\mu(\boldsymbol{x}_i), \sigma_0^2)$, where $\sigma_0 = 0.4$, and

$$\begin{aligned}\mu(\boldsymbol{x}_i) &= 4 - 0.1x_{i,1} + 0.7x_{i,2} + 0.3x_{i,3} - 0.07x_{i,2}^2 + 0.1x_{i,1}x_{i,2} \\ &\quad +0.2x_{i,2}x_{i,3} - 0.18x_{i,1}x_{i,2}x_{i,3}\end{aligned}$$

- Two scenarios: $n = 200$ observations without censoring and $n = 200$ with 25% censoring.

Simulation Study

Comparison

Simulation Study

Comparison

- AFT regression models using either lognormal or Weibull distributions by assuming

Simulation Study

Comparison

- AFT regression models using either lognormal or Weibull distributions by assuming

$$\begin{aligned}\log(Y_i) &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} \\ &\quad + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i.\end{aligned}$$

Simulation Study

Comparison

- AFT regression models using either lognormal or Weibull distributions by assuming

$$\begin{aligned}\log(Y_i) &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} \\ &\quad + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i.\end{aligned}$$

- Two flexible semiparametric survival methods: model the baseline survival using
  - a Polya Trees (PT) prior (Hanson and Johnson, 2002)

Simulation Study

Comparison

- AFT regression models using either lognormal or Weibull distributions by assuming

$$\begin{aligned}\log(Y_i) &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} \\ &\quad + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i.\end{aligned}$$

- Two flexible semiparametric survival methods: model the baseline survival using
  - a Polya Trees (PT) prior (Hanson and Johnson, 2002)
  - a transformed Bernstein polynomials (TBP) prior (Zhou and Hanson, 2018)

Simulation Study

Comparison

- AFT regression models using either lognormal or Weibull distributions by assuming

$$\begin{aligned}\log(Y_i) &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} \\ &\quad + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i.\end{aligned}$$

- Two flexible semiparametric survival methods: model the baseline survival using
  - a Polya Trees (PT) prior (Hanson and Johnson, 2002)
  - a transformed Bernstein polynomials (TBP) prior (Zhou and Hanson, 2018)
- Two nonparametric survival methods

Simulation Study

Comparison

- AFT regression models using either lognormal or Weibull distributions by assuming

$$\begin{aligned}\log(Y_i) &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} \\ &\quad + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i.\end{aligned}$$

- Two flexible semiparametric survival methods: model the baseline survival using
  - a Polya Trees (PT) prior (Hanson and Johnson, 2002)
  - a transformed Bernstein polynomials (TBP) prior (Zhou and Hanson, 2018)
- Two nonparametric survival methods
  - random forests (RF) (Ishwaran et al., 2008)

Simulation Study

Comparison

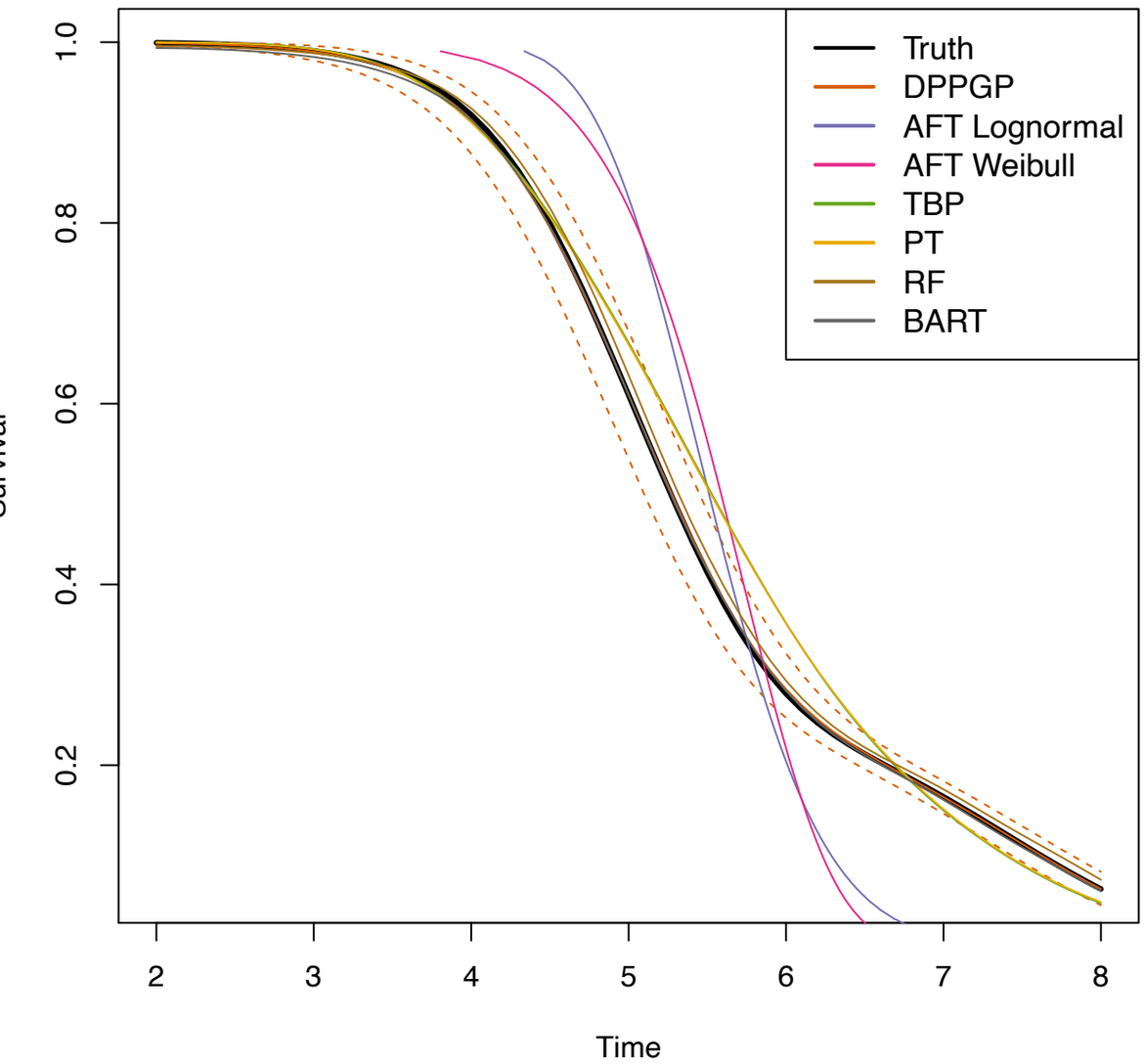
- AFT regression models using either lognormal or Weibull distributions by assuming

$$\begin{aligned}\log(Y_i) &= \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,2}^2 + \beta_5 x_{i,1} x_{i,2} \\ &\quad + \beta_6 x_{i,2} x_{i,3} + \beta_7 x_{i,1} x_{i,3} + \sigma \epsilon_i.\end{aligned}$$

