What's New in SAS/STAT®15.1

Changbin Guo, Clay Thompson, Fang Chen
SAS Institute Inc.
fangk.chen@sas.com

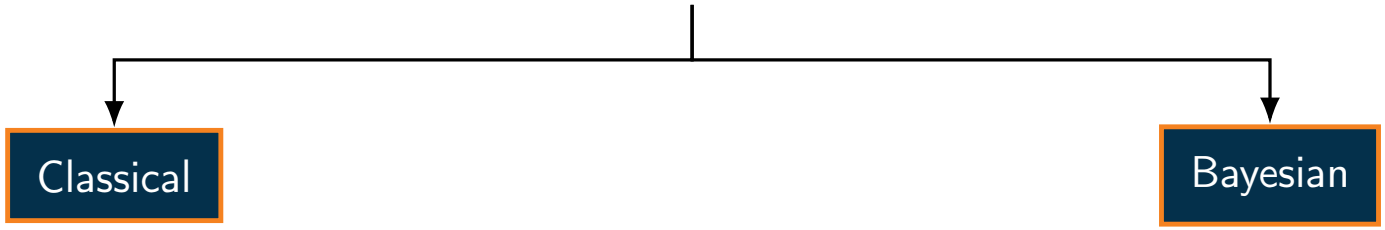
BASUG 2019

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

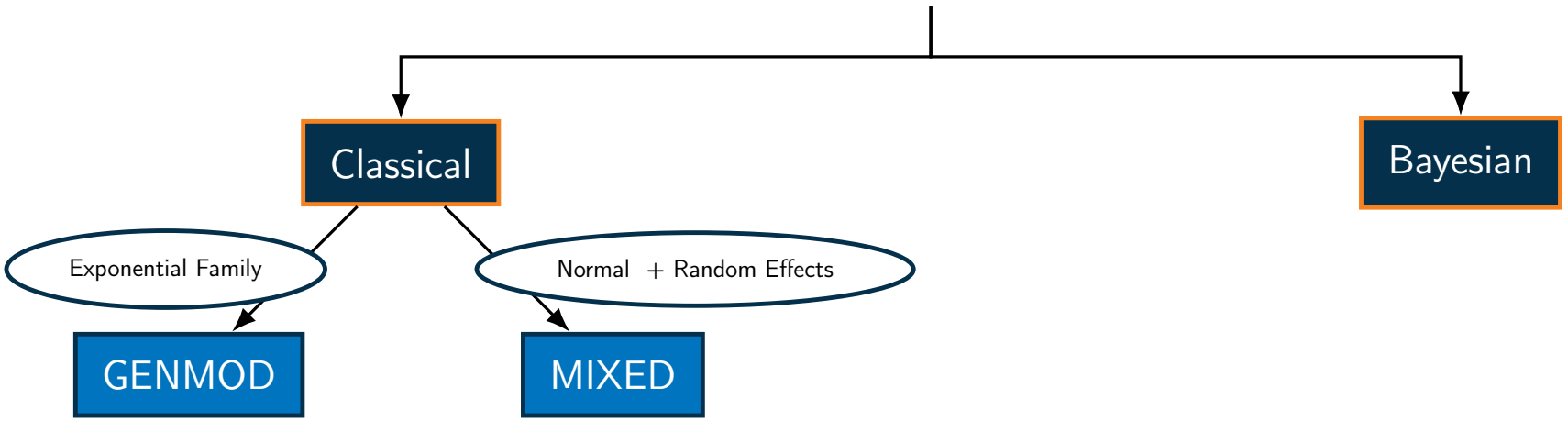
SAS/STAT 15.1 Features Three New Procedures

Procedure	Description
BGLIMM	Bayesian generalized linear mixed models
RMSTREG	Time-to-event analysis based on restricted mean survival time
CAUSALGRAPH	Graphical causal models for guiding causal inference

