

Causal Learning and Inference: A Practical Guide — Part II

Adèle H. Ribeiro

Institute of Medical Informatics
University of Münster

<https://adele.github.io/>

adele.ribeiro@uni-muenster.de

Júlia M. P. Soler

Institute of Mathematics and Statistics
University of São Paulo

<https://www.ime.usp.br/pavan/>

pavan@ime.usp.br

69th Annual Meeting of the
Brazilian Region of the International Society of Biometrics (RBras)
August 4-5, 2025



FAPESP, Process #2022/11963-3



Federal Ministry
of Education
and Research

BMBF, Process #BD605629

What We Covered on Day 1:

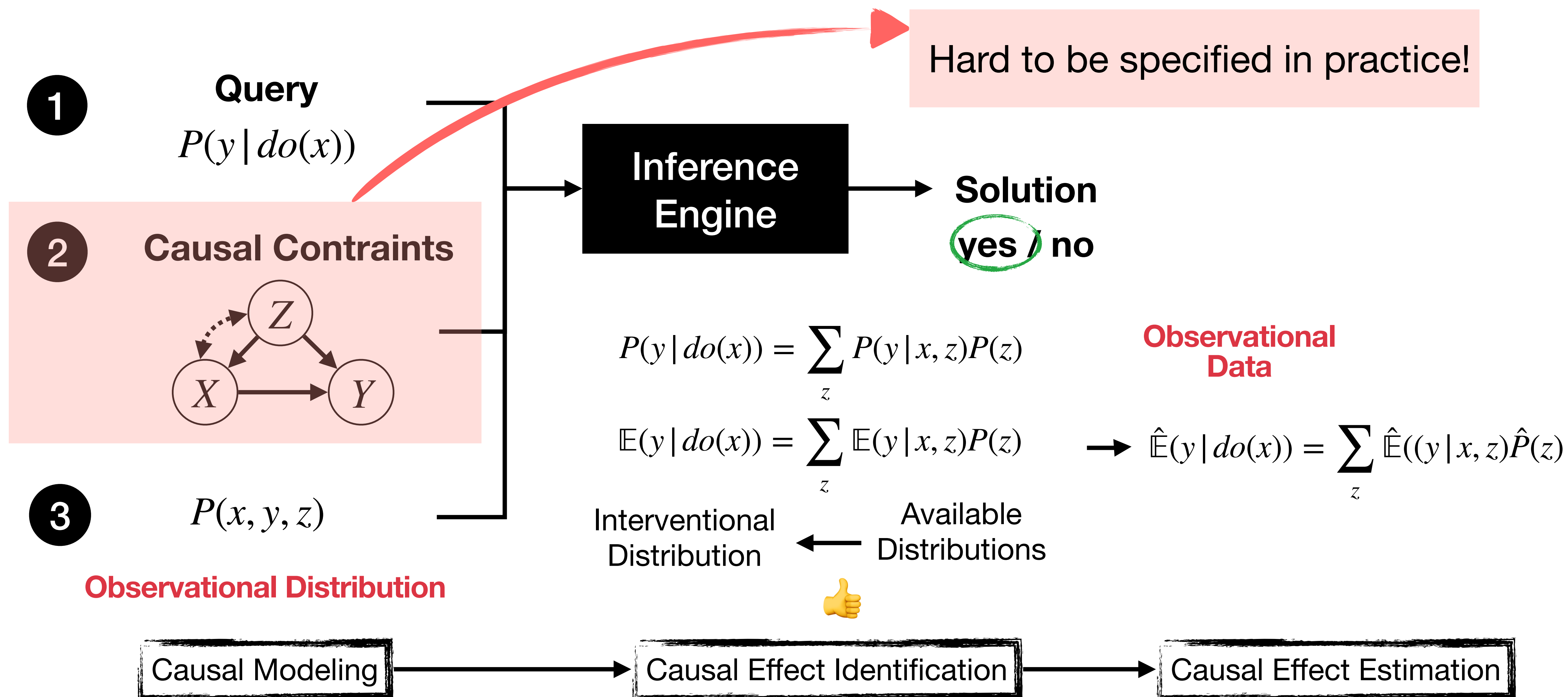
Classical causal inference pipeline from a causal diagram:

Causal modeling: Structural Causal Models and Causal Bayesian Networks / d-Separation

Effect identification given a **causal diagram**: Backdoor Adjustment & ID Algorithm

Coding Exercises

Classical Causal Pipeline - Judea Pearl's framework



Outline



Can we infer the causal diagram from
observational data?