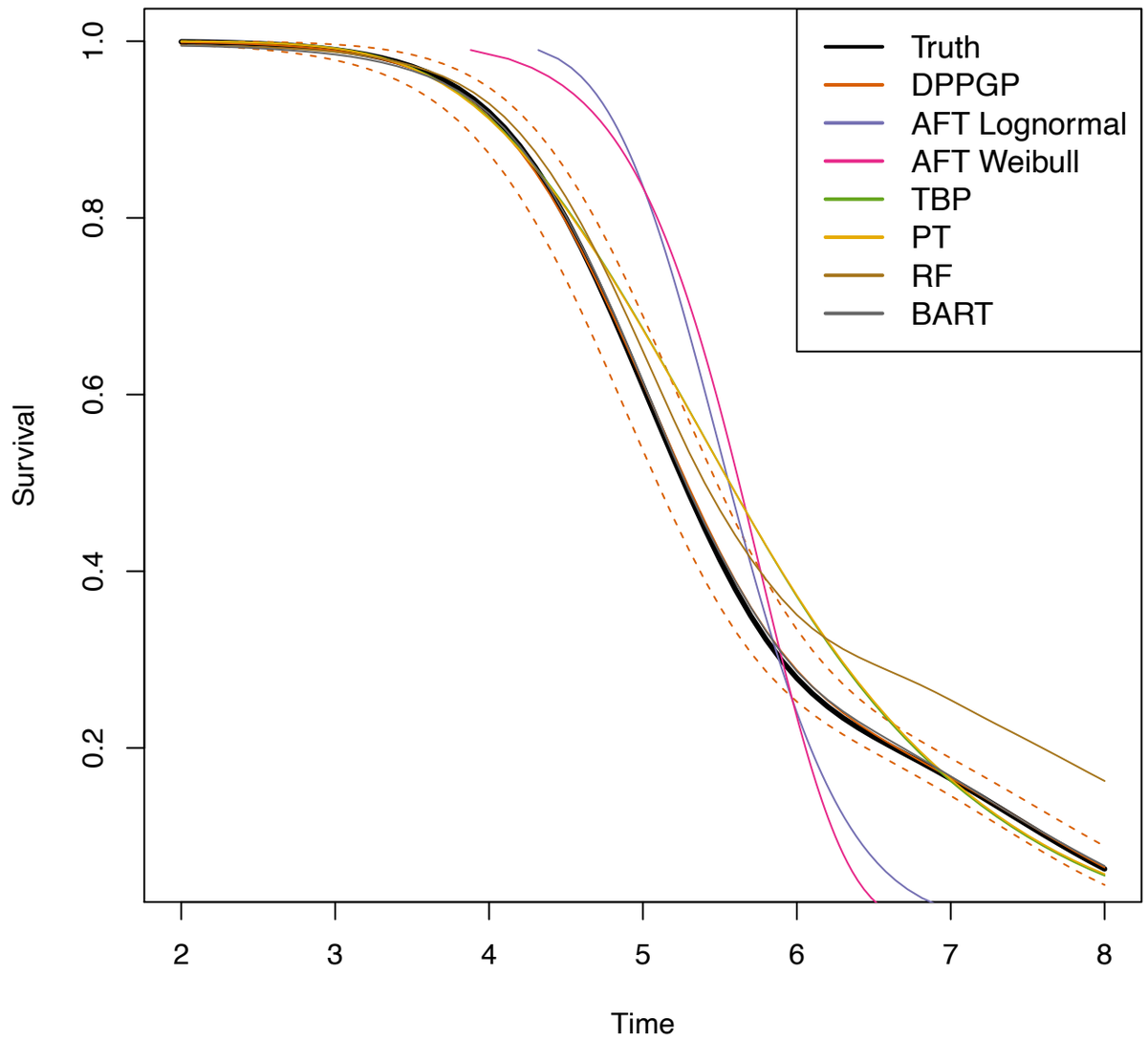
- Two flexible semiparametric survival methods: model the baseline survival using
  - a Polya Trees (PT) prior (Hanson and Johnson, 2002)
  - a transformed Bernstein polynomials (TBP) prior (Zhou and Hanson, 2018)
- Two nonparametric survival methods
  - random forests (RF) (Ishwaran et al., 2008)
  - Bayesian additive regression trees (BART) (Sparapani et al., 2016)