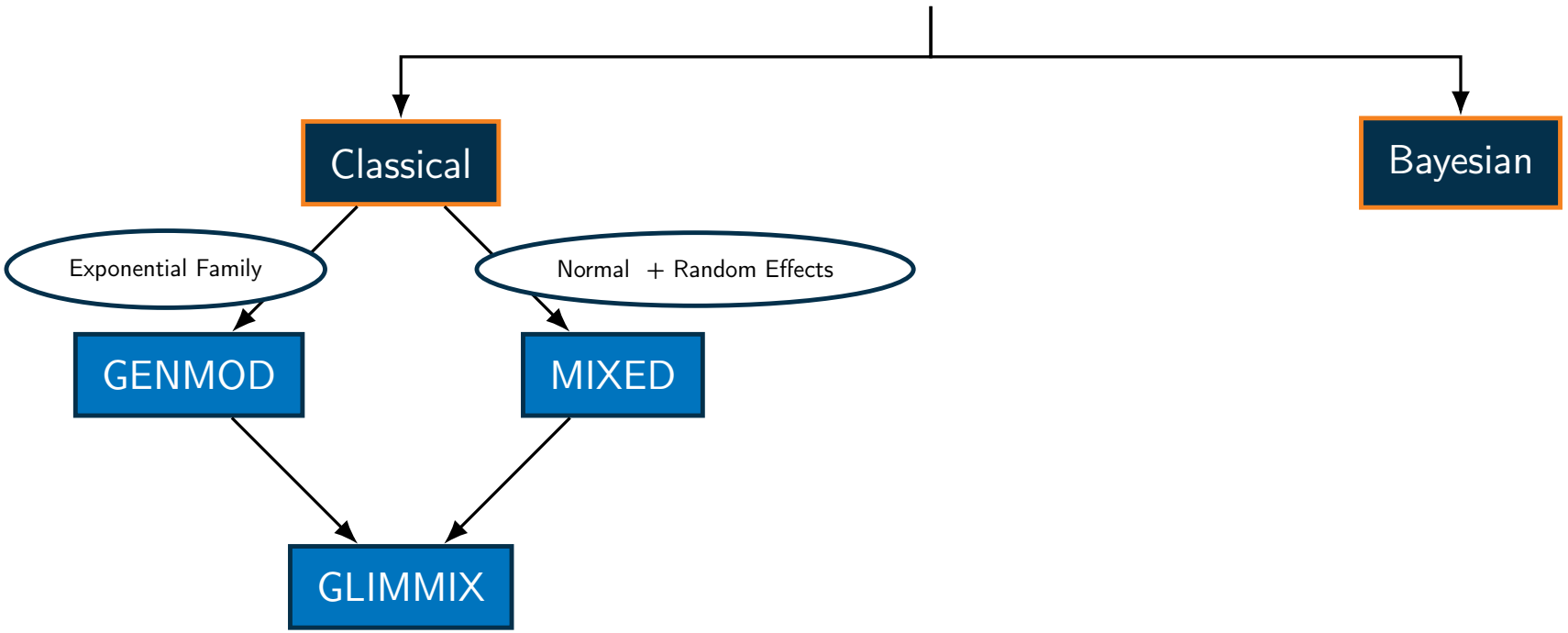
Tools in SAS/STAT for Mixed Models



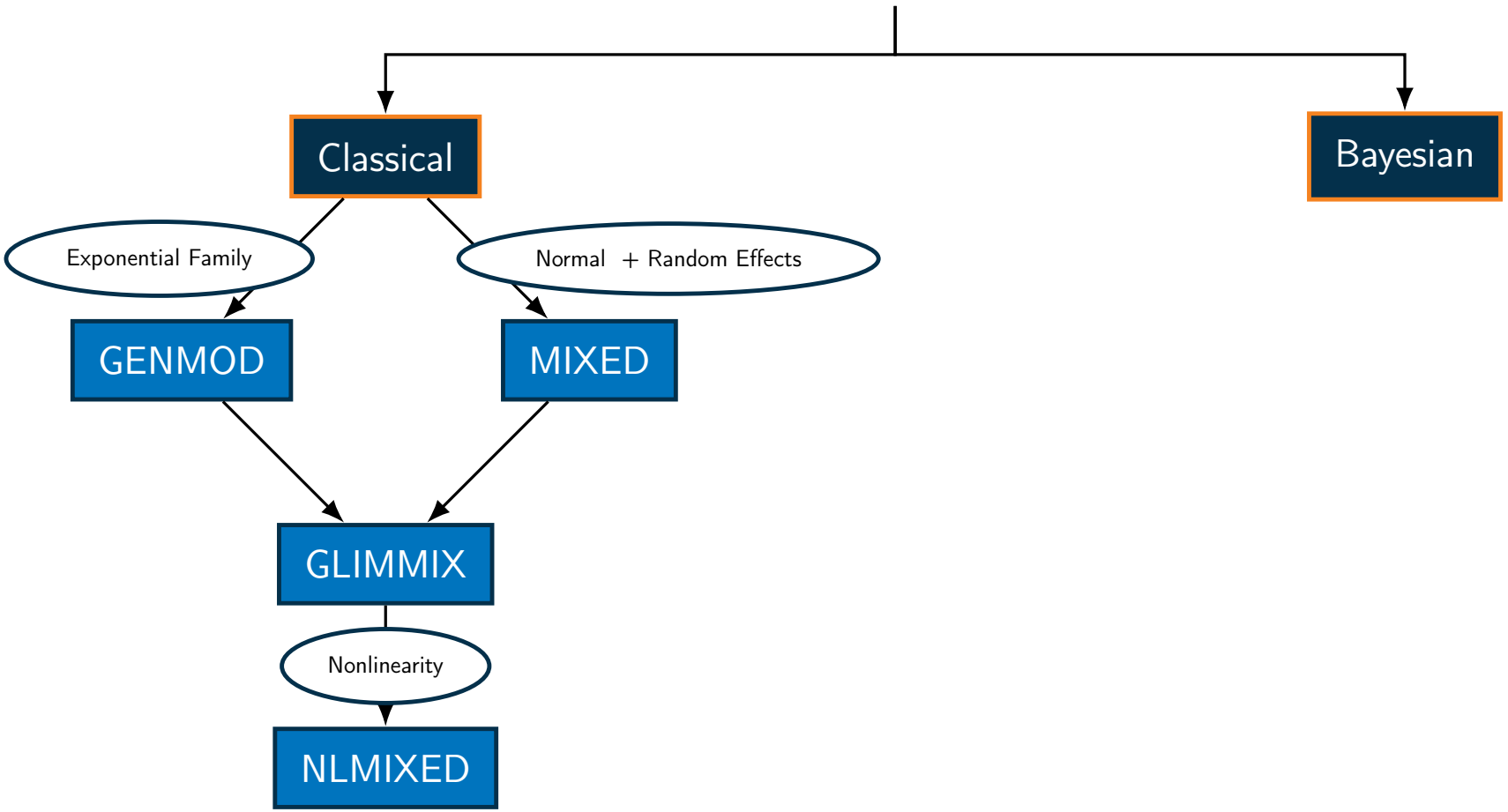
Tools in SAS/STAT for Mixed Models



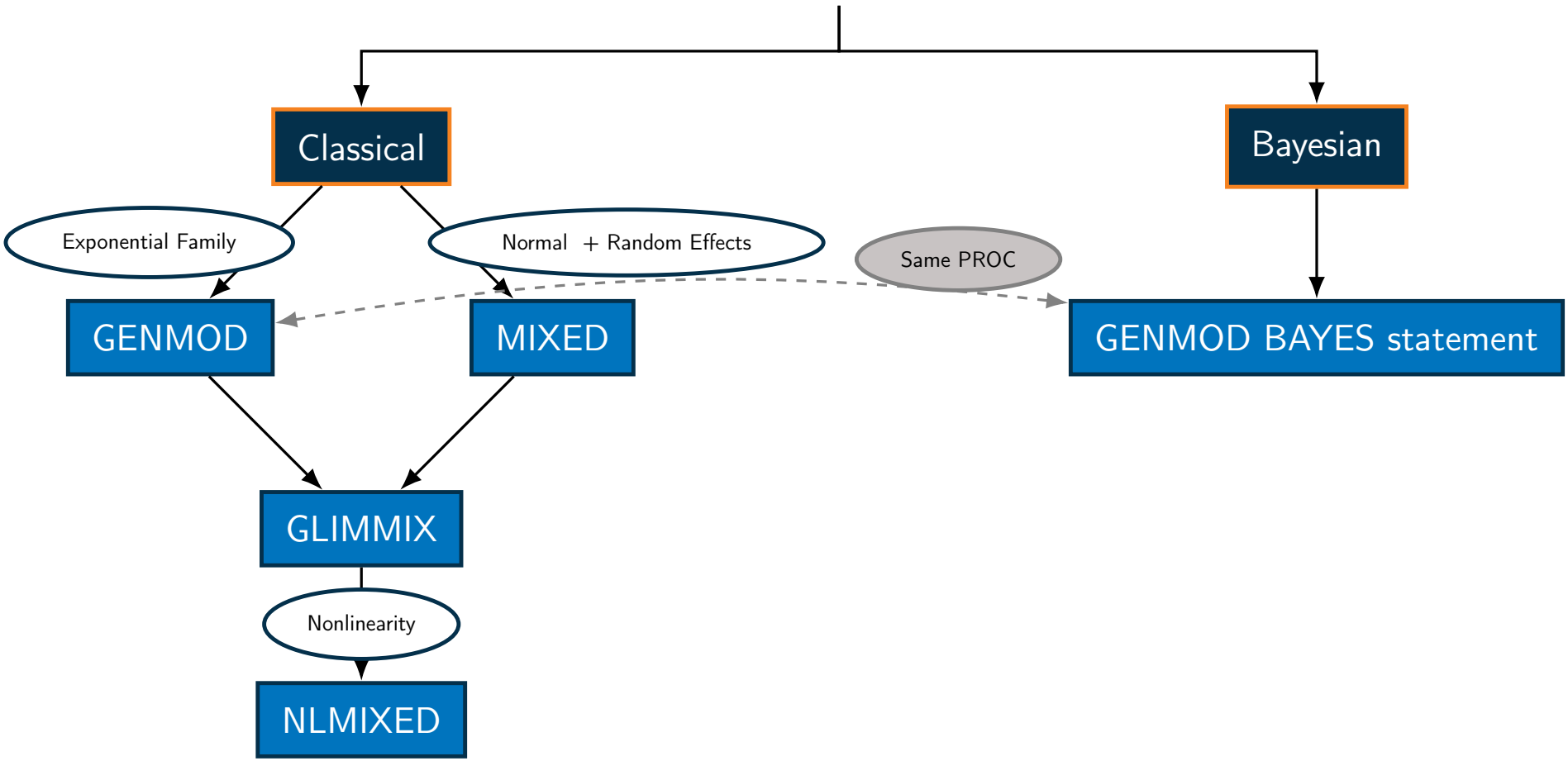
Tools in SAS/STAT for Mixed Models



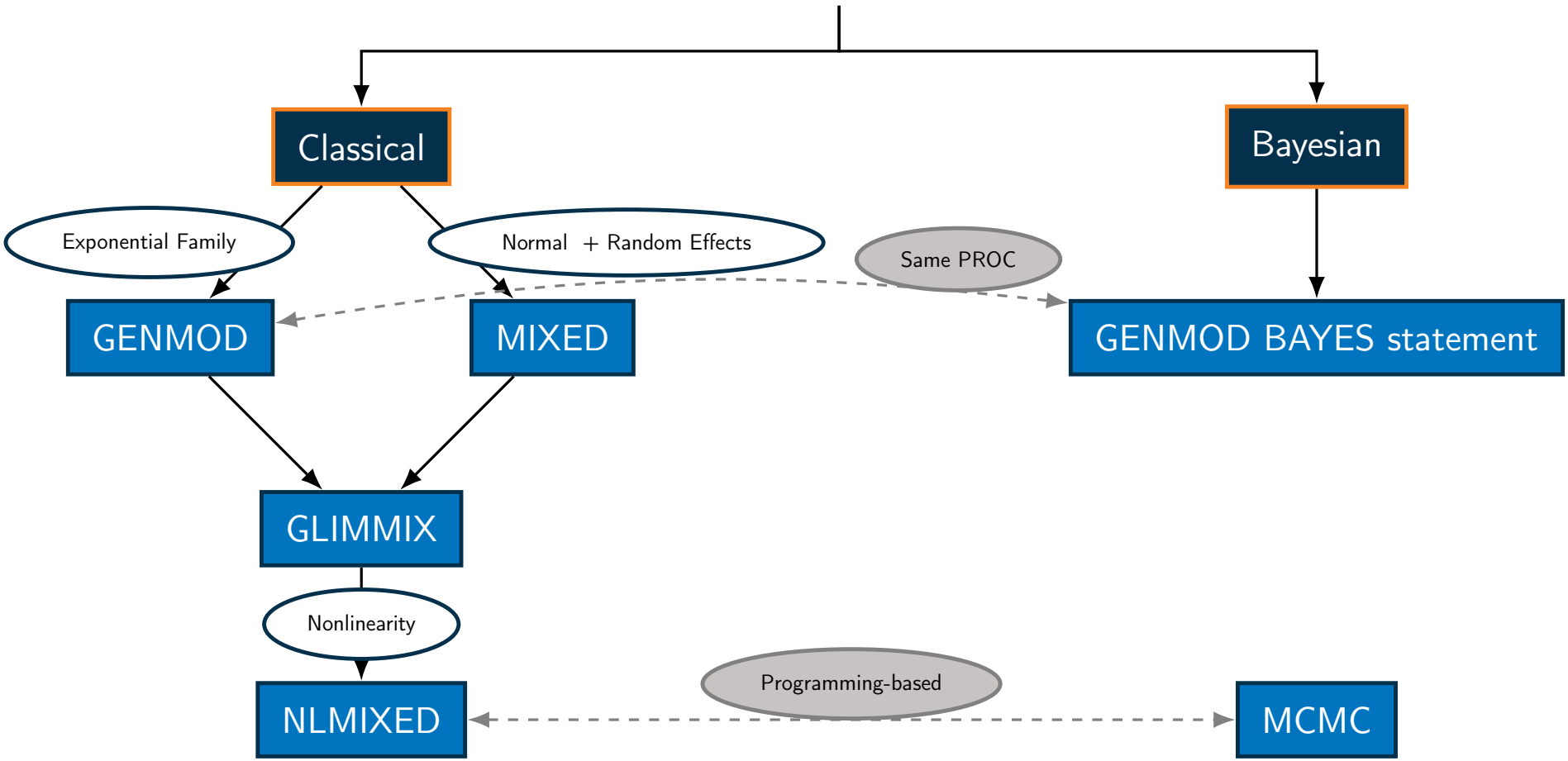
Tools in SAS/STAT for Mixed Models



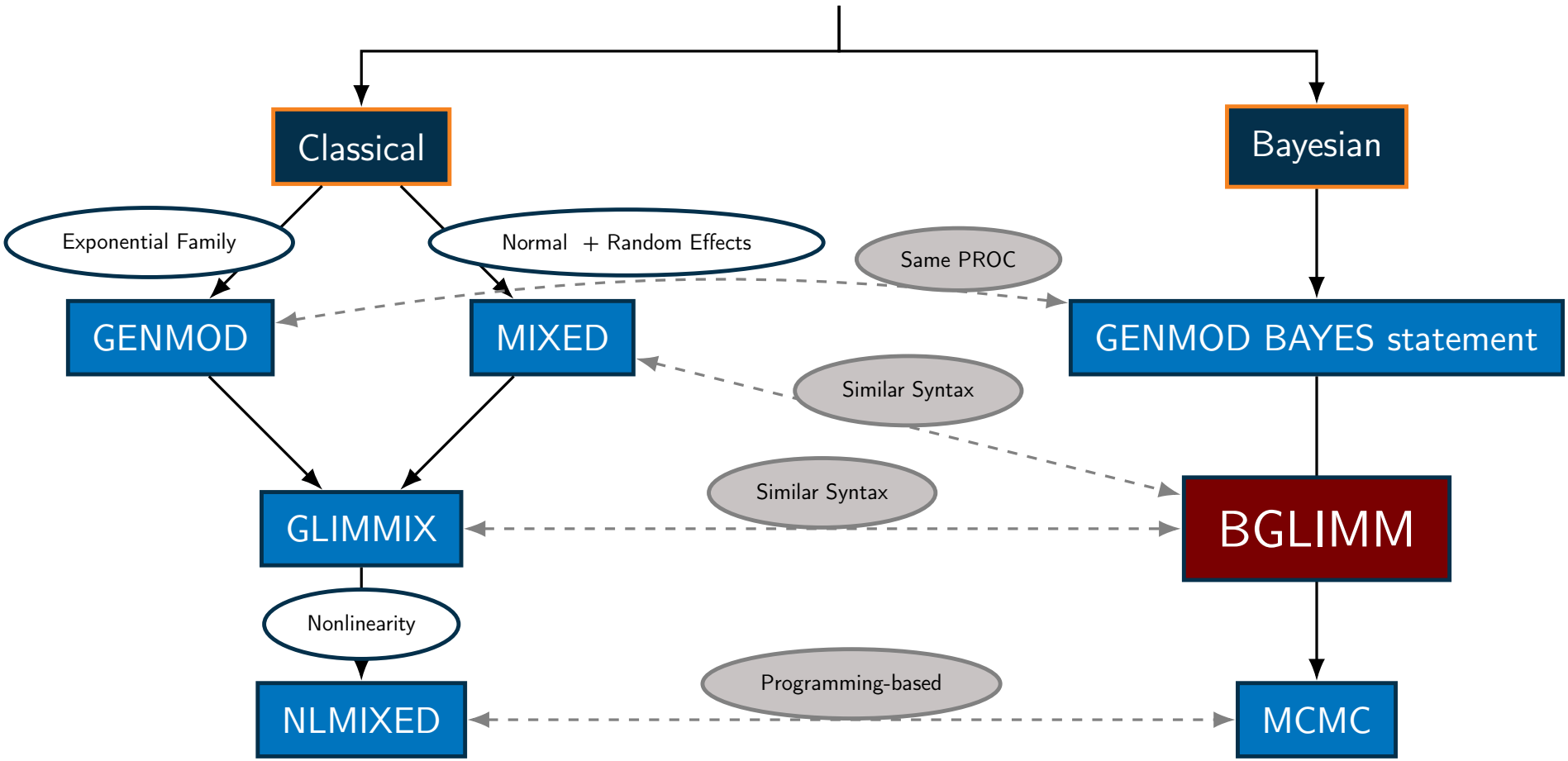
Tools in SAS/STAT for Mixed Models



Tools in SAS/STAT for Mixed Models



Tools in SAS/STAT for Mixed Models



Easy Access to Bayesian Generalized Linear Mixed Models

```
proc bglimm data=Fev;
    class Drug Patient Hour;
    model FEV = BaseVal Drug Hour;
    random int / subject=Patient;
    repeated hour /subject=Patient(Drug) type=un;
run;
```

- Full Bayesian solution (MCMC-based)

Easy Access to Bayesian Generalized Linear Mixed Models

```
proc bglimm data=Fev;
    class Drug Patient Hour;
    model FEV = BaseVal Drug Hour;
    random int / subject=Patient;
    repeated hour /subject=Patient(Drug) type=un;
run;
```

- Full Bayesian solution (MCMC-based)
- Easy transition for MIXED/GLIMMIX users

Easy Access to Bayesian Generalized Linear Mixed Models

```
proc bglimm data=Fev;
    class Drug Patient Hour;
    model FEV = BaseVal Drug Hour;
    random int / subject=Patient;
    repeated hour /subject=Patient(Drug) type=un;
run;
```

- Full Bayesian solution (MCMC-based)
- Easy transition for MIXED/GLIMMIX users
- Optimal algorithms, fast

Easy Access to Bayesian Generalized Linear Mixed Models

```
proc bglimm data=Fev;
    class Drug Patient Hour;
    model FEV = BaseVal Drug Hour;
    random int / subject=Patient;
    repeated hour /subject=Patient(Drug) type=un;
run;
```

- Full Bayesian solution (MCMC-based)
- Easy transition for MIXED/GLIMMIX users
- Optimal algorithms, fast
- Multilevel models, repeated measures, missing data, and many more features

Shi, Amy and Chen, Fang (2019) "Introducing the BGLIMM Procedure for Bayesian Generalized Linear Mixed Models" In Proceedings of the SAS Global Forum 2019 Conference, Cary, NC: SAS Institute Inc.

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

Introduction

- Time-to-event is a common endpoint in clinical trials
- Conventional statistical methods include:
  1. Kaplan-Meier method for estimating the survival function
  2. Log-rank test for comparing different groups
  3. Cox proportional hazards (PH) model for comparing different groups and adjusting for other covariates/stratification factors

Introduction

- Time-to-event is a common endpoint in clinical trials
- Conventional statistical methods include:
  1. Kaplan-Meier method for estimating the survival function
  2. Log-rank test for comparing different groups
  3. Cox proportional hazards (PH) model for comparing different groups and adjusting for other covariates/stratification factors

SAS procedures:

- PROC LIFETEST for KM and log-rank test
- PROC PHREG for Cox PH model

Situations When the PH Assumption May be Violated

- Crossing survival
- Delayed treatment effect
  - Late separation of survival curves
- Early treatment effect
  - No survival difference at late times
- Rare events

Non PH can Lead to Problems

- Log-rank test is no longer the most powerful test compared to alternatives
- For Cox regression, the hazard ratio (HR) estimate may not be very meaningful in quantifying the differences between groups
- Alternative non-hazard based methods are needed

Survival Mean

- Definition:

$$\mu = E[T] = \int_0^{\inf} tf(t)dt = \int_0^{\inf} S(t)dt$$

- Equivalent to the area under the survival curve
- However, this may not be estimable unless $S(t)$ is estimated for the whole space $[0, \inf)$

Restricted Mean Survival Time

- Consider the restricted time interval $[0, \tau]$, and let

$$R = \min(T, \tau)$$

Restricted Mean Survival Time

- Consider the restricted time interval $[0, \tau]$, and let

$$R = \min(T, \tau)$$

- RMST definition:

$$\text{RMST}_\tau = E(R) = E(T|T \leq \tau) = \int_0^\tau S(u)du$$

Restricted Mean Survival Time

- Consider the restricted time interval $[0, \tau]$, and let

$$R = \min(T, \tau)$$

- RMST definition:

$$\text{RMST}_\tau = E(R) = E(T|T \leq \tau) = \int_0^\tau S(u)du$$

- Equivalent to the area under the survival curve up to time τ
- Can be estimated nonparametrically using the KM estimates

RMST Characteristics

- Interpretability
  - In randomized trials, RMST can be interpreted as expected event-free time over the restricted period between randomization and a predefined time horizon

RMST Characteristics

- Interpretability
  - In randomized trials, RMST can be interpreted as expected event-free time over the restricted period between randomization and a predefined time horizon
- Effective summery of the survival profile
  - Set τ to be the largest follow-up time

RMST Characteristics

- Interpretability
  - In randomized trials, RMST can be interpreted as expected event-free time over the restricted period between randomization and a predefined time horizon
- Effective summery of the survival profile
  - Set τ to be the largest follow-up time
- Assumption-free

RMST Characteristics

- Interpretability
  - In randomized trials, RMST can be interpreted as expected event-free time over the restricted period between randomization and a predefined time horizon
- Effective summery of the survival profile
  - Set τ to be the largest follow-up time
- Assumption-free
- Need to pre-specify τ

You can Make the Same Types of Statistical Inferences Based on the RMST

	Classical Setting	**RMST**
Estimation	$S(t)$	RMST_τ
Group Comparison	$H_0 : S_1(t) = S_2(t)$	$H_0 : \text{RMST}_{\tau,1} = \text{RMST}_{\tau,2}$
Modeling	PH: $h(t) = h_0(t)\exp(X'\beta)$ AFT: $\log(T) = \beta_0 + X'\beta + \epsilon$	LM: $\text{RMST}_\tau = \beta_0 + X'\beta$ LLM: $\log(\text{RMST}_\tau) = \beta_0 + X'\beta$

- PH: Proportional Hazards Model (Cox)
- AFT: Accelerated Failure Time Model
- LM: Linear Model
- LLM: Log-Linear Model

SAS/STAT 15.1 Offers New Functionality for Analyzing the RMST

1. New RMST option in PROC LIFETEST
   - Nonparametric estimation
   - Group comparison
     - STRATA statement

SAS/STAT 15.1 Offers New Functionality for Analyzing the RMST

1. New RMST option in PROC LIFETEST
   - Nonparametric estimation
   - Group comparison
     - STRATA statement
2. New PROC RMSTREG
   - Fits linear and log-linear models to the RMST
   - Implements specialized methods based on estimating equations
   - Supports a variety of post-fitting analysis

RMST Estimation

- Nonparametric estimator:

$$\widehat{\text{RMST}}_\tau = \sum_{j=0}^{D} \hat{S}(t_j)(t_{j+1} - t_j)$$

RMST Estimation

- Nonparametric estimator:

$$\widehat{\text{RMST}}_\tau = \sum_{j=0}^{D} \hat{S}(t_j)(t_{j+1} - t_j)$$

- Variance formula can be derived based on the Greenwood formula

$$\text{Var}(\widehat{\text{RMST}}_\tau) = \frac{m}{m-1} \sum_{j=1}^{D} \left[\sum_{i=j}^{D} \hat{S}(t_i)(t_{i+1} - t_i) \right]^2 \frac{d_j}{Y_j(Y_j - d_j)}$$

where $m = \sum_{j=1}^{D} d_j$.