Day 2:

Causal discovery: from **heterogenous** data types; **independent** and **dependent** observations

Effect identification given a **partial ancestral graph (PAG)**: Generalized Adjustment & PAG-ID

Coding Exercises

Markov Equivalence Class

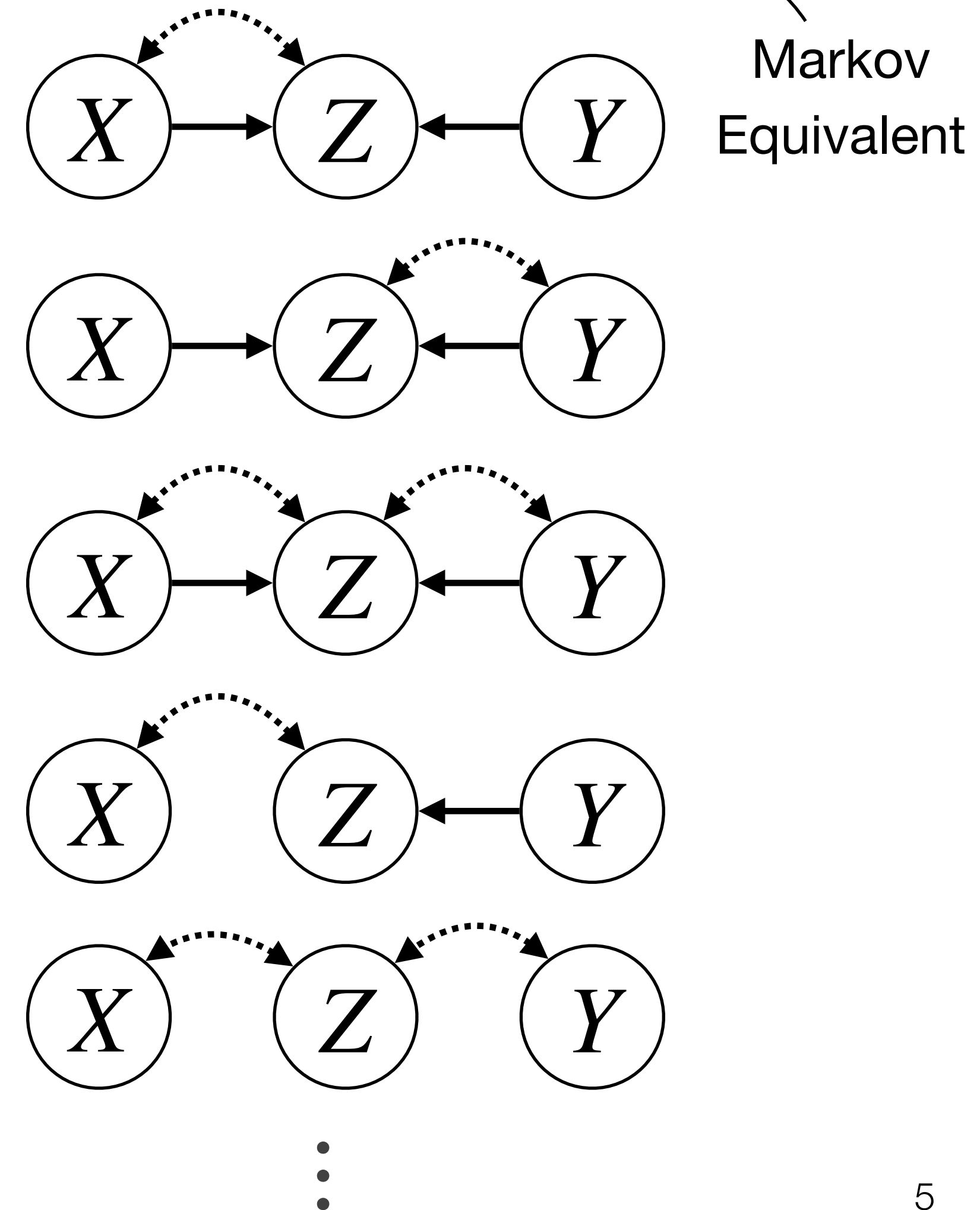
Distribution

$P(X, Y, Z)$
 with $P(Y|X) = P(Y)$
 i.e., $X \perp\!\!\!\perp Y$

Factorization

$$\begin{aligned} P(x, y, z) &= P(z|x, y)P(x|y)P(y) \\ &= P(z|x, y)P(x)P(y) \end{aligned}$$