Simulation Study

n=200 without censoring

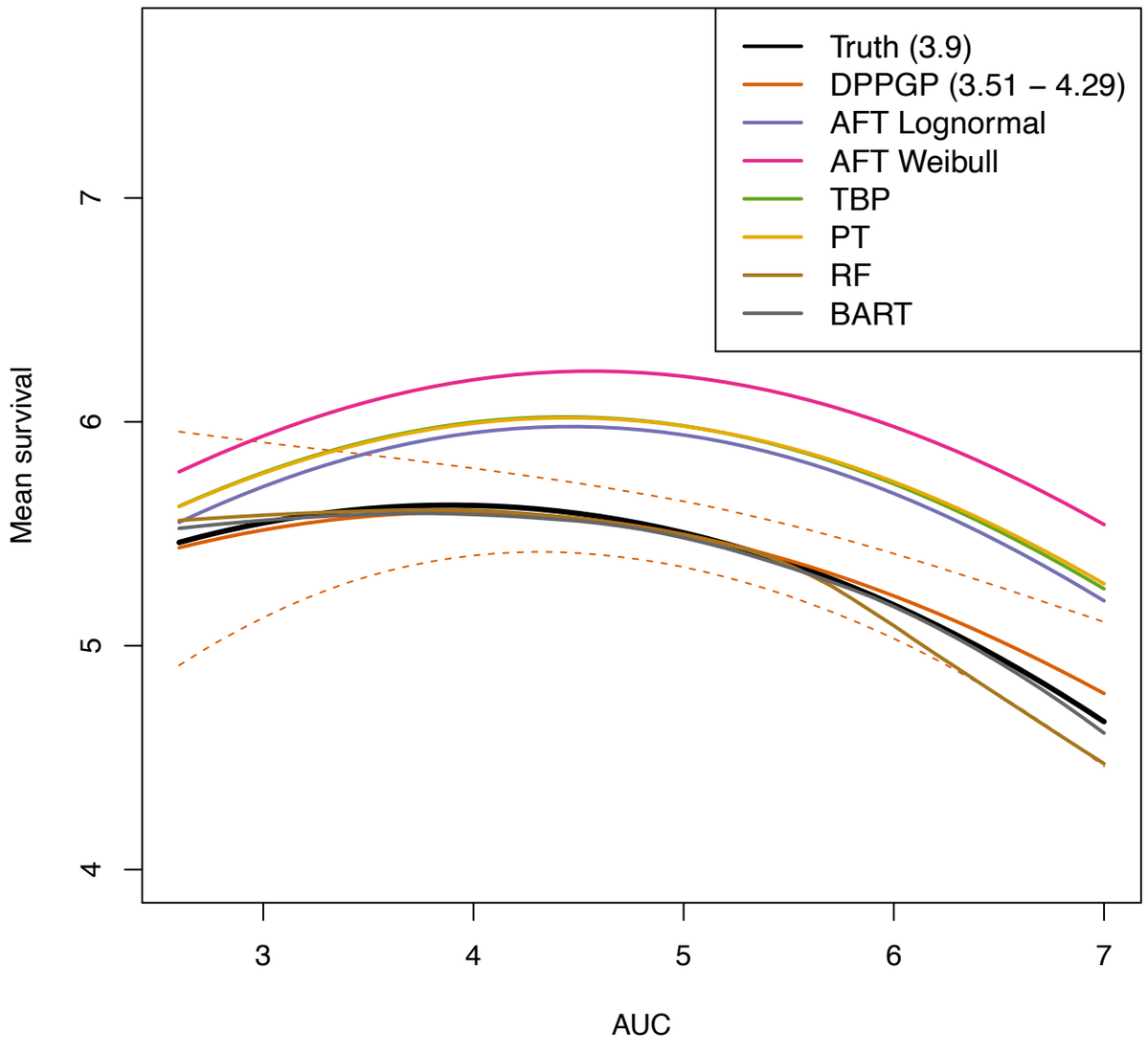


n=200 with 25% censoring

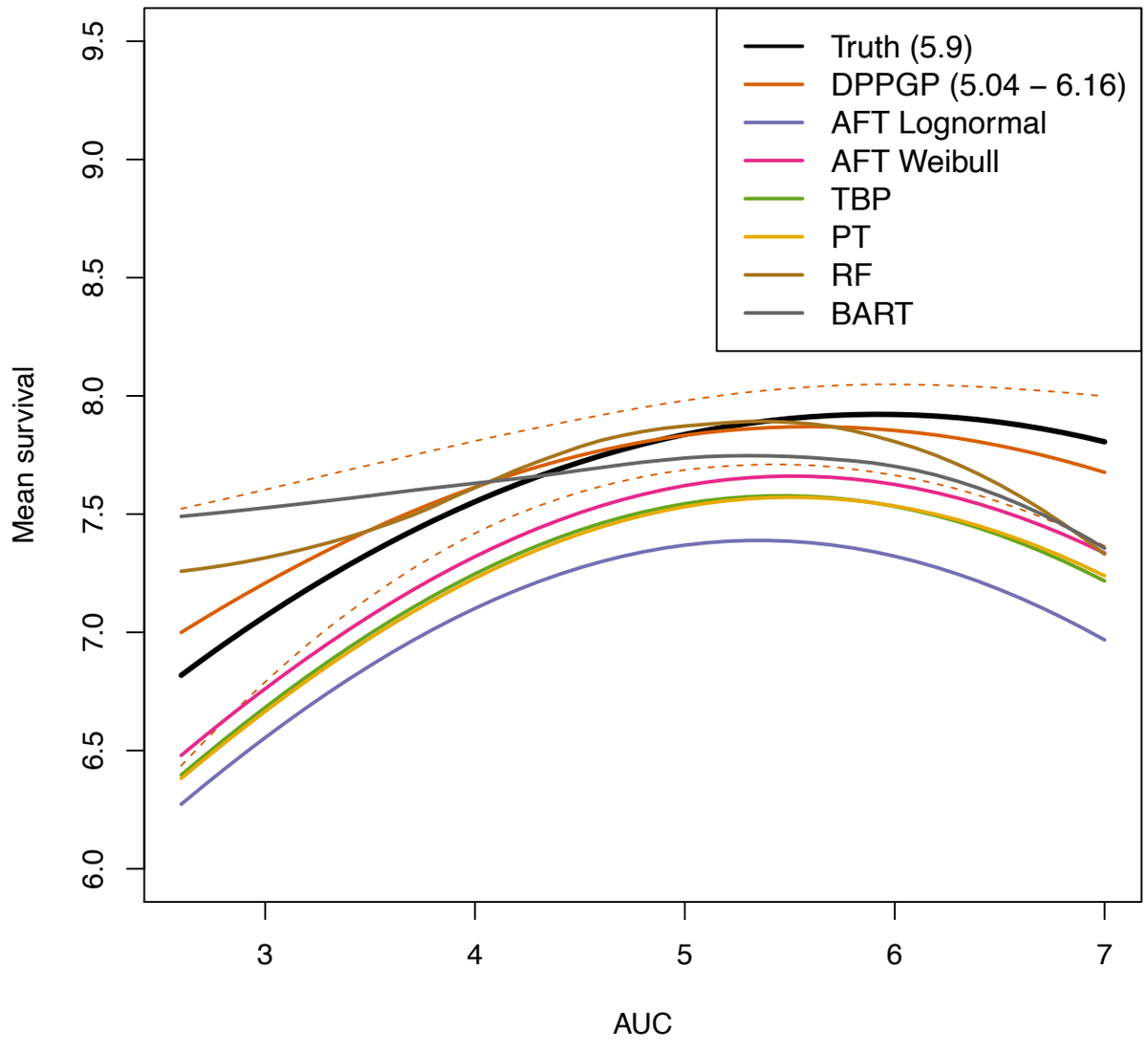


Simulation Study

n=200 without censoring



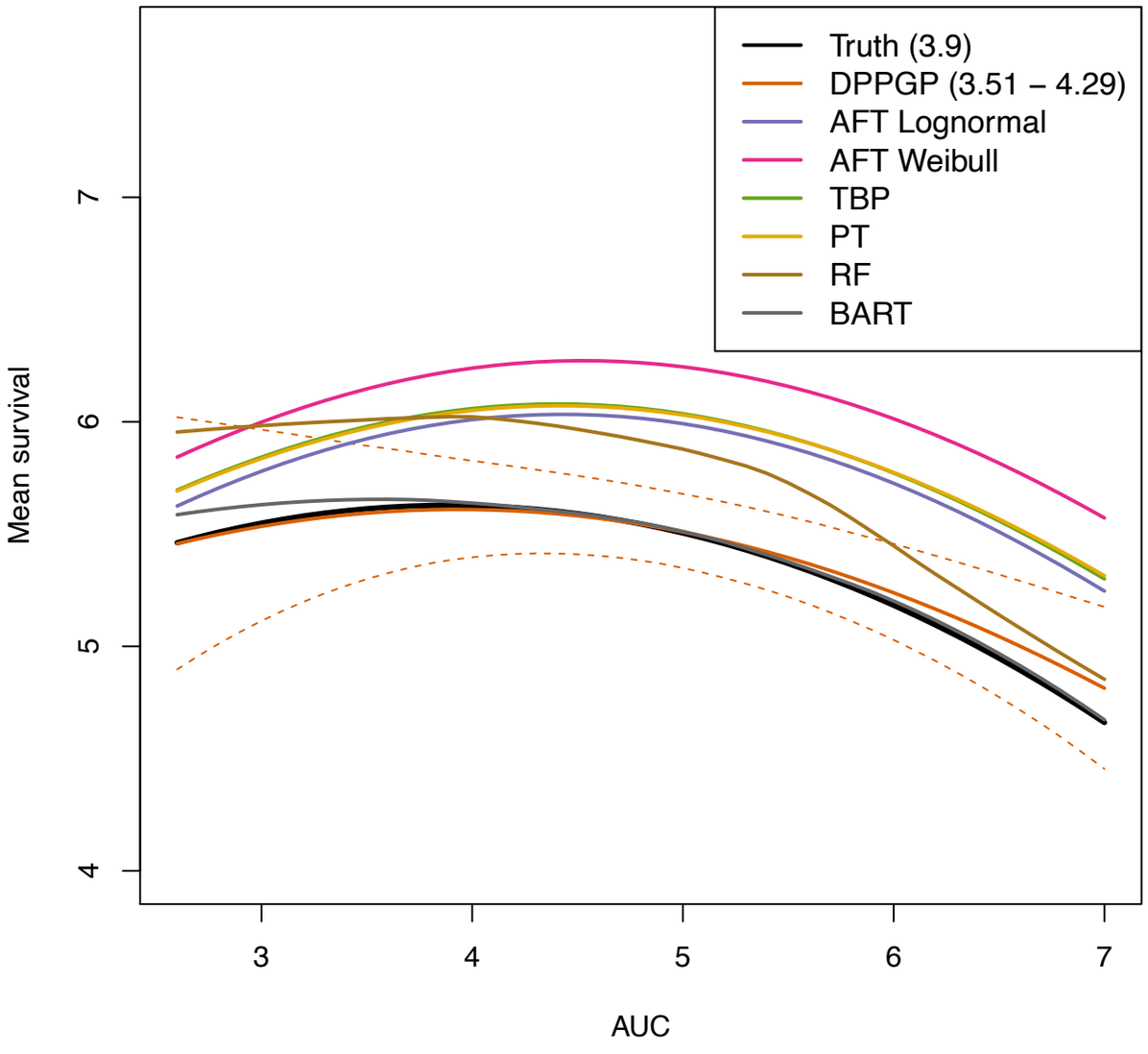
CR=0



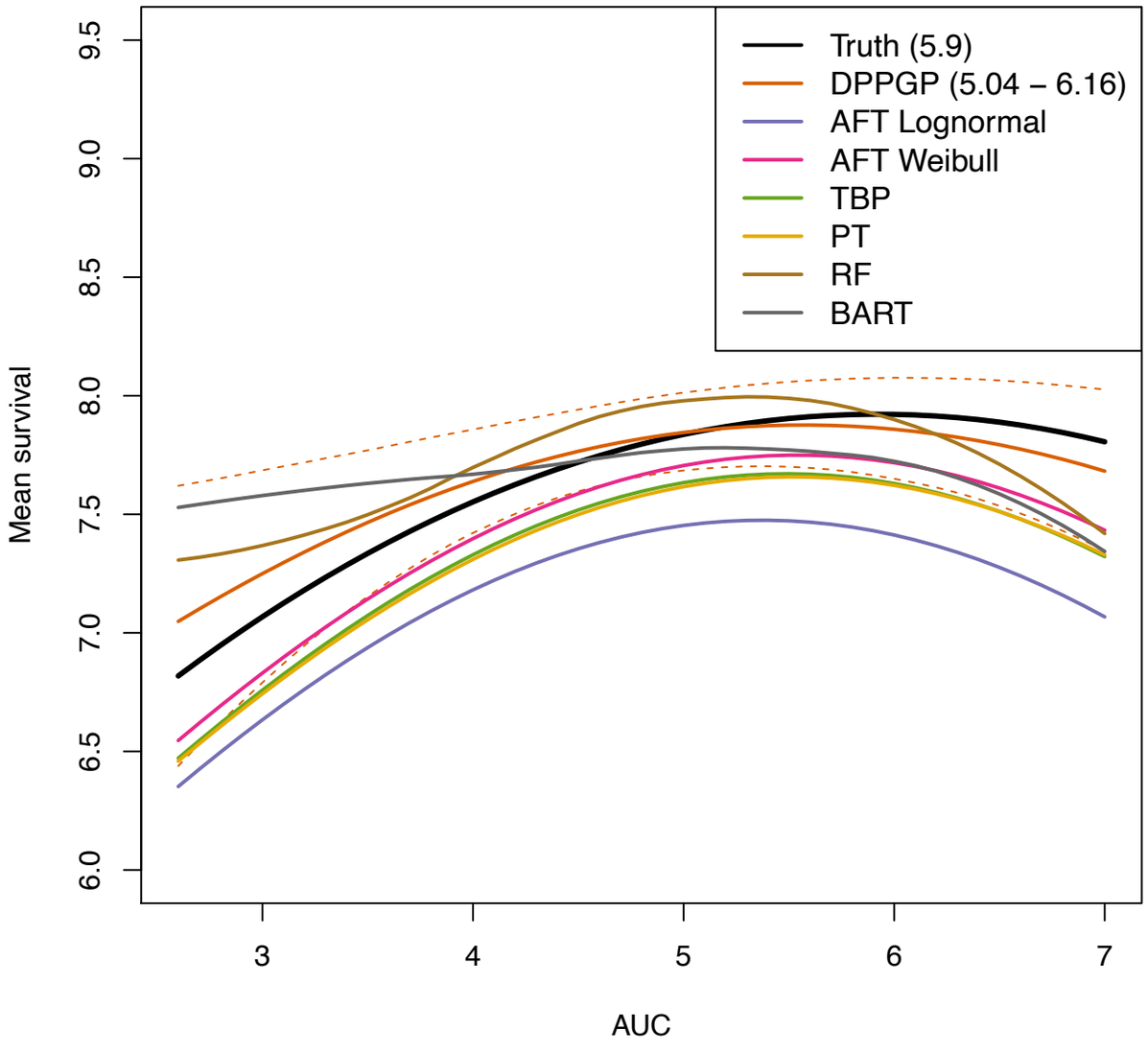
CR=1

Simulation Study

n=200 with 25% censoring

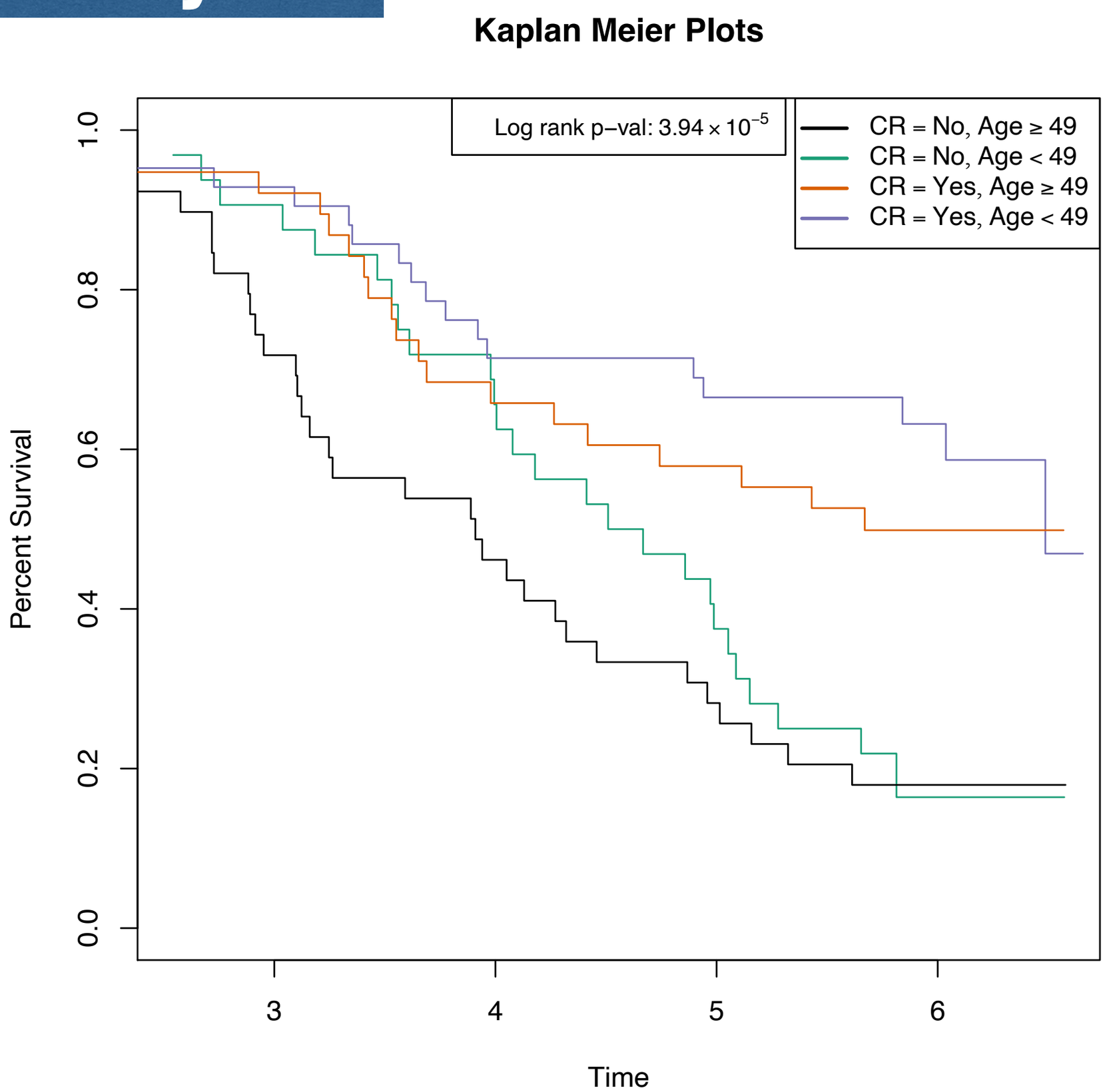