Two Sample Comparisons

- Confidence limits:

$$\widehat{\text{RMST}}_{\tau,1} - \widehat{\text{RMST}}_{\tau,2} \pm z_{\alpha/2}\sqrt{\text{Var}(\widehat{\text{RMST}}_{\tau,1}) + \text{Var}(\widehat{\text{RMST}}_{\tau,2})}$$

  - $z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the standard normal distribution
- Null hypothesis: $H_0 : \text{RMST}_{\tau,1} = \text{RMST}_{\tau,2}$
- Alternative hypothesis $H_1 : \text{RMST}_{\tau,1} \neq \text{RMST}_{\tau,2}$
- Test statistic:

$$T = \frac{\widehat{\text{RMST}}_{\tau,1} - \widehat{\text{RMST}}_{\tau,2}}{\sqrt{\text{Var}(\widehat{\text{RMST}}_{\tau,1}) + \text{Var}(\widehat{\text{RMST}}_{\tau,2})}}$$

- Asymptotically, T follows the standard normal distribution

Prognosis of Lung Cancer Patients

- Data: Kalbfleisch and Prentice (1980)
- Follow up study of lung cancer patients
- Outcome variable `SurvTime`: survival time in days
- Treatment variable `Treatment`: value 0 for standard therapy and value 1 for test therapy
- Censoring indicator `Censor`: value 1 indicating a censored time and the value 0 indicating an event time

Other Variables

- `Cell`: type of cancer cell
- `Therapy`: type of therapy: standard or test
- `Prior`: prior therapy: 0 = no, 10 = yes
- `Age`: age in years
- `DiagTime`: time in months from diagnosis to entry into the trial
- `Kps`: performance status

Data

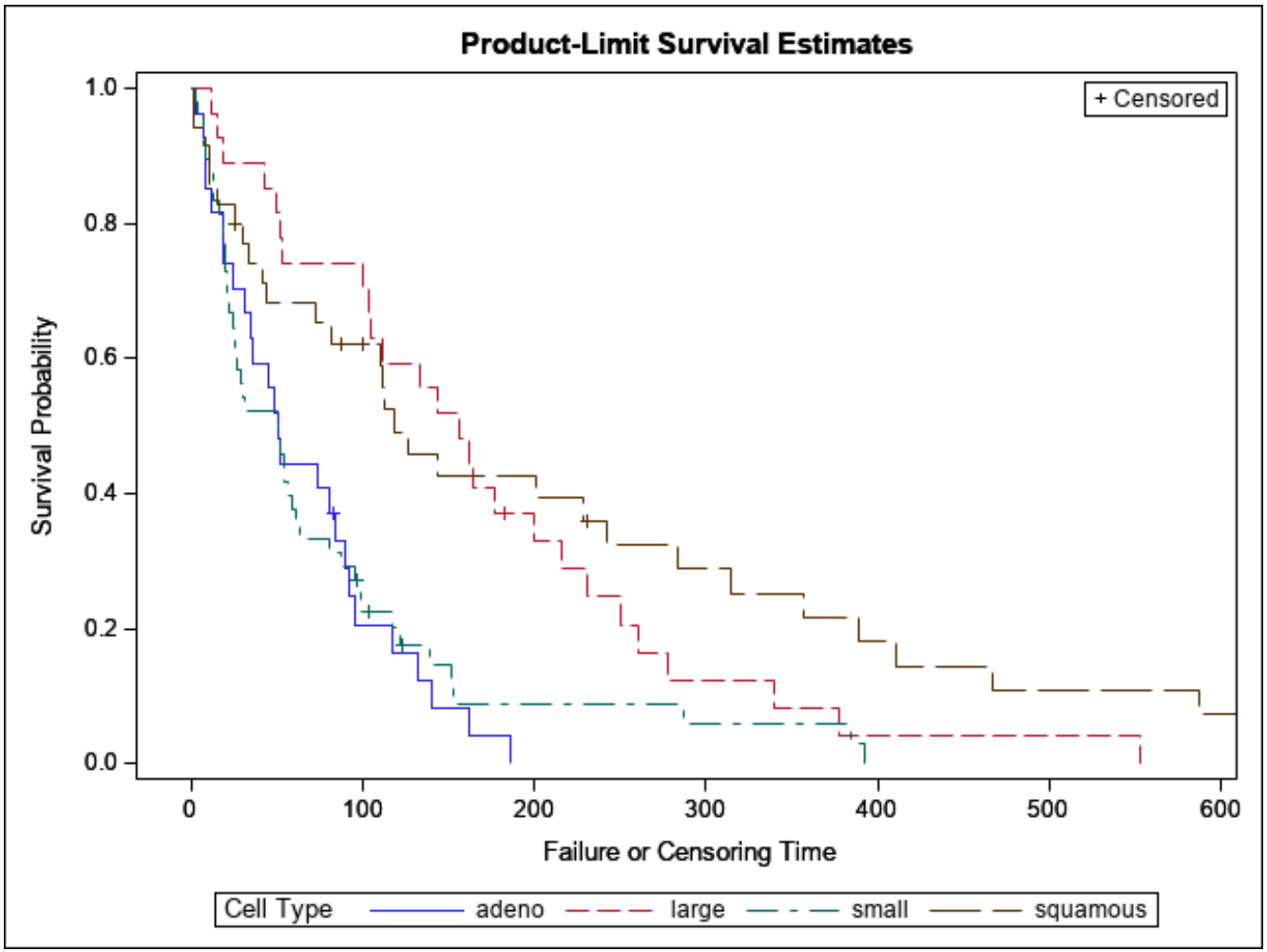
```
data VALung;
   drop check m;
   retain Therapy Cell;
   infile cards column=column;
   length Check $ 1;
   label SurvTime='Failure or Censoring Time'
         Kps='Karnofsky Index'
         DiagTime='Months till Randomization'
         Age='Age in Years'
         Prior='Prior Treatment?'
         Cell='Cell Type'
         Therapy='Type of Treatment'
         Treatment='Treatment Indicator';
 ... more lines ...
 52 60 4 45 0  164 70 15 68 10   19 30  4 39 10   53 60 12 66 0
 15 30 5 63 0   43 60 11 49 10  340 80 10 64 10  133 75  1 65 0
111 60 5 64 0  231 70 18 67 10  378 80  4 65  0   49 30  3 37 0
;
```

New Options in PROC LIFETEST Perform Nonparametric Analysis for the RMST

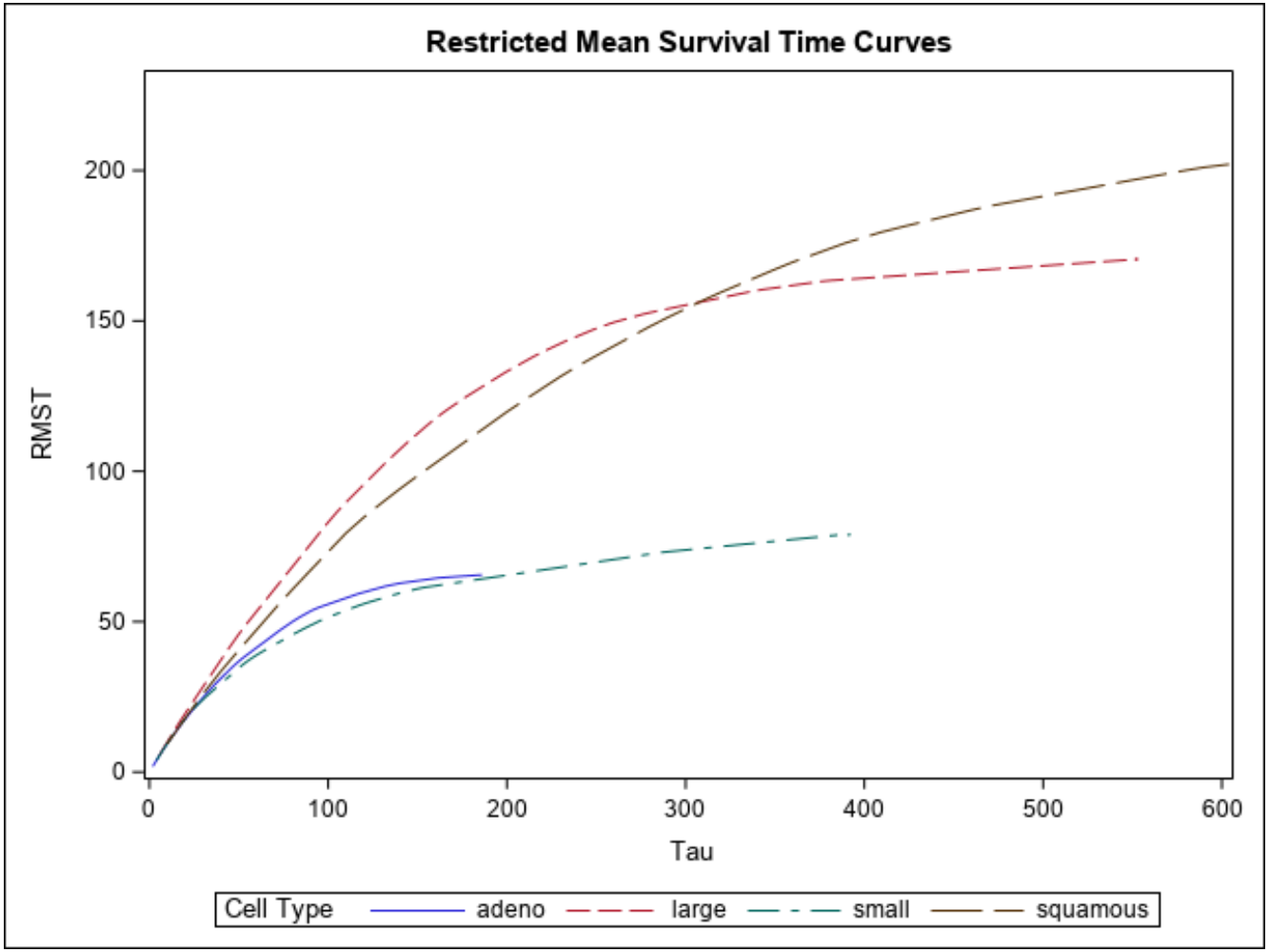
```
proc lifetest data=VALung plots=(rmst s) rmst maxtime=600;
   time SurvTime*Censor(1);
   strata Cell;
run;
```

- RMST requests RMST estimates
- PLOTS=RMST plots the RMST curve
- MAXTIME=600 sets the plotting limit to 600

Standard LIFETEST Output: Kaplan-Meier Curves



RMST Curve Plots the Estimated RMST Against Different τ Values



LIFETEST Results: Homogeneity Tests and Summary Statistics

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	25.4037	3	<.0001
Wilcoxon	19.4331	3	0.0002
-2Log(LR)	33.9343	3	<.0001

Summary of the Number of Censored and Uncensored Values

Stratum	Cell	Total	Failed	Censored	Percent Censored
1	adeno	27	26	1	3.70
2	large	27	26	1	3.70
3	small	48	45	3	6.25
4	squamous	35	31	4	11.43
Total		137	128	9	6.57

Results from RMST Analysis are Displayed After the Existing Analysis in PROC LIFETEST

RMST Analysis Information

Tau	186

RMST Estimates

Stratum	Cell Type	Estimate	Standard Error
1	adeno	65.55556	9.9303
2	large	128.0370	11.9858
3	small	64.20647	8.2859
4	squamous	113.8040	12.3703

RMST Test of Equality

Source	Chi-Square	DF	Pr > ChiSq
Strata	28.4427	3	<.0001

Determine the Value of τ?

- Suppose that there are K groups under comparison
- Let t_k be the largest follow-up time (event or censoring) for the kth group
- $\tau = \min_{k=1,\cdots,K} \{t_k\}$
- Note the log-rank test utilizes all data up to $\min(\tau, t_{\max})$, where $t_{\max}$ is the largest event time.