Bayesian Networks



Markov Equivalence Class

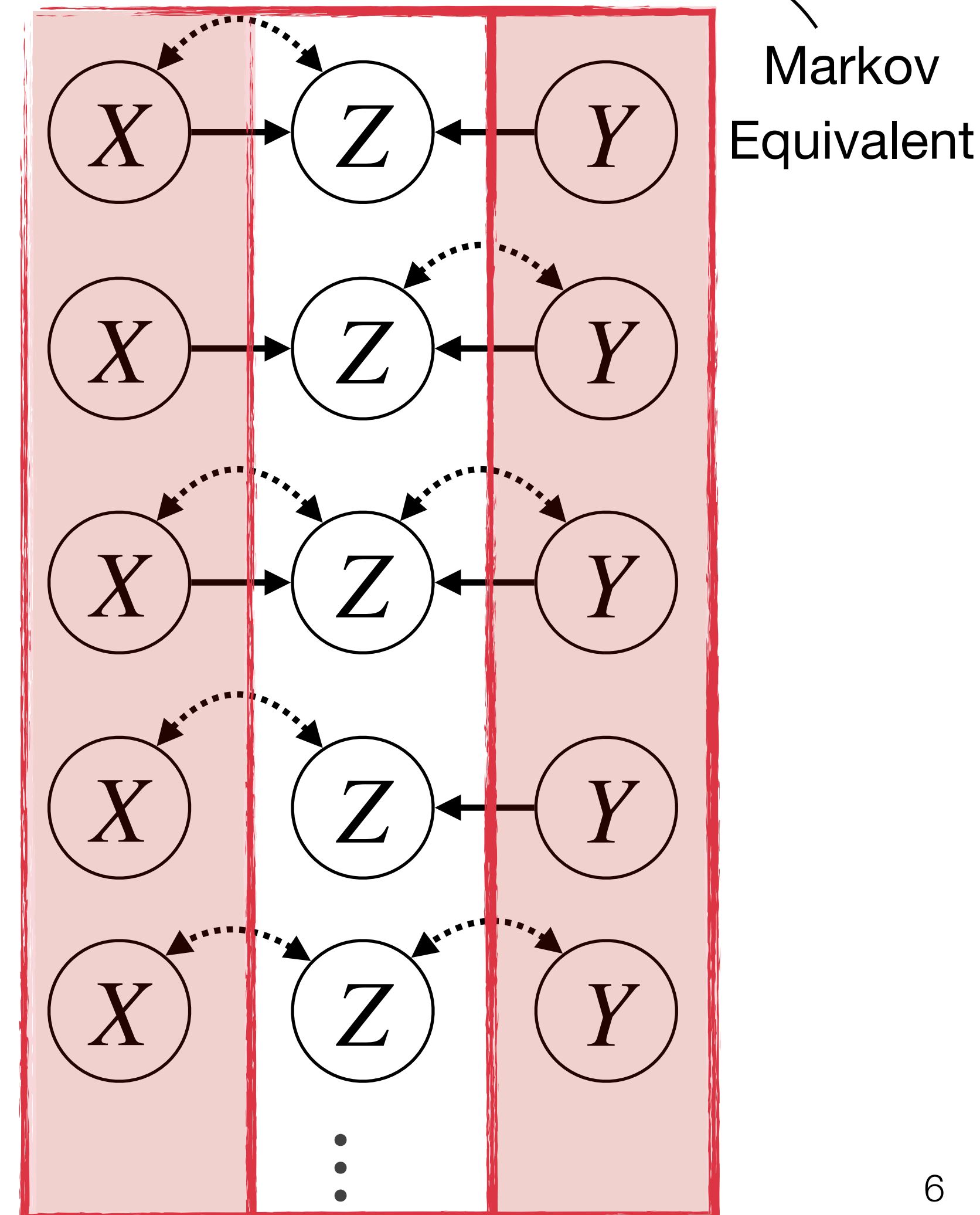
Distribution

$P(X, Y, Z)$
 with $P(Y|X) = P(Y)$
 i.e., $X \perp\!\!\!\perp Y$

Factorization

$$P(x, y, z) = P(z|x, y)P(x|y)P(y) \\ = P(z|x, y)P(x)P(y)$$

Bayesian Networks



All models imply *only* $X \perp\!\!\!\perp Y$ and
 Z is always a *collider* in such models,
 Note: Z is never an ancestor of X or Y