CR=0



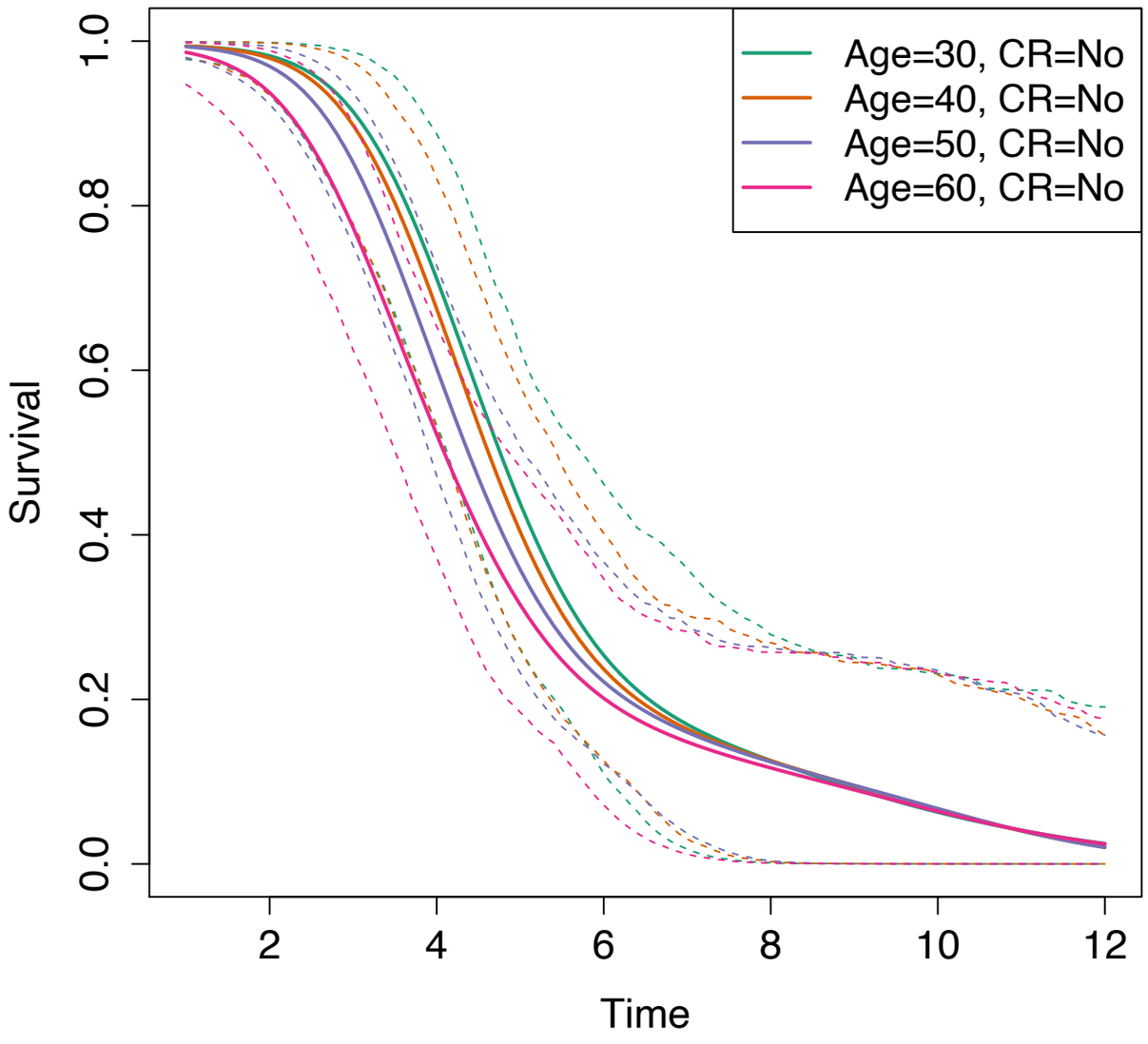
CR=1

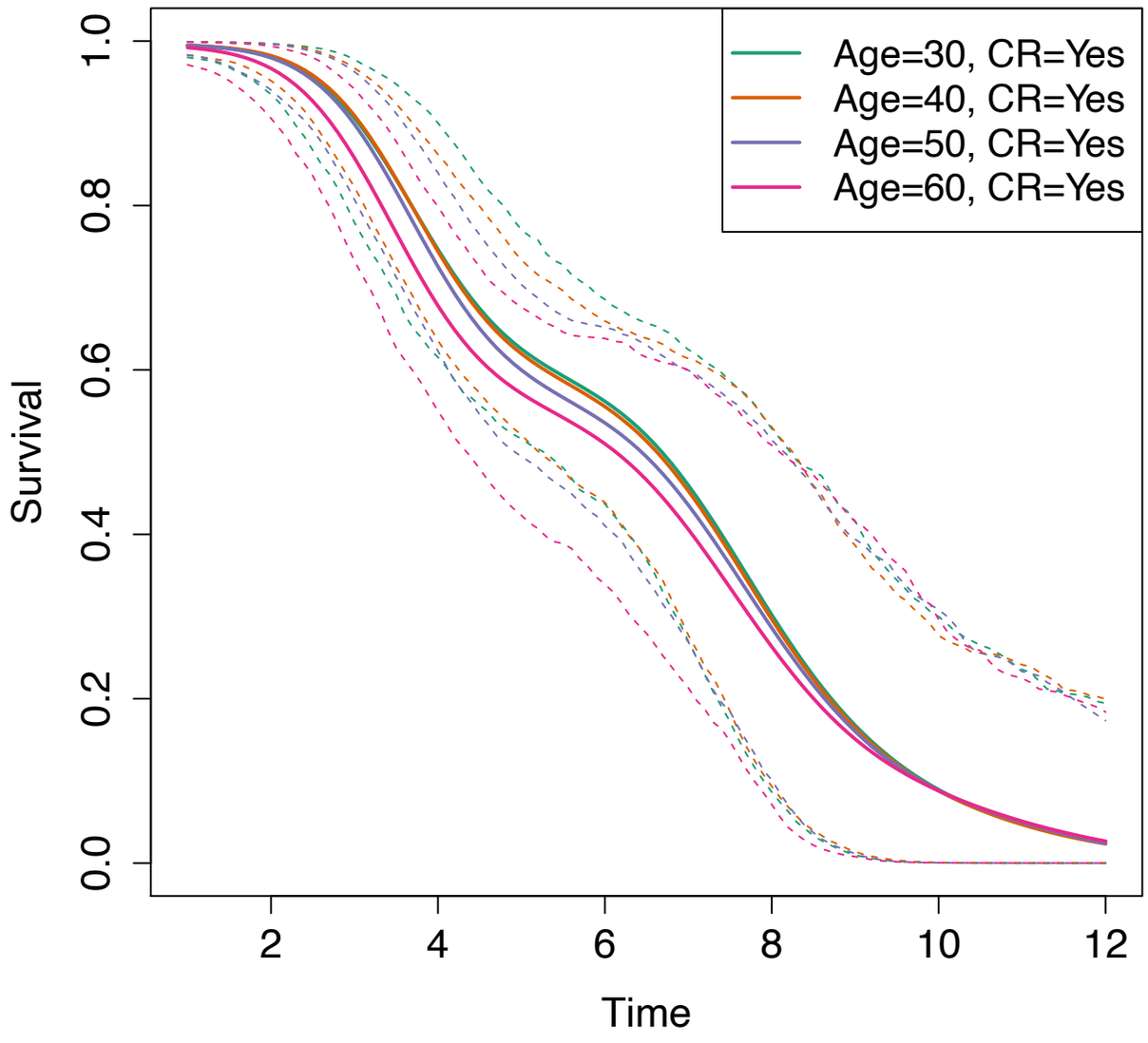
Data Analysis



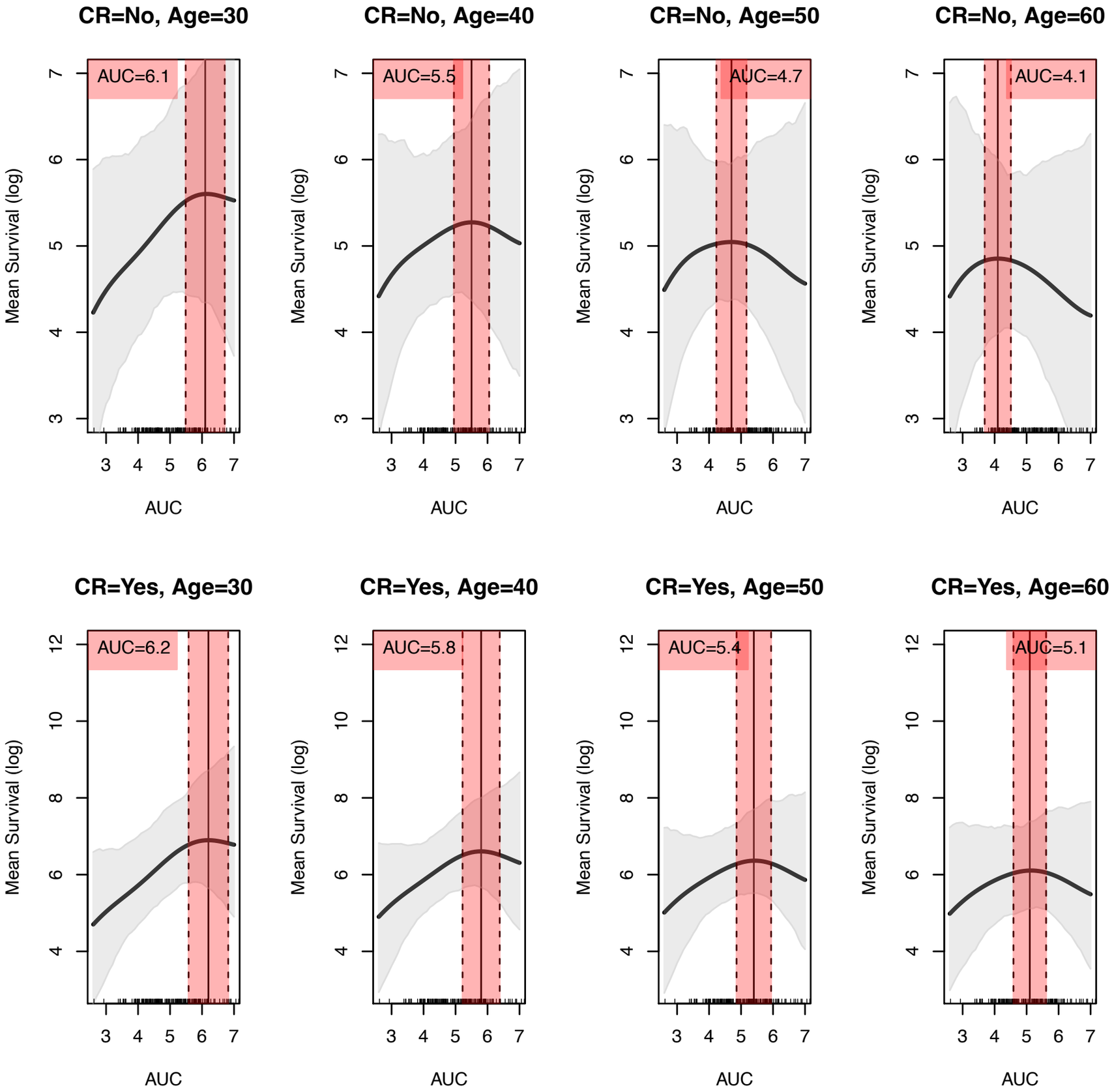
Data Analysis

AUC=5

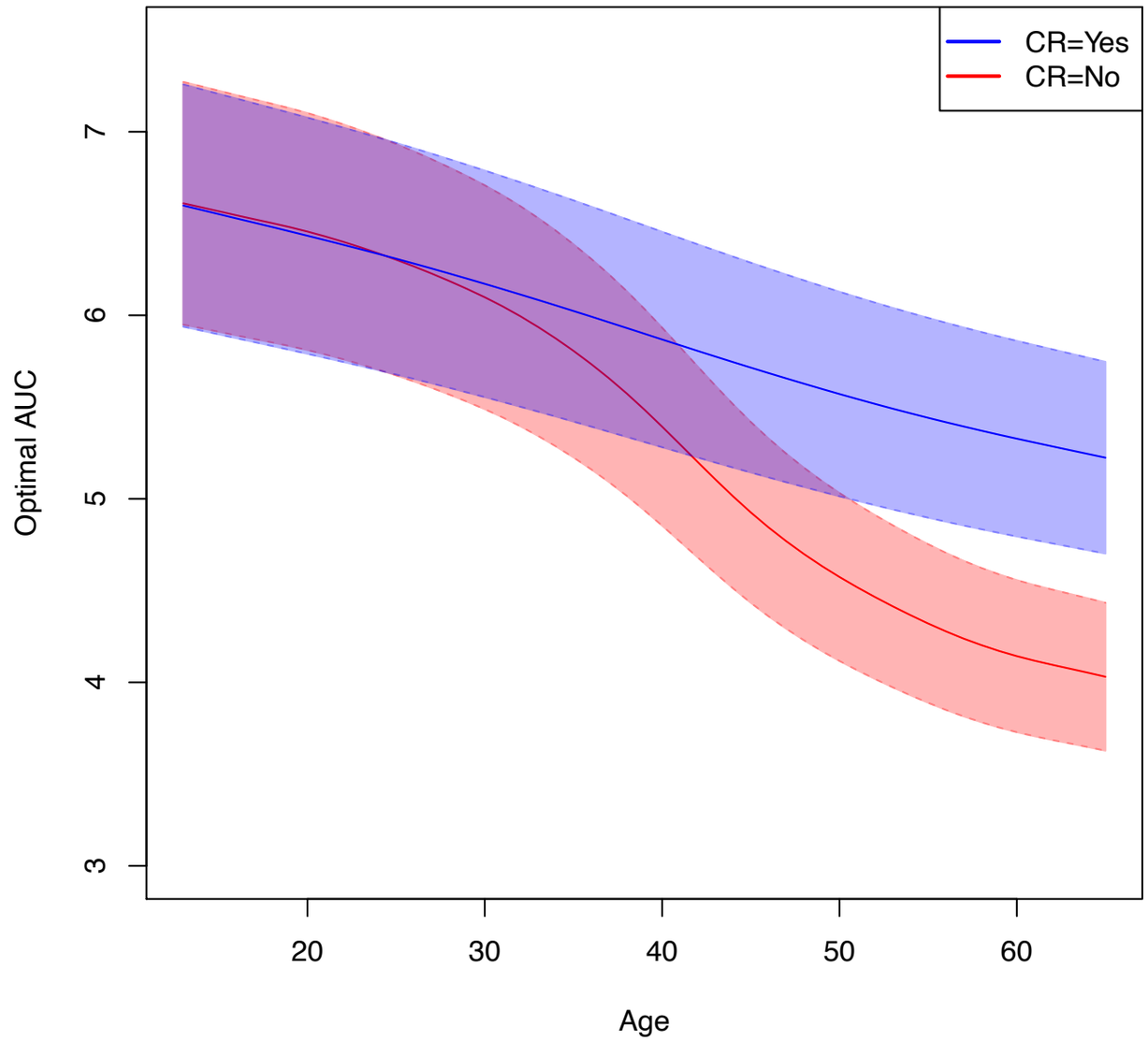




Data Analysis

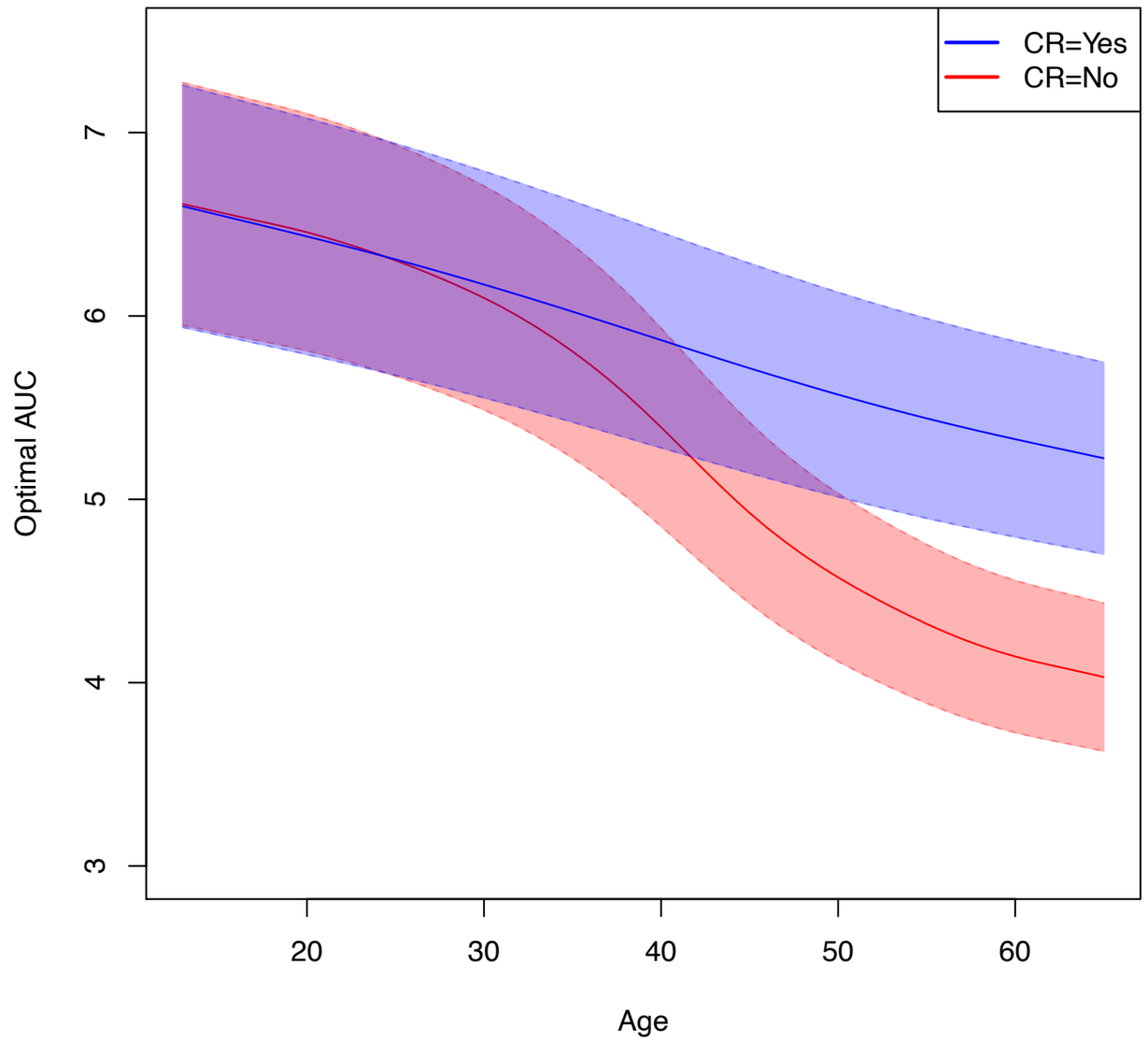


Data Analysis



Xu et al., 2019

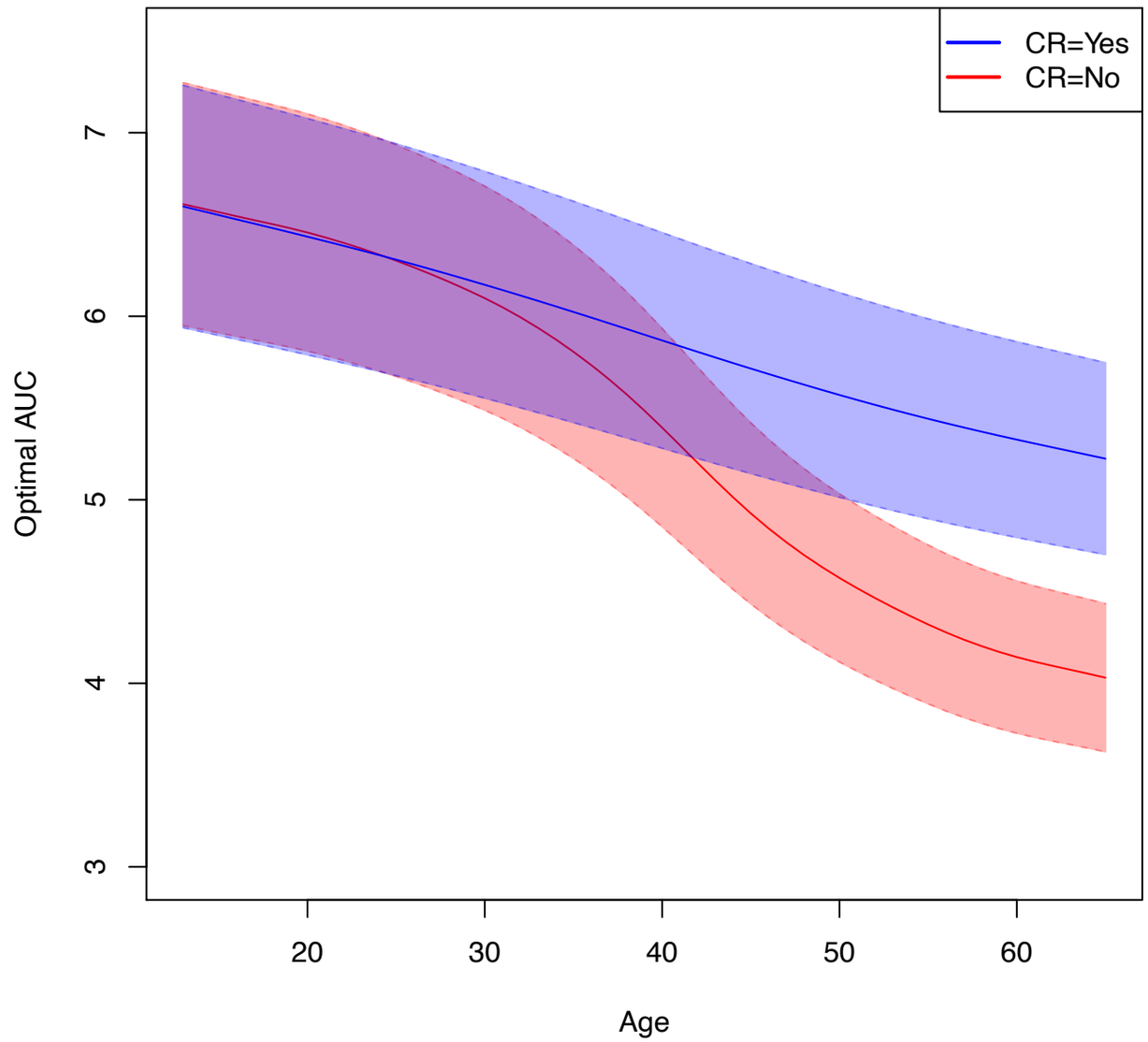