Use the DIFF= Option to Request Pairwise Comparisons

```
proc lifetest data=VALung rmst;
   time SurvTime*Censor(1);
   strata Cell / diff=all adj=sidak;
run;
```

- DIFF=ALL computes all paired differences of the RMST among the groups
- ADJ=SIDAK adjusts the p-values using Sidak's method

The p-Values are Properly Inflated to Account for Multiple Comparisons

Restricted Mean Survival Time Comparisons
Adjustment for Multiple Comparisons: Sidak

Stratum Comparison		Difference	Standard Error	Chi-Square	Pr > ChiSq Unadjusted	Adjusted
adeno	large	-62.4815	15.5650	16.1141	<.0001	0.0004
adeno	small	1.349087	12.9332	0.0109	0.9169	1.0000
adeno	squamous	-48.2485	15.8630	9.2511	0.0024	0.0140
large	small	63.83057	14.5710	19.1901	<.0001	<.0001
large	squamous	14.23303	17.2245	0.6828	0.4086	0.9572
small	squamous	-49.5975	14.8889	11.0967	0.0009	0.0052

Use DIFF=CONTROL to Choose a Control Group

```
proc lifetest data=VALung rmst;
   time SurvTime*Censor(1);
   strata Cell / diff=control('small') adj=sidak;
run;
```

Restricted Mean Survival Time Comparisons
Adjustment for Multiple Comparisons: Sidak

Stratum			Standard		Pr > ChiSq	
Comparison		Difference	Error	Chi-Square	Unadjusted	Adjusted
adeno	small	1.349087	12.9332	0.0109	0.9169	0.9994
large	small	63.83057	14.5710	19.1901	<.0001	<.0001
squamous	small	49.59754	14.8889	11.0967	0.0009	0.0026

Use the TAU= Option to Specify the Time Horizon

```
proc lifetest data=VALung rmst(tau=100);
   time SurvTime*Censor(1);
   strata Cell;
run;
```

Results Exhibit Similar Pattern as Previous Analysis

RMST Analysis Information

Tau	100

RMST Estimates

Stratum	Cell Type	Estimate	Standard Error
1	adeno	55.80247	6.8516
2	large	83.07407	5.7566
3	small	51.49653	5.1985
4	squamous	73.22857	6.3892

RMST Test of Equality

Source	Chi-Square	DF	Pr > ChiSq
Strata	20.0426	3	0.0002

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

PROC RMSTREG Performs Regression Modeling of the RMST

- Handles right-censored data
- Fits linear and log-linear models to the RMST
- Implements specialized estimation methods
  1. Pseudo value (PV) regression (default)
  2. Inverse probability censoring weighting (IPCW)
- Supports post-fitting analysis similar to generalized linear models
  - LSMEANS, ESTIMATE, LSMESTIMATE, SLICE

PROC RMSTREG is a New Procedure for Modeling Time-to-Event Data

Procedure	Focus	Model Type	Estimation Method
PROC LIFEREG	Time to Event	Accelerated Failure Time Models	Likelihood
PROC PHREG	Hazard Function	Proportional Hazards Models	Partial Likelihood
PROC RMSTREG	Restricted Mean Survival Time	Generalized Linear Models	Estimating Equations

RMST Regression versus Cox Regression

- Cox regression focuses on the hazard function
  - Interpret the model effects as hazard ratios
  - Relies on the proportional hazards assumption
  - Important to check the PH assumption

RMST Regression versus Cox Regression

- Cox regression focuses on the hazard function
  - Interpret the model effects as hazard ratios
  - Relies on the proportional hazards assumption
  - Important to check the PH assumption
- RMST regression models the RMST directly
  - Quantifies the effects in the RMST scale
  - Robust to non-PH

The Syntax of PROC RMSTREG Resembles PROC PHREG and PROC LIFEREG

```
PROC RMSTREG;
    BY variables ;
    CLASS variables;
    FREQ variable;
    MODEL response<*censor(list)> = < effects >;
    ESTIMATE 'label' effect values;
    LSMEANS < model-effects >;
    LSMESTIMATE model-effect;
    SLICE model-effect;
RUN;
```

What Factors Affect Mortality in Liver Cirrhosis?

- Data: Lin, Wei, and Ying (1993)
- 418 patients who have primary biliary cirrhosis (PBC)
- 161 had died as of the date of data listing
- Outcome variable `Time`: follow-up time (years)
- Event indicator `Status`: value 1 for death time and 0 for censored time;

Additional Variables

- `Age`: age from birth to study registration (years)
- `Albumin`: serum albumin level (g/dl)
- `Bilirubin`: serum bilirubin level (g/dl)
- `Edema`: edema presence
- `Protime`: prothrombin time (seconds)

Data

```
data Liver;
   input Time Status Age Albumin Bilirubin Edema Protime @@;
   label Time="Follow-Up Time in Years";
   Time= Time / 365.25;
datalines;
400 1 58.7652 2.60 14.5 1.0 12.2
4500 0 56.4463 4.14 1.1 0.0 10.6
1012 1 70.0726 3.48 1.4 0.5 12.0
... more lines ...
1055 0 57.0000 3.42 1.6 0.0 9.9
691 0 58.0000 3.75 0.8 0.0 10.4
976 0 53.0000 3.29 0.7 0.0 10.6
;
```

You can Fit a Linear model of the RMST at a Specific Time

```
proc rmstreg data=liver tau=10;
   class Edema;
   model Time*Status(0) = Age Bilirubin Edema
                / link=linear method=pv;
run;
```

$$\text{RMST}_\tau = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Bilirubin} + \beta_3 \cdot I(\text{Edema} = 0) + \beta_4 \cdot I(\text{Edema} = 0.5)$$

Information on the Input Data and Methods

The RMSTREG Procedure

Model Information	
Data Set	WORK.LIVER
Time Variable	Time
Censoring Variable	Status
Censoring Value(s)	0
Link Function	Linear
Estimation Method	Pseudo Value
Tau Value	10

Summary of the Number of Event and Censored Values

Total	Event	Censored
418	161	257

Results of the Fitted RMST Linear Model

Type 3 Analysis of Effects

Effect	DF	Chi-Square	Pr > ChiSq
Age	1	23.4052	<.0001
Bilirubin	1	88.8223	<.0001
Edema	2	19.4651	<.0001

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	9.0588	1.1201	6.8635	11.2540	65.41	<.0001
Age		1	-0.0686	0.0142	-0.0964	-0.0408	23.41	<.0001
Bilirubin		1	-0.3614	0.0383	-0.4366	-0.2862	88.82	<.0001
Edema	0	1	3.0259	0.7353	1.5849	4.4670	16.94	<.0001
Edema	0.5	1	1.7778	0.8658	0.0807	3.4748	4.22	0.0401
Edema	1	0	0.0000					

Pseudo Value Regression

- Generic method of fitting generalized linear models (Andersen, Klein, and Rosthoj 2003)
- Let $D_1, \cdots, D_n$ be independent and identically distributed (i.i.d.) quantities
- Let $\theta = E[f(D_i)]$ be the parameter of interest
- Suppose $\hat{\theta}$ is an unbiased estimator of θ
- Let $x_1, \cdots, x_n$ be i.i.d. samples of covariates
- The ith pseudo-observation of θ is computed as

$$\hat{\theta} = n\hat{\theta} - (n-1)\hat{\theta}^{-i}$$

- where $\hat{\theta}^{-i}$ is the jackknife leave-one-out estimator for θ based on $\{D_j : j \neq i\}$

Pseudo Value Regression

- Model: $g(\theta_i) = \beta' x_i$, where $g()$ is a link function
- Solve the estimating equations:

$$U(\beta) = \sum_{i=1}^{n} U_i(\beta) = \sum_{i=1}^{n} \left(\frac{\partial \theta_i}{\partial \beta} \right)' V_i^{-1} (\hat{\theta}_i - \theta_i)$$

  - where V_i is a working covariance matrix.
- Let $\hat{\beta}$ be a solution of the estimating equations
- The variance of $\hat{\beta}$ can be estimated via a sandwich estimator:

$$\Sigma = I_0^{-1} I_1 I_0^{-1}$$

$$I_0 = \sum_{i=1}^{n} \left(\frac{\partial \theta_i}{\partial \beta} \right)' V_i^{-1} (\hat{\theta}_i - \theta_i) \qquad I_1 = \sum_{i=1}^{n} U_i(\hat{\beta})' U_i(\hat{\beta})$$

Applying Pseudo Value Regression to RMST Models

- Let $\theta_i = \text{RMST}_\tau^i = E(T_i \wedge \tau | x_i)$
- Let

$$V_i = \begin{cases} \theta_i = \text{RMST}_\tau^i, & \text{if } g(u) = \log(u). \\ 1, & g(u) = u. \end{cases}$$

- RMST_τ can be unbiasedly estimated nonparametrically using the Kaplan-Meier estimates

Use LINK=LOG to Fit the Log-Linear Model

```
proc rmstreg data=liver tau=10;
   class Edema;
   model Time*Status(0) = Age Bilirubin Edema
                          / link=log method=pv;
run;
```

$$\log(\text{RMST}_\tau) = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Bilirubin} + \beta_3 \cdot I(\text{Edema} = 0) + \beta_4 \cdot I(\text{Edema} = 0.5)$$

Results of Fitted Log-Linear Model

The RMSTREG Procedure

Type 3 Analysis of Effects

Effect	DF	Chi-Square	Pr > ChiSq
Age	1	23.2684	<.0001
Bilirubin	1	53.9905	<.0001
Edema	2	13.8173	0.0010

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	1.7815	0.3241	1.1462	2.4168	30.21	<.0001
Age		1	-0.0098	0.0020	-0.0138	-0.0058	23.27	<.0001
Bilirubin		1	-0.0848	0.0115	-0.1074	-0.0621	53.99	<.0001
Edema	0	1	0.9362	0.2952	0.3576	1.5148	10.06	0.0015
Edema	0.5	1	0.7356	0.3083	0.1314	1.3398	5.69	0.0170
Edema	1	0	0.0000					

Use ESTIMATE Statement to Obtain Model-Based RMST Ratios

```
proc rmstreg data=liver tau=10;
   class Edema;
   model Time*Status(0) = Age Bilirubin Edema
                          / link=log method=pv;
   estimate '0 vs 0.5' Edema 1 -1 0 / exp cl;
   estimate '0.5 vs 1' Edema 0 1 -1 / exp cl;
   estimate '0 vs 1'   Edema 1 0 -1 / exp cl;
run;
```

$$\log(\text{RMST}_\tau) = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Bilirubin} + \beta_3 \cdot I(\text{Edema} = 0) + \beta_4 \cdot I(\text{Edema} = 0.5)$$

Results: Covariate-Adjusted RMST Ratios

	Estimate									
Label	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
0 vs 0.5	0.2006	0.09844	2.04	0.0416	0.05	0.007668	0.3936	1.2222	1.0077	1.4822

	Estimate									
Label	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
0.5 vs 1	0.7356	0.3083	2.39	0.0170	0.05	0.1314	1.3398	2.0867	1.1404	3.8182

	Estimate									
Label	Estimate	Standard Error	z Value	Pr > |z|	Alpha	Lower	Upper	Exponentiated	Exponentiated Lower	Exponentiated Upper
0 vs 1	0.9362	0.2952	3.17	0.0015	0.05	0.3576	1.5148	2.5503	1.4299	4.5487

Inverse Probability Censoring Weighting Method

- Data: $\{(U_i, \Delta_i, x_i) : i = 1, \cdots, n\}$ where $U_i = \min(T_i, C_i)$, and $\Delta_i = I(T_i \leq C_i)$
- Define:

$$\begin{aligned} R_i &= T_i \wedge \tau \\ \text{RMST}_\tau^i &= E(R_i) \\ \hat{\Delta}_i &= I(R_i \leq C_i) \end{aligned}$$

- Let $\hat{G}(t)$ be the Kaplan-Meier estimate of the survival function of the censoring variable
- Compute subject-wise weights as:

$$w_i = \frac{\hat{\Delta}_i}{\hat{G}(R_i)}$$

Inverse Probability Censoring Weighting Method

- Model: $g(\text{RMST}_\tau^i) = \beta' x_i$, where $g(\cdot)$ is a link function
- Solve the estimating equations:

$$U(\beta) = \sum_{i=1}^{n} U_i(\beta) = \sum_{i=1}^{n} w_i(R_i - g^{-1}(\beta' x_i))x_i = 0$$

  where w_i is the weight for subject i.
- Let $\hat{\beta}$ a solution of the estimating equations
- The variance of $\hat{\beta}$ can be estimated via a sandwich estimator.

What Factors Affect HIV Mortality?

- Data: Hosmer and Lemeshow (1999)
- Patients were followed after a confirmed HIV-positive diagnosis
- Outcome variable `Time`: follow-up time (days)
- Event indicator `Status`: value 1 for death time and 0 for censored time
- `Drug`: a value of 1 for prior intravenous drug use and 0 otherwise
- `Age`: patient's age (years) at the beginning of the follow-up

Data

```
data HIV;
   input Time Age Drug Status;
   label Time="Follow-Up Time in Months";
datalines;
5 46 0 1
6 35 1 0
8 30 1 1
... more lines ...
1 34 1 1
;
```

Specify MEHTOD=IPCW to Use the IPCW Method

```
proc rmstreg data=hiv tau=48;
   class Drug;
   model Time*Status(0) = Drug Age / link=linear method=ipcw;
run;
```

$$\text{RMST}_\tau = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot I(\text{Drug} = 1)$$

Model information

The RMSTREG Procedure

Model Information	
Data Set	WORK.HIV
Time Variable	Time
Censoring Variable	Status
Censoring Value(s)	0
Link Function	Linear
Estimation Method	IPCW
Tau Value	48

Both Age and Drug Use are Strong Predictors of Mortality for a Time Horizon of 48 Months

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	44.6305	8.8479	27.2889	61.9720	25.44	<.0001
Drug	0	1	12.5483	2.6029	7.4468	17.6499	23.24	<.0001
Drug	1	0	0.0000					
Age		1	-1.0748	0.2289	-1.5234	-0.6262	22.05	<.0001

Compare to page 60

Fit a Log-Linear Model using IPCW Method

```
proc rmstreg data=hiv tau=48;
   class Drug;
   model Time*Status(0) = Drug Age / link=log method=ipcw;
run;
```

$$\log(\text{RMST}_\tau) = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot I(\text{Drug} = 1)$$

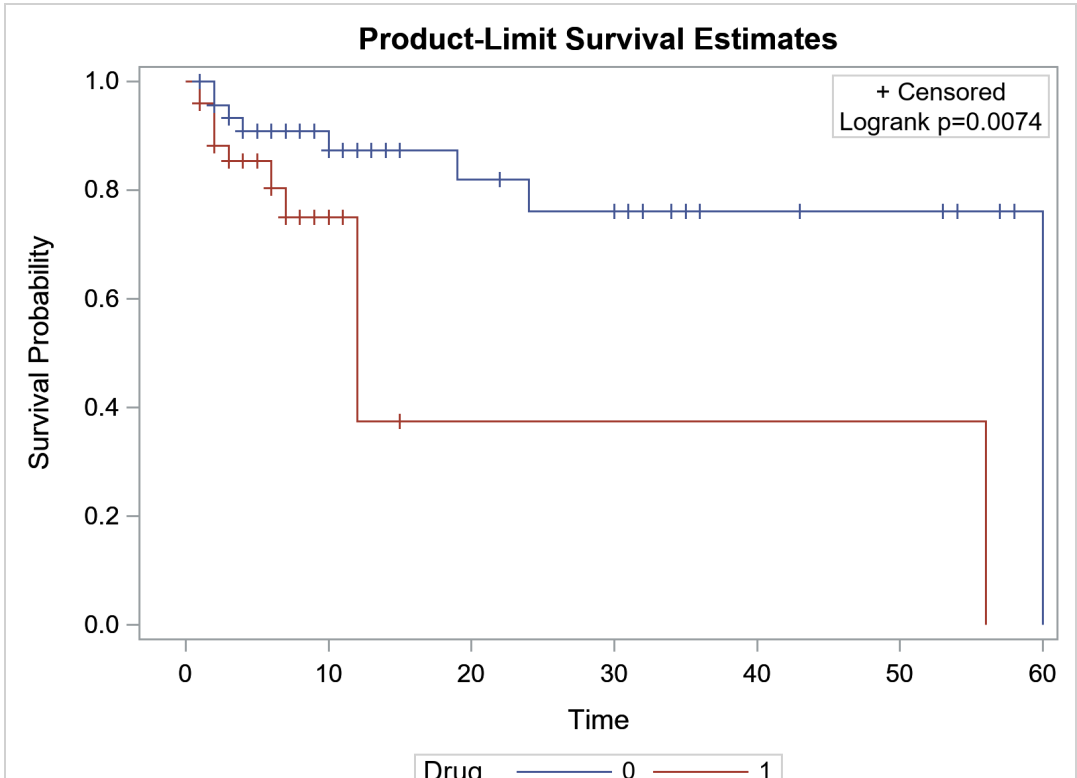
Results of the Log-Linear Model Agree with the Linear Model

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.6169	0.6984	3.2480	5.9858	43.70	<.0001
Drug	0	1	1.0711	0.2291	0.6220	1.5201	21.86	<.0001
Drug	1	0	0.0000					
Age		1	-0.0821	0.0175	-0.1164	-0.0478	22.03	<.0001

Is the Censoring Distribution Homogeneous Between the Two Drug Use Groups?

```
proc lifetest data=hiv plot=s(test);
   strata Drug;
   time Time*Status(1);
run;
```



You can Account for Censoring Inhomogeneity by Using the STRATA= Option

```
proc rmstreg data=hiv tau=48;
   class Drug;
   model Time*Status(0) = Drug Age / link=linear
                                     method=ipcw(strata=Drug)
run;
```

Stratification on Censoring does not Change the Conclusions

The RMSTREG Procedure

Analysis of Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	47.5159	10.5939	26.7522	68.2795	20.12	<.0001
Drug	0	1	11.4049	2.5667	6.3743	16.4354	19.74	<.0001
Drug	1	0	0.0000					
Age		1	-1.1380	0.2654	-1.6582	-0.6179	18.39	<.0001

▸ Compare to page 55, without stratification

Pseudo Value Regression versus IPCW Method

- Both techniques can be applied to fit RMST models
- Common assumption: censoring is noninformative to survival
- IPCW method assumes additionally that the censoring distribution can be properly estimated

$$T|(C,X) \quad \text{vs} \quad (T,C)|X$$

- Pseudo value regression is robust to dependent censoring
- IPCW method can be more efficient when censoring is properly accounted

RMST Methods Offer Useful Alternatives to Classical Hazard-Based Methods

- RMST can be more reliably estimated than mean and median
- When non proportional hazards occur, RMST methods can still be applied to obtain interpretable results
- Use the RMST option in PROC LIFETEST for nonparametric analyses
- Use PROC RMSTREG to perform regression modeling of the RMST

Guo, Changbin (2019) "Modeling Nonproportional Hazards by Performing Restricted Mean Survival Time Analysis" In Proceedings of the SAS Global Forum 2019 Conference, Cary, NC: SAS Institute Inc.

Guo, Changbin (2019) "Modeling Nonproportional Hazards by Performing Restricted Mean Survival Time Analysis" In Proceedings of the SAS Global Forum 2019 Conference, Cary, NC: SAS Institute Inc.

Guo, Changbin and So, Ying (2018) "Cause-Specific Analysis of Competing Risks using the PHREG Procedure" In Proceedings of the SAS Global Forum 2018 Conference, Cary, NC: SAS Institute Inc.

Guo, Changbin, So, Ying, and Woosung, Jang (2017) "Evaluating Predictive Accuracy of Survival Models with PROC PHREG" In Proceedings of the SAS Global Forum 2017 Conference, Cary, NC: SAS Institute Inc.

Outline

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

Causal Analysis in SAS/STAT

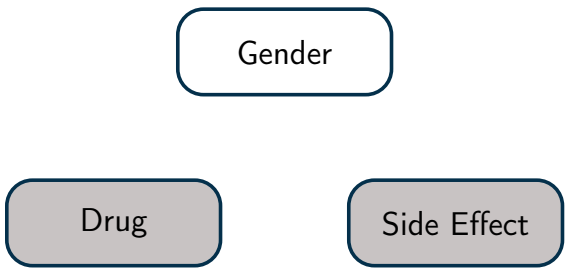
Procedure	Description
PSMATCH	Create and match propensity scores
CAUSALTRT	Estimate the causal effects on a binary treatment
CAUSALMED	Estimate causal mediation effects
CAUSALGRAPH	Examine the structure of causal graphical models

PROC CAUSALGRAPH is a pre-analysis procedure that helps you find proper adjustment sets to make valid causal interpretations.

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

When Should You Adjust for a Covariate?

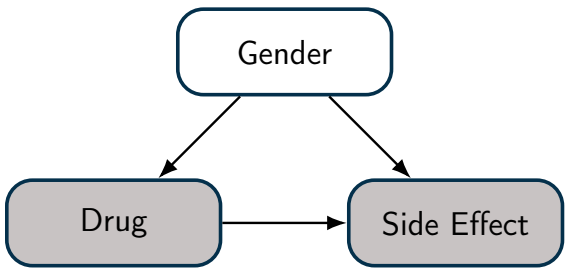
The Adjustment Decision Requires a Causal Model



Obs	Gender	Drug	SideEffect
1	Female	No	Yes
2	Female	Yes	No
3	Male	Yes	Yes
4	Female	Yes	No
5	Male	Yes	No
6	Male	Yes	Yes
7	Female	No	Yes
8	Female	Yes	Yes
9	Male	Yes	No
10	Female	No	No

When Should You Adjust for a Covariate?

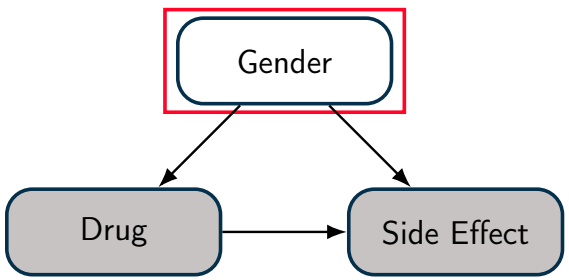
The Adjustment Decision Requires a Causal Model



Obs	Gender	Drug	SideEffect
1	Female	No	Yes
2	Female	Yes	No
3	Male	Yes	Yes
4	Female	Yes	No
5	Male	Yes	No
6	Male	Yes	Yes
7	Female	No	Yes
8	Female	Yes	Yes
9	Male	Yes	No
10	Female	No	No

When Should You Adjust for a Covariate?

The Adjustment Decision Requires a Causal Model

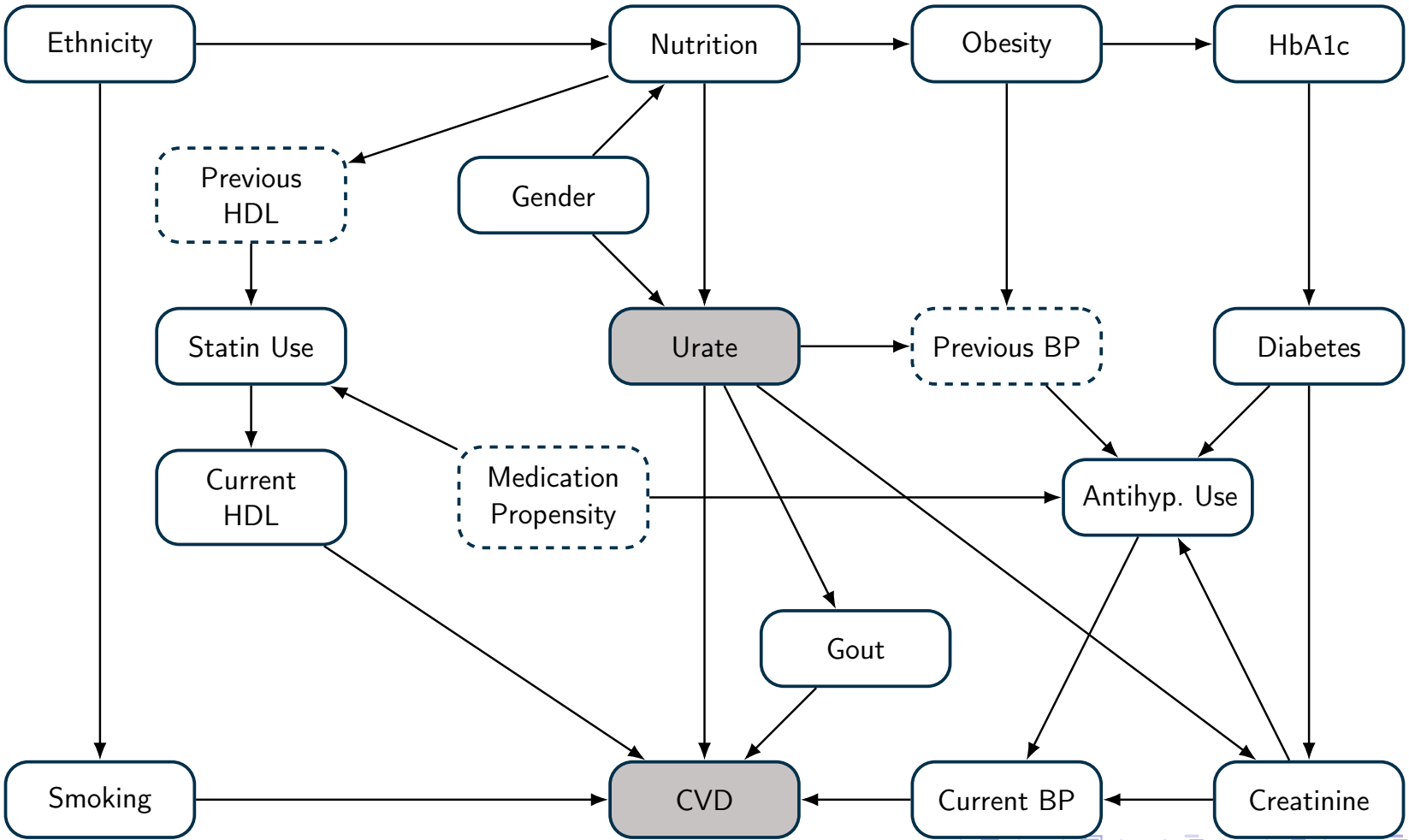


Obs	Gender	Drug	SideEffect
1	Female	No	Yes
2	Female	Yes	No
3	Male	Yes	Yes
4	Female	Yes	No
5	Male	Yes	No
6	Male	Yes	Yes
7	Female	No	Yes
8	Female	Yes	Yes
9	Male	Yes	No
10	Female	No	No

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

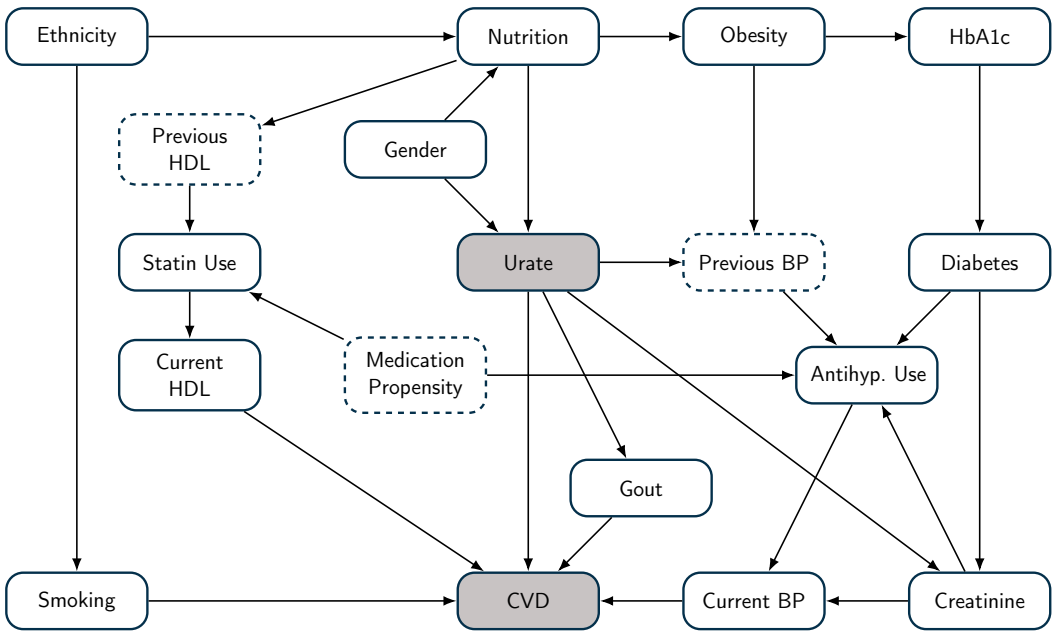
A Causal Model for the Effect of Serum Urate on Cardiovascular Disease

Adapted from Thornley et al. (2013)



Specifying an Analysis with the CAUSALGRAPH Procedure

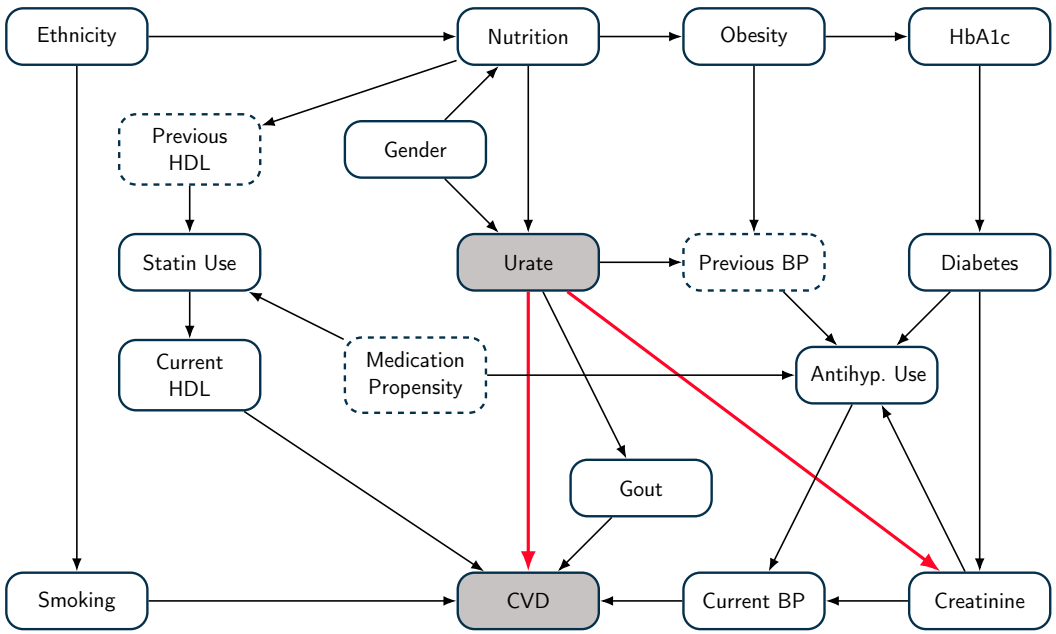
```
proc causalgraph compact minimal;
  model "Thor13"
```




```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

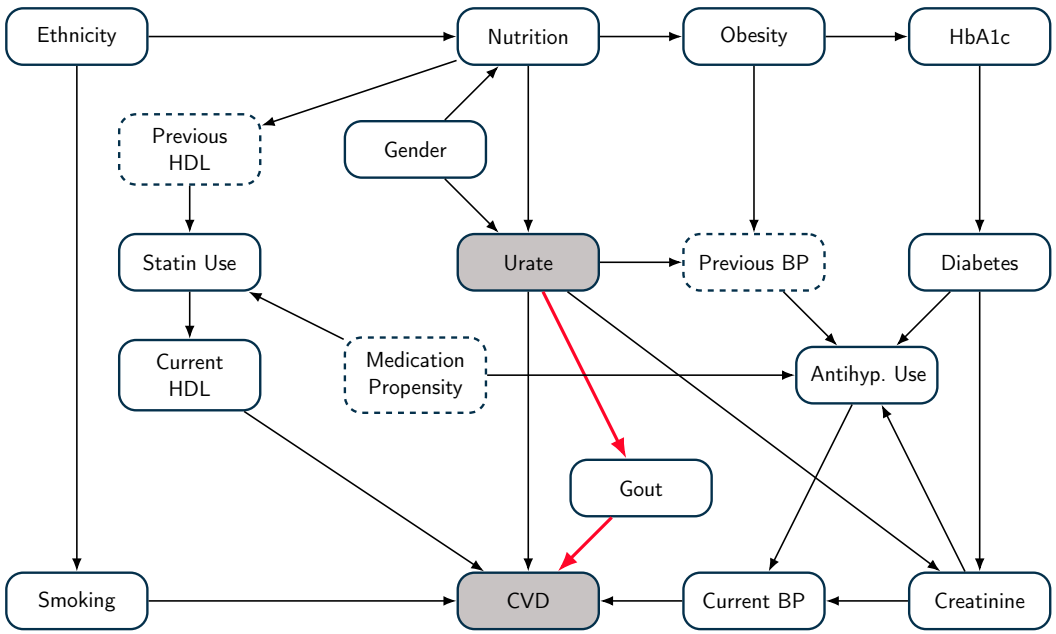
```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
```




```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

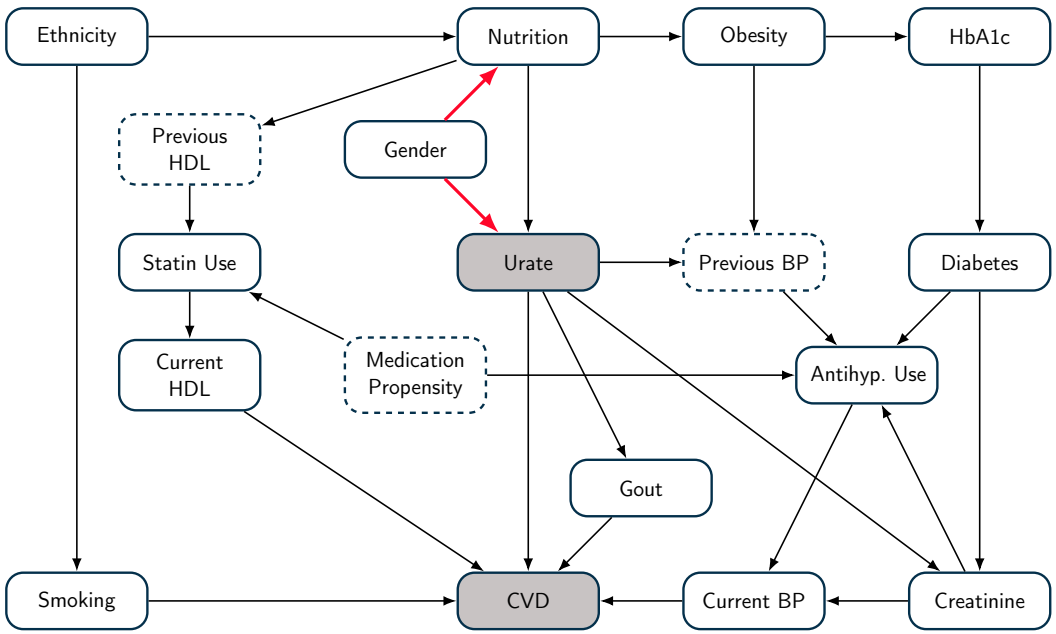
```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
```




```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

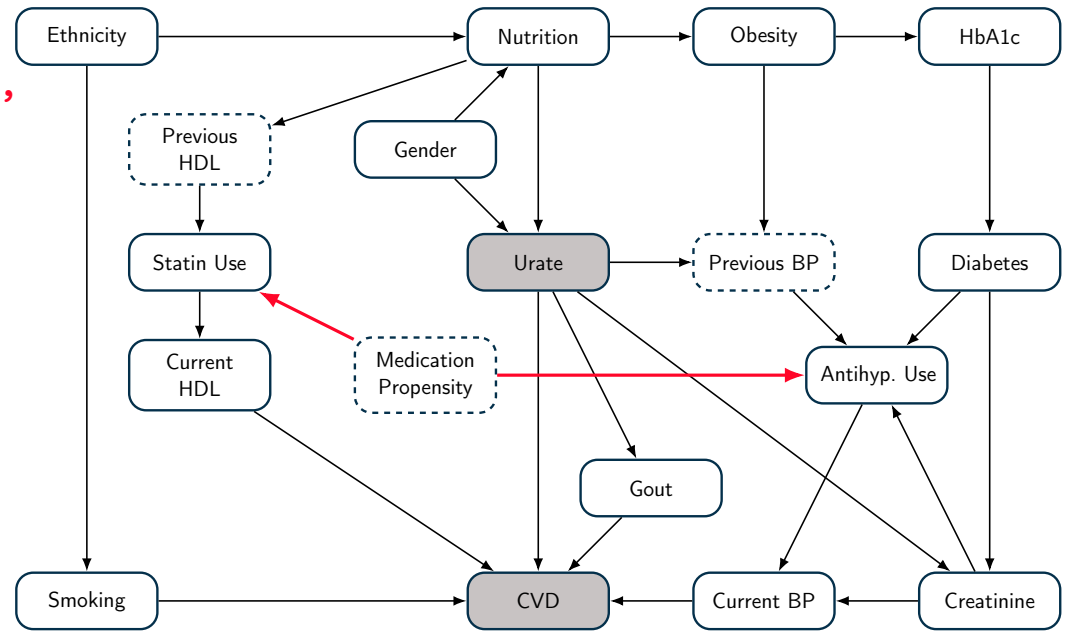
```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
    Gender => Nutrition Urate,
```




```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

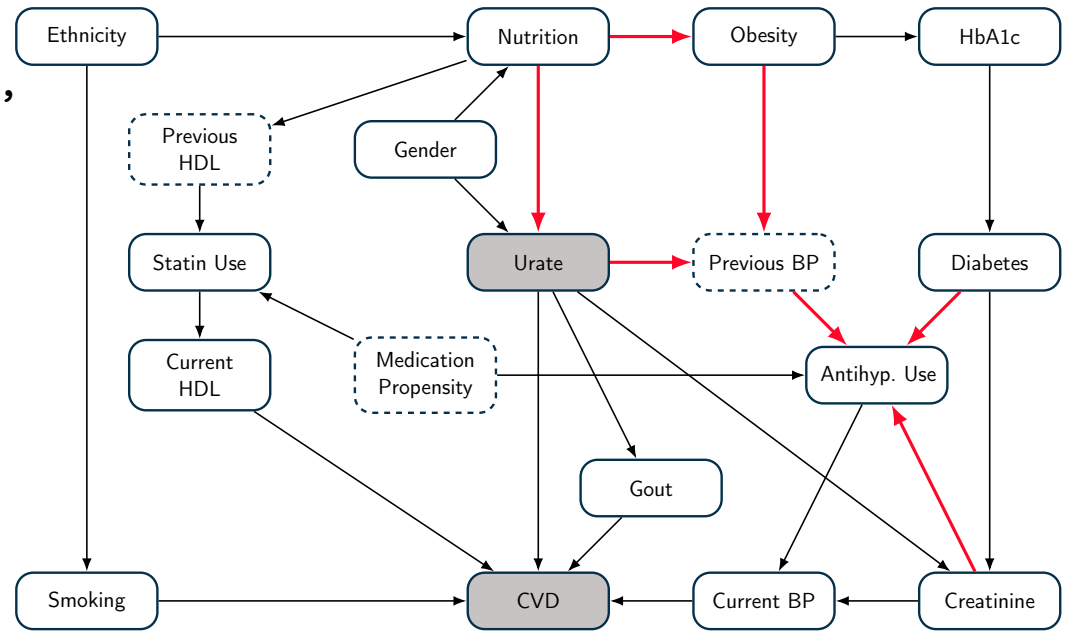
```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
    Gender => Nutrition Urate,
    StatinUse AntiHypUse <= MedProp,
```




```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

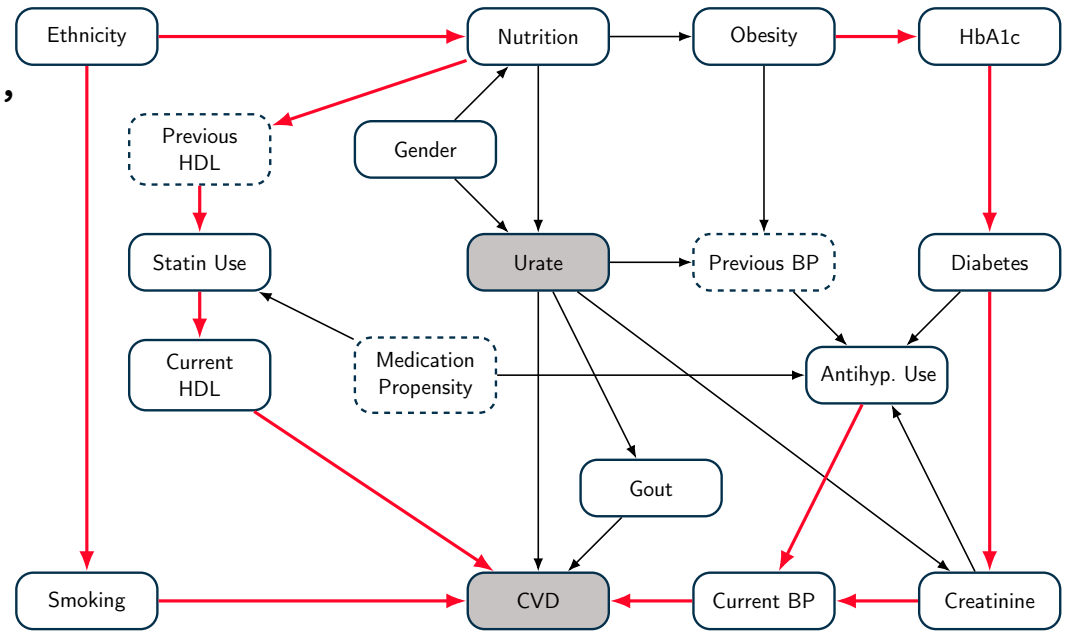
```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
    Gender => Nutrition Urate,
    StatinUse AntiHypUse <= MedProp,
    Nutrition => Obesity Urate =>
      PreviousBP => AntiHypUse <=
      Diabetes Creatinine,
```




```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
    Gender => Nutrition Urate,
    StatinUse AntiHypUse <= MedProp,
    Nutrition => Obesity Urate =>
      PreviousBP => AntiHypUse <=
      Diabetes Creatinine,
    Nutrition => PreviousHDL =>
       StatinUse => CurrentHDL =>
       CVD <= CurrentBP <=
       AntiHypUse Creatinine,
    Obesity => HbA1c => Diabetes
     => Creatinine,
    Nutrition <= Ethnicity;
    => Smoking => CVD;
```

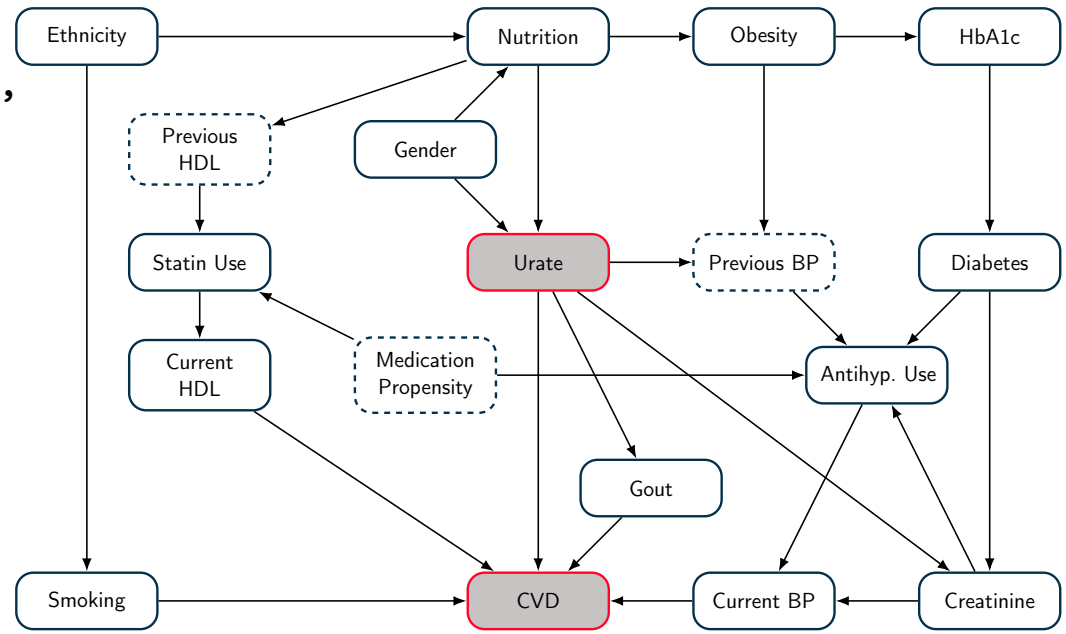



```
run;
```

Specifying an Analysis with the CAUSALGRAPH Procedure

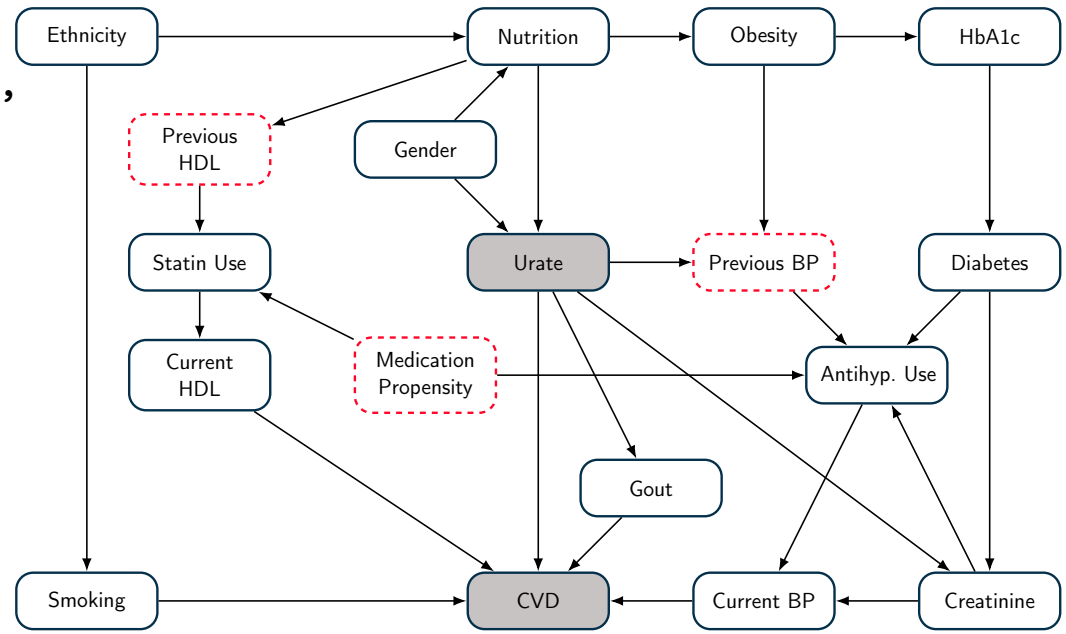
```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
    Gender => Nutrition Urate,
    StatinUse AntiHypUse <= MedProp,
    Nutrition => Obesity Urate =>
      PreviousBP => AntiHypUse <=
      Diabetes Creatinine,
    Nutrition => PreviousHDL =>
       StatinUse => CurrentHDL =>
       CVD <= CurrentBP <=
       AntiHypUse Creatinine,
    Obesity => HbA1c => Diabetes
     => Creatinine,
    Nutrition <= Ethnicity;
    => Smoking => CVD;
  identify Urate => CVD;


run;
```



Specifying an Analysis with the CAUSALGRAPH Procedure

```
proc causalgraph compact minimal;
  model "Thor13"
    Urate => CVD, Urate => Creatinine,
    Urate => Gout => CVD,
    Gender => Nutrition Urate,
    StatinUse AntiHypUse <= MedProp,
    Nutrition => Obesity Urate =>
      PreviousBP => AntiHypUse <=
      Diabetes Creatinine,
    Nutrition => PreviousHDL =>
       StatinUse => CurrentHDL =>
       CVD <= CurrentBP <=
       AntiHypUse Creatinine,
    Obesity => HbA1c => Diabetes
     => Creatinine,
    Nutrition <= Ethnicity;
    => Smoking => CVD;
  identify Urate => CVD;
  unmeasured PreviousBP PreviousHDL
   MedProp;
run;
```



Results from PROC CAUSALGRAPH

Minimal Adjustment Sets

Covariate Adjustment Sets for Thor13

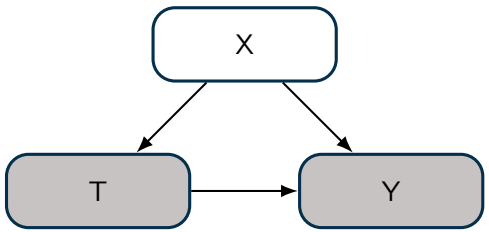
Causal Effect of Urate on CVD

			Covariates			
	Size	**Minimal**	**Ethnicity**	**Gender**	**Nutrition**	**Smoking**
1	2	Yes	*		*	
2	2	Yes		*	*	
3	2	Yes			*	*

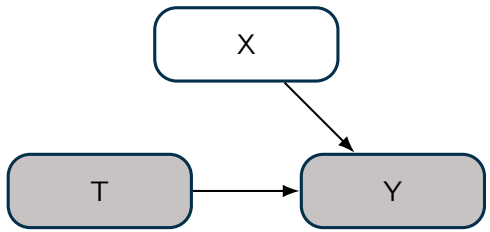
- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

Subtle Difficulties Arise in the Absence of Ideal Randomization

Nonrandomized Study

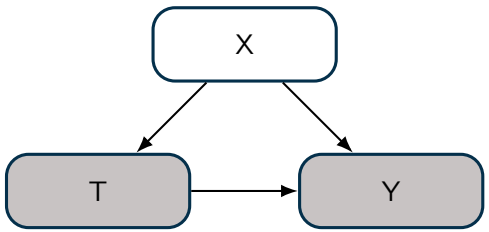


Randomized Study

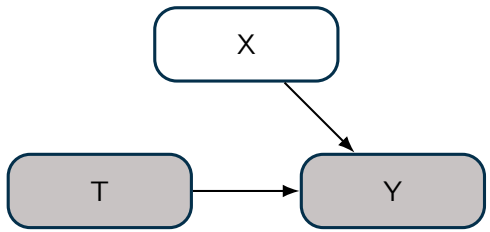


Subtle Difficulties Arise in the Absence of Ideal Randomization

Nonrandomized Study



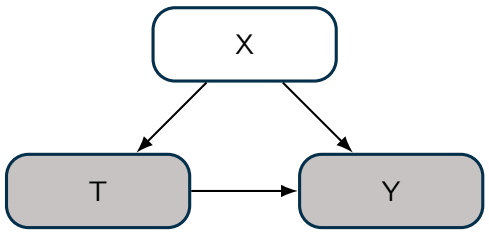
Randomized Study



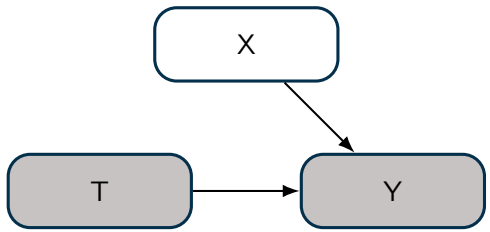
- Non-experimental studies
  - Observational study
  - Retrospective analysis

Subtle Difficulties Arise in the Absence of Ideal Randomization

Nonrandomized Study



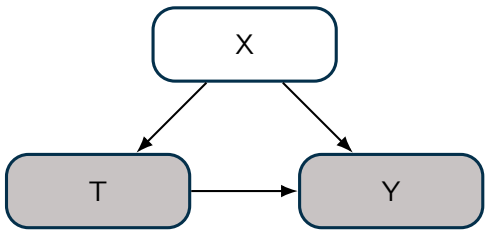
Randomized Study



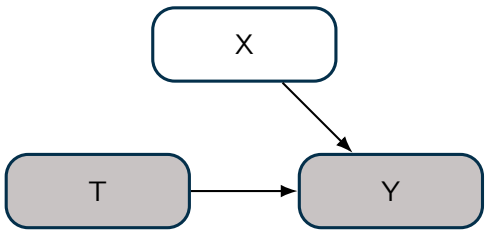
- Non-experimental studies
  - Observational study
  - Retrospective analysis
- Experimental studies
  - Noncompliance
  - Selective enrollment

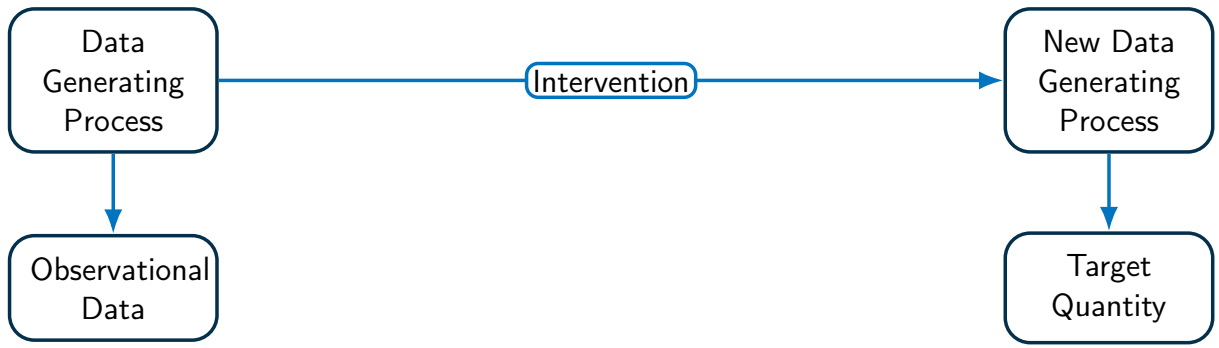
The Fundamental Challenge of Causal Analysis

Nonrandomized Study



Randomized Study



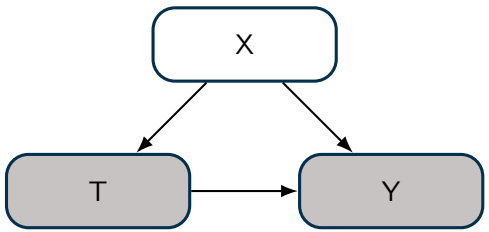


$P(T, X, Y)$

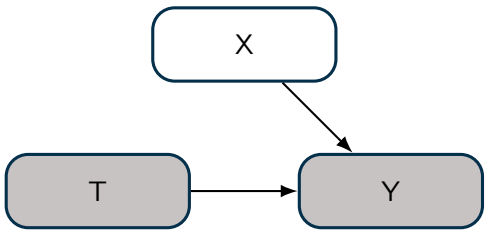
$P(Y|do(T))$

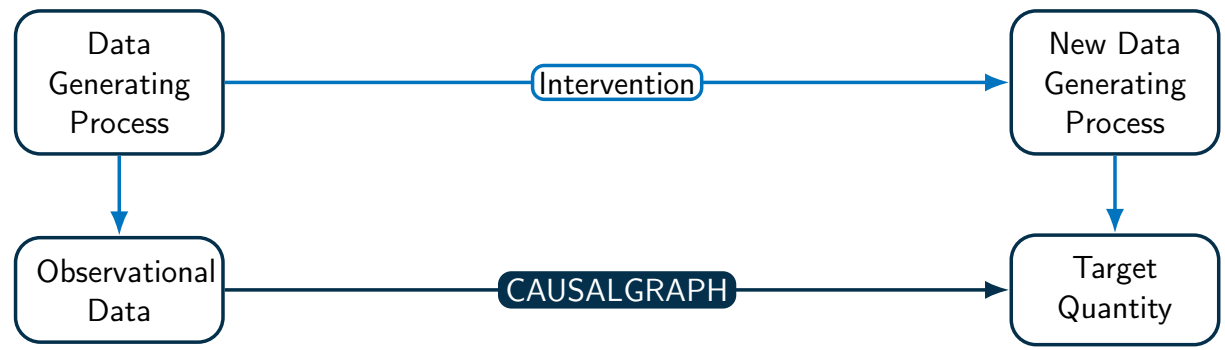
The Fundamental Challenge of Causal Analysis

Nonrandomized Study



Randomized Study





$P(T, X, Y)$

$P(Y|do(T))$

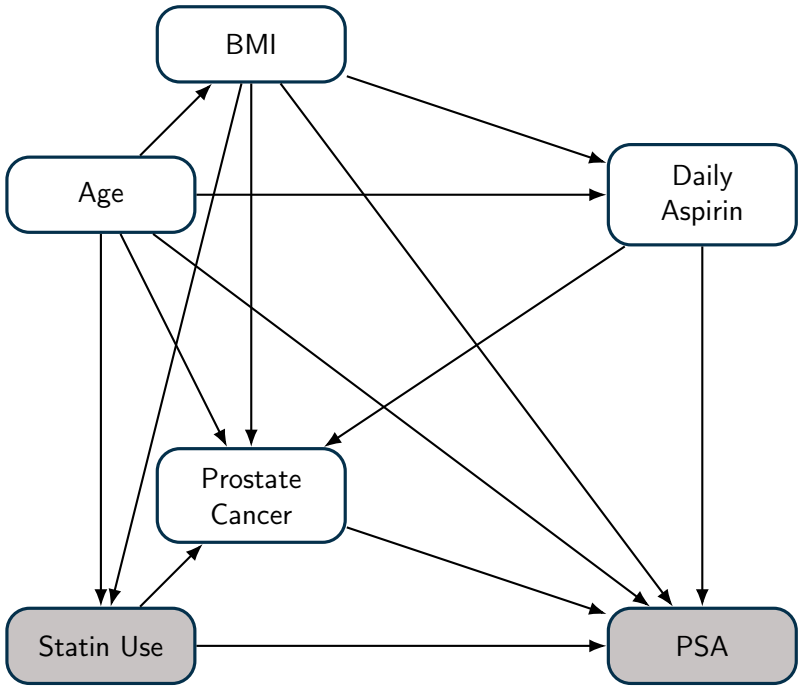
- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

Causal Analysis of Observational Data

A Causal Model for the Effect of Statin Use on Prostate-Specific Antigen

Adapted from Ferro et al. (2015)

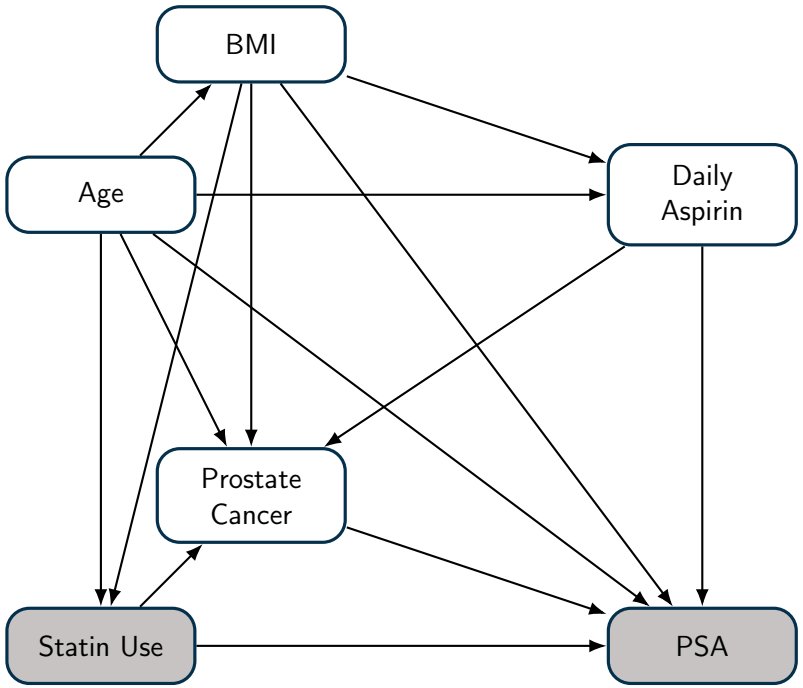
Obs	Age	BMI	Aspirin	StatinUse	Cancer	PSA
1	67.8219	26.5741	0	1	0	4.19232
2	56.3722	26.4821	0	0	0	4.92659
3	57.8478	26.4788	0	0	1	6.97768
4	74.1587	26.6207	1	0	1	6.88858
5	69.4976	26.3082	0	0	0	5.89829
6	71.6984	26.0927	1	0	0	4.87417
7	61.0182	26.8058	0	0	0	6.14052
8	69.9975	27.0862	0	0	0	4.87205
9	56.6626	26.4457	0	0	0	4.55604
10	56.7639	24.1488	0	0	1	5.45989



A Causal Model for the Effect of Statin Use on Prostate-Specific Antigen

Adapted from Ferro et al. (2015)

```
proc causalgraph;
  model "Statin Effect on PSA"
    BMI ==> Cancer StatinUse
      PSA Aspirin,
    StatinUse ==> Cancer PSA,
    Cancer ==> PSA,
    Age ==> BMI StatinUse
      Aspirin PSA Cancer,
    Aspirin ==> PSA Cancer;
  identify StatinUse ==> PSA;
run;
```



There Are Two Valid Adjustment Sets for the Causal Effect

True ATE: -0.608 ng/mL

```
proc causalgraph;
  model "Statin Effect on PSA"
    BMI ==> Cancer StatinUse
      PSA Aspirin,
    StatinUse ==> Cancer PSA,
    Cancer ==> PSA,
    Age ==> BMI StatinUse
      Aspirin PSA Cancer,
    Aspirin ==> PSA Cancer;
  identify StatinUse ==> PSA;
run;
```

Covariate Adjustment Sets for PSA Model

Causal Effect of StatinUse on PSA

			Covariates			
	Size	**Minimal**	**Age**	**Aspirin**	**BMI**	**Cancer**
1	2	Yes	*		*	
2	3	No	*	*	*	

There Are Two Valid Adjustment Sets for the Causal Effect

True ATE: -0.608 ng/mL

```
proc causalgraph;
  model "Statin Effect on PSA"
    BMI ==> Cancer StatinUse
      PSA Aspirin,
    StatinUse ==> Cancer PSA,
    Cancer ==> PSA,
    Age ==> BMI StatinUse
      Aspirin PSA Cancer,
    Aspirin ==> PSA Cancer;
  identify StatinUse ==> PSA;
run;

proc causaltrt data=StatinPSA;
  class StatinUse Cancer
Aspirin;
  psmodel StatinUse(event='1')
    = Age BMI;
  model PSA;
run;
```

Covariate Adjustment Sets for PSA Model

Causal Effect of StatinUse on PSA

			Covariates		
Size	**Minimal**	**Age**	**Aspirin**	**BMI**	**Cancer**
1	2 Yes	*		*	
2	3 No	*	*	*	

Analysis of Causal Effect

Parameter	Treatment Level	Estimate	Robust Std Err	Wald 95% Confidence Limits		Z	Pr > |Z|
POM	**1**	5.3265	0.0289	5.2698	5.3832	184.17	<.0001
POM	**0**	5.9585	0.0147	5.9297	5.9874	405.20	<.0001
ATE		-0.6320	0.0320	-0.6947	-0.5693	-19.76	<.0001

There Are Two Valid Adjustment Sets for the Causal Effect

True ATE: -0.608 ng/mL

```
proc causalgraph;
  model "Statin Effect on PSA"
    BMI ==> Cancer StatinUse
      PSA Aspirin,
    StatinUse ==> Cancer PSA,
    Cancer ==> PSA,
    Age ==> BMI StatinUse
      Aspirin PSA Cancer,
    Aspirin ==> PSA Cancer;
  identify StatinUse ==> PSA;
run;

proc causaltrt data=StatinPSA;
  class StatinUse Cancer
Aspirin;
  psmodel StatinUse(event='1')
    = Age Aspirin BMI;
  model PSA;
run;
```

Covariate Adjustment Sets for PSA Model

Causal Effect of StatinUse on PSA

			Covariates		
Size	**Minimal**	**Age**	**Aspirin**	**BMI**	**Cancer**
1 2	Yes	*		*	
2 3	No	*	*	*	

Analysis of Causal Effect

Parameter	Treatment Level	Estimate	Robust Std Err	Wald 95% Confidence Limits		Z	Pr > |Z|
POM	**1**	5.3349	0.0285	5.2791	5.3907	187.37	<.0001
POM	**0**	5.9556	0.0146	5.9271	5.9841	409.08	<.0001
ATE		-0.6207	0.0311	-0.6817	-0.5597	-19.93	<.0001

There Are Two Valid Adjustment Sets for the Causal Effect

True ATE: -0.608 ng/mL

```
proc causalgraph;
  model "Statin Effect on PSA"
    BMI ==> Cancer StatinUse
      PSA Aspirin,
    StatinUse ==> Cancer PSA,
    Cancer ==> PSA,
    Age ==> BMI StatinUse
      Aspirin PSA Cancer,
    Aspirin ==> PSA Cancer;
  identify StatinUse ==> PSA;
run;

proc ttest data=StatinPSA
order=data;
  class StatinUse;
  var PSA;
run;
```

Covariate Adjustment Sets for PSA Model

Causal Effect of StatinUse on PSA

			Covariates		
Size	**Minimal**	**Age**	**Aspirin**	**BMI**	**Cancer**
1 2	Yes	*		*	
2 3	No	*	*	*	

StatinUse	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		5.4194	5.3670	5.4717	0.9155	0.8800	0.9541
0		5.9308	5.9019	5.9598	0.9139	0.8939	0.9349
Diff (1-2)	Pooled	-0.5115	-0.5712	-0.4517	0.9143	0.8967	0.9326
Diff (1-2)	Satterthwaite	-0.5115	-0.5713	-0.4516			

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

Causal Graph Analysis with the CAUSALGRAPH Procedure

New in SAS/STAT 15.1

- Many real-world studies are nonrandomized
  - Causal analysis is still an important task
  - Subtle difficulties arise in such analyses
- The CAUSALGRAPH procedure is a powerful tool to facilitate such analyses
  - Articulate a set of causal assumptions
  - Build a causal model
  - Test model assumptions
  - Assess the identifiability of a causal effect
- CAUSALGRAPH results are easily incorporated into an analysis workflow

Important Features of the CAUSALGRAPH Procedure

- Define causal models in the form of **directed acyclic graphs** (DAGs)
- Specify **unmeasured**, **latent**, or **excluded** variables
- Multiple **criteria** for identifying causal effects
  - Constructive backdoor criterion (Van der Zander, Liskiewicz, and Textor, 2014)
  - Backdoor criterion (Pearl, 2009)
  - Instrumental variables criterion (Van der Zander, Liskiewicz, and Textor, 2015)
- **List** valid identification strategies or **test** a user-proposed strategy

Additional Features of the CAUSALGRAPH Procedure

- Explore **causal and noncausal paths** between treatments and outcomes
- Enumerate observationally **testable assumptions** in a model
- Analyze any static treatment regime
  - **Multiple treatments**, **multiple outcomes**
- Analyze **multiple models** simultaneously

Outline

- Overview
- Restricted Mean Survival Time (RMST)
  - Introduction
  - RMST Regression Analysis
- Causal Analysis of Observational Data
  - What is Causal Analysis?
  - Example 1: Finding Valid Adjustment Sets
  - Randomized vs. Nonrandomized (Observational) Studies
  - Example 2: Using Observational Data to Estimate a Causal Effect
  - Summary Remarks
- Additional Functionalities

New Features in SAS/STAT 15.1

- PROC ICPHREG: semi-parametric PH model to interval-censored data

  `BASE=UNSPECIFIED`

New Features in SAS/STAT 15.1

- PROC ICPHREG: semi-parametric PH model to interval-censored data

  `BASE=UNSPECIFIED`

- Regression Procedures: Internal spline knots

New Features in SAS/STAT 15.1

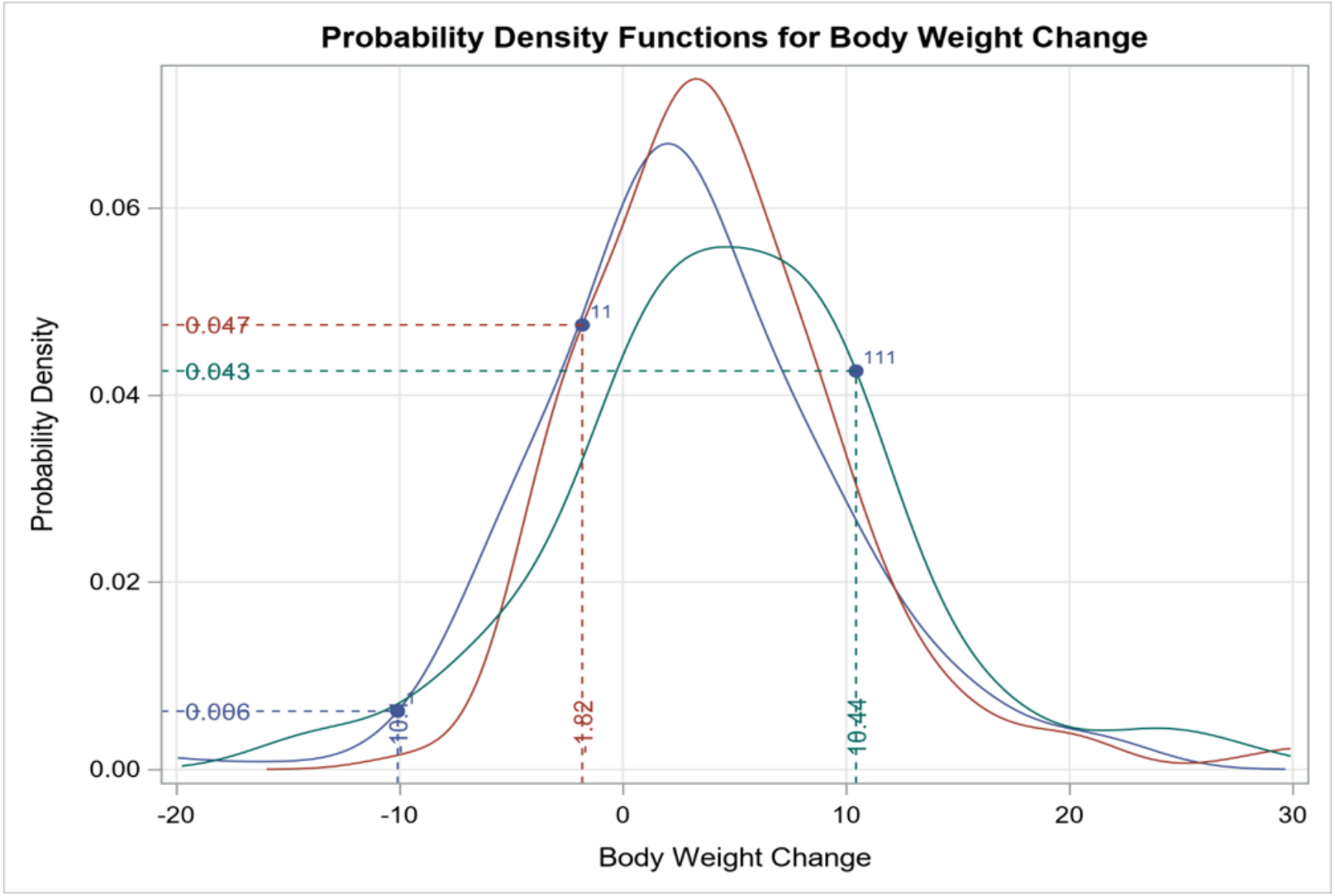
- PROC ICPHREG: semi-parametric PH model to interval-censored data

  `BASE=UNSPECIFIED`

- Regression Procedures: Internal spline knots
- PROC FREQ and PROC SURVEYFREQ: Estimates and confidence limits for sensitivity and specificity

  SENSPEC option in the TABLE Statement

Conditional Distribution Analysis (QUANTREG)



Additional highlights of SAS/STAT 15.1

Procedure	New Feature
CAUSALMED	WEIGHT Statement
PSMATCH	PSWEIGHT Statement
MCMC & NLMIXED	Enhanced compartment model capabilities
Survey ROCs	New options for replication methods
TTEST	Bootstrap graphics