Learning the Markov Equivalence Class

Causal Discovery:

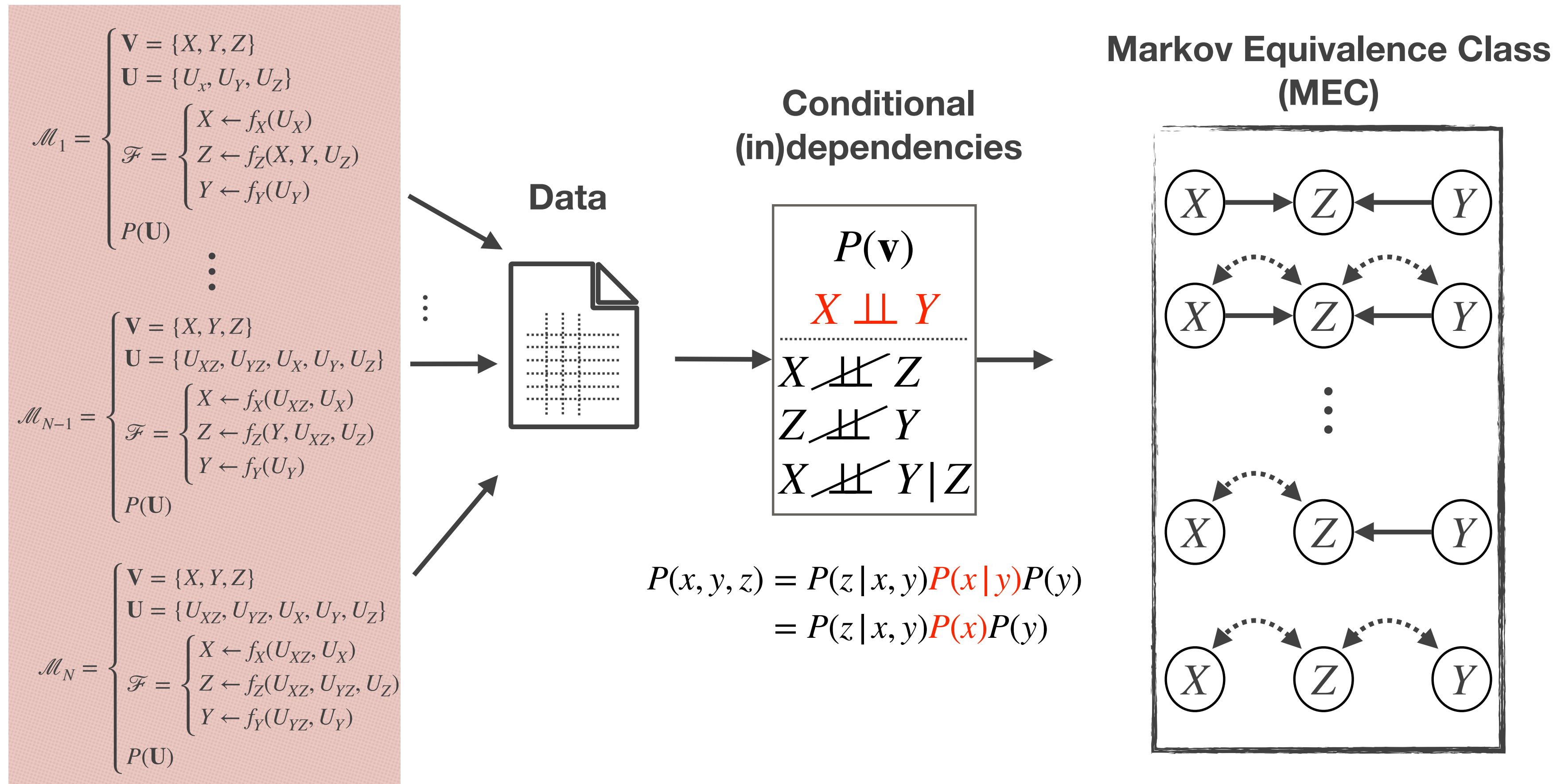
Many models are statistically indistinguishable without additional parametric / distributional assumptions.

In non-parametric settings, causal discovery algorithms can only learn a graphical representation of its *Markov equivalence class* (MEC)!

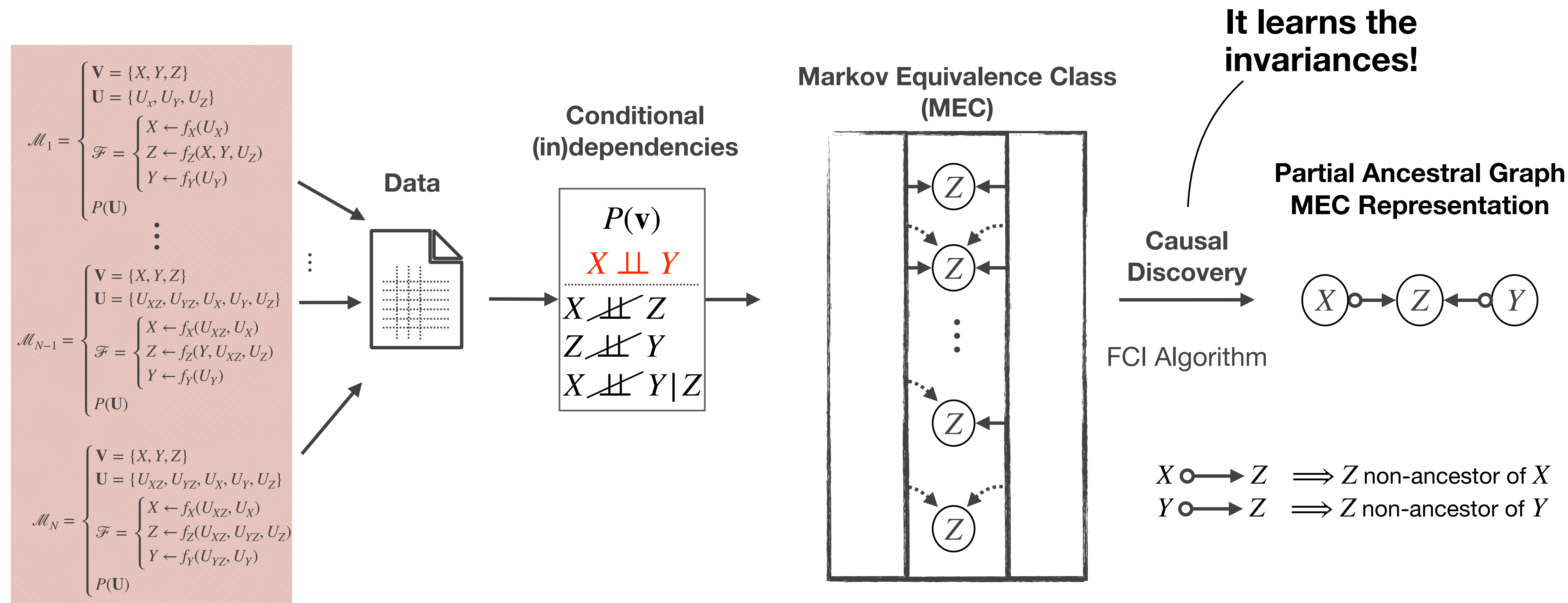
Fast Causal Inference (FCI): Sound and complete causal discovery algorithm, even in the presence of unobserved confounders and selection bias.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. [Link](#)

Constraint-Based Causal Discovery Algorithms



Constraint-Based Causal Discovery Algorithms



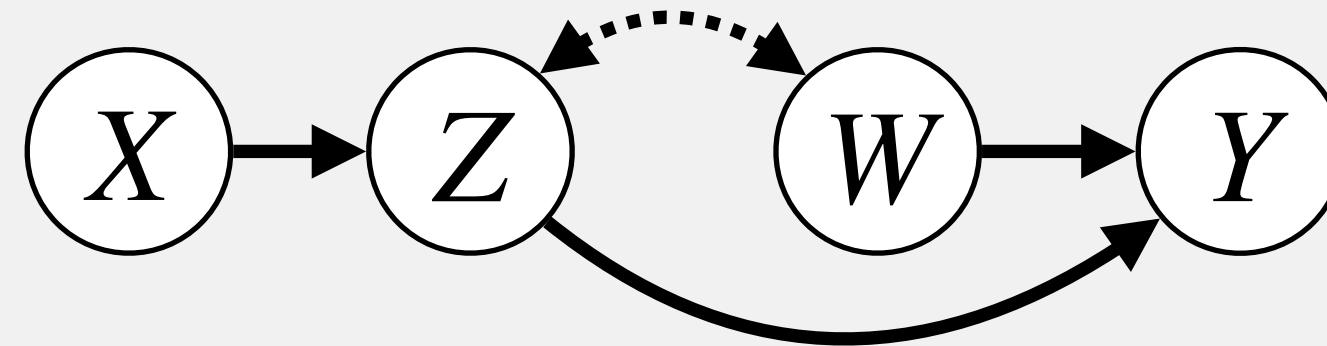
Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896. [Link](#)

9

FCI Algorithm - Pipeline

Unknown Reality

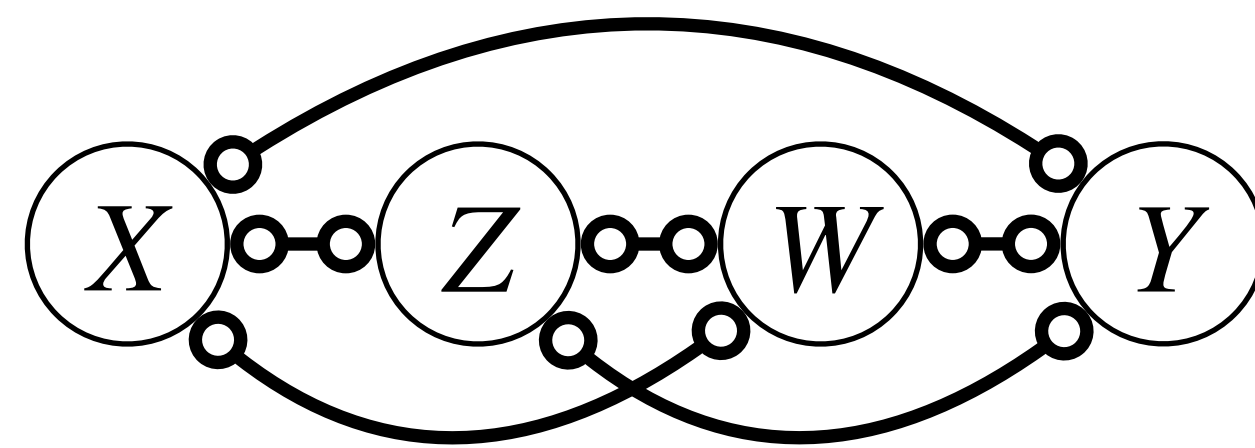
True causal diagram



$$X \perp\!\!\!\perp W$$

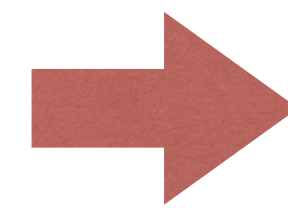
$$X \perp\!\!\!\perp Y | Z, W$$

Implied by the ADMG using d-separation



Complete Graph

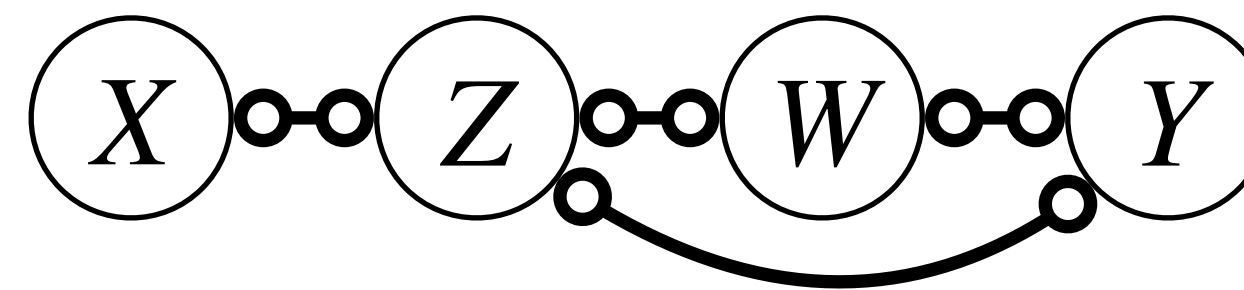
Conditional Independence Tests



$$X \perp\!\!\!\perp W$$

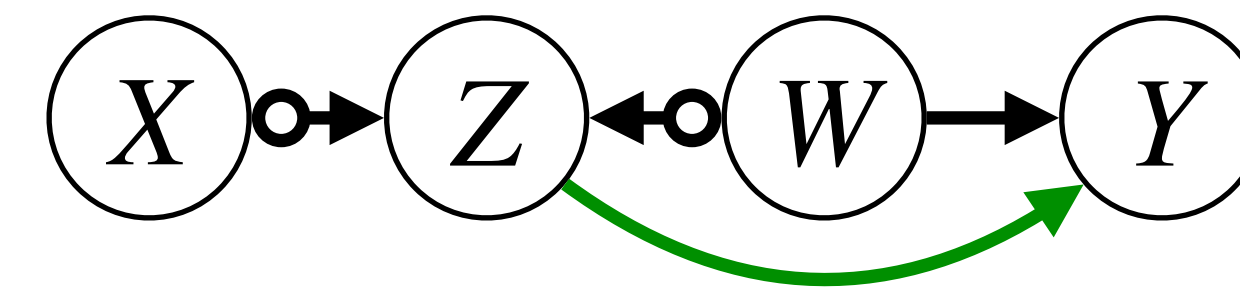
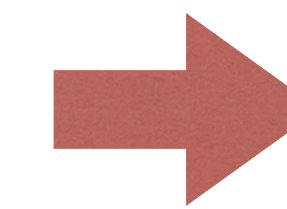
$$X \perp\!\!\!\perp Y | Z, W$$

By **faithfulness**, are correctly observed in the data



Skeleton

FCI Rules
(R0) + (R1 to R10)



Partial Ancestral Graph (PAG)

Implied by the PAG using m-separation

$$X \perp\!\!\!\perp W$$

$$X \perp\!\!\!\perp Y | Z, W$$

$A \circ \rightarrow B \implies B$ non-ancestor of A

$A \rightarrow B \implies A$ ancestor of B

$A \leftrightarrow B \implies$ spurious association

$A - B \implies$ selection bias

Z is not an ancestor of X or W .

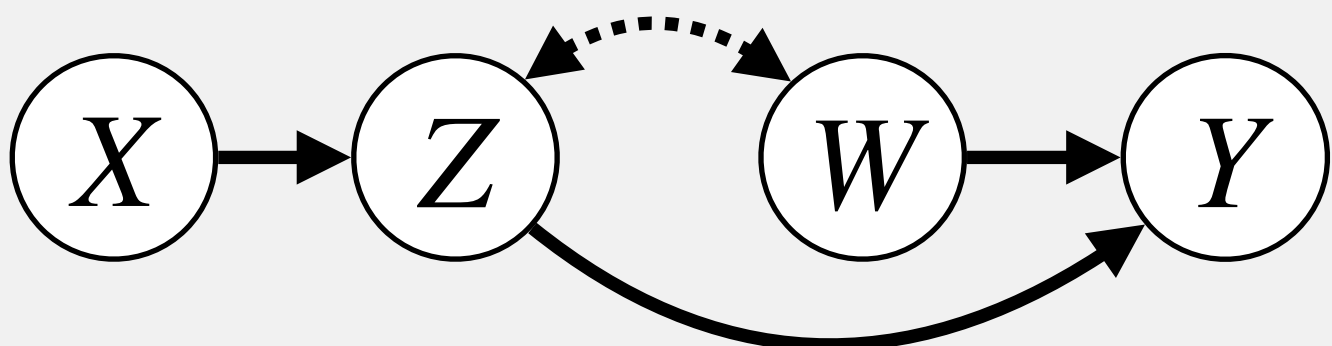
Z and W are ancestors of Y .

Z is not confounded with Y .

Applying FCI's Rules

Unknown Reality