Data Analysis



- The optimal targeted busulfan dose interval goes down with Age.

Xu et al., 2019

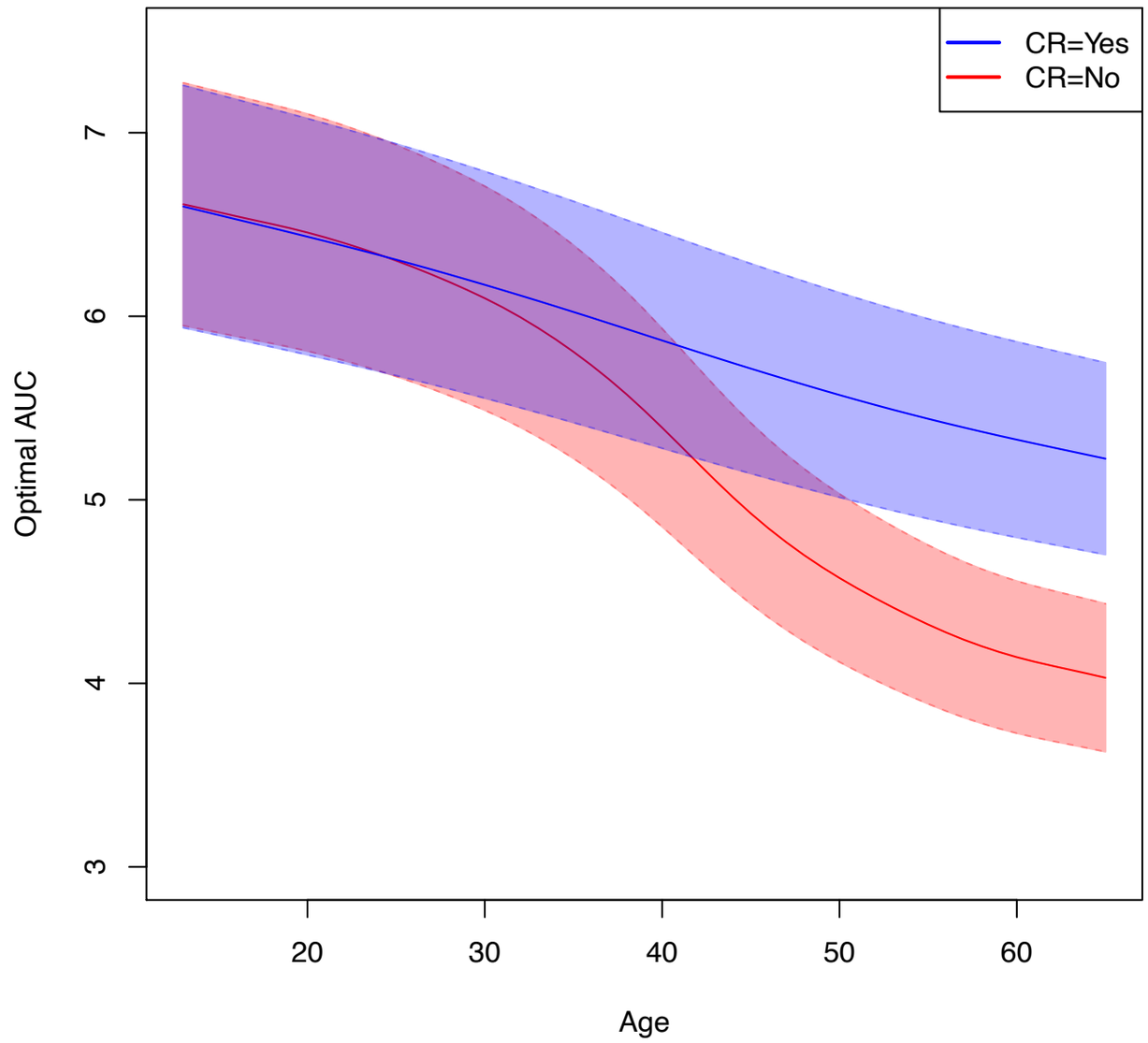
Data Analysis



- The optimal targeted busulfan dose interval goes down with Age.
- For Age ≤ 30, the optimal targeted IV busulfan dose intervals are identical for patients in CR or with active disease (no CR).

Xu et al., 2019

Data Analysis



- The optimal targeted busulfan dose interval goes down with Age.
- For Age ≤ 30, the optimal targeted IV busulfan dose intervals are identical for patients in CR or with active disease (no CR).
- For Age > 30, the optimal targeted dose intervals for [CR = No] are well below the intervals for [CR = Yes], with complete separation for Age > 55.

Xu et al., 2019

Take-home Messages

Take-home Messages

- This precision IV busulfan dosing may be applied worldwide in alloSCT to improve survival.

Take-home Messages

- This precision IV busulfan dosing may be applied worldwide in alloSCT to improve survival.
- The DDP-GP model is a tool for robust Bayesian nonparametric survival regression analysis that may be applied widely.

Take-home Messages

- This precision IV busulfan dosing may be applied worldwide in alloSCT to improve survival.
- The DDP-GP model is a tool for robust Bayesian nonparametric survival regression analysis that may be applied widely.

The R package DDPGPSurv can be downloaded from

https://cran.r-project.org/web/packages/DDPGPSurv/index.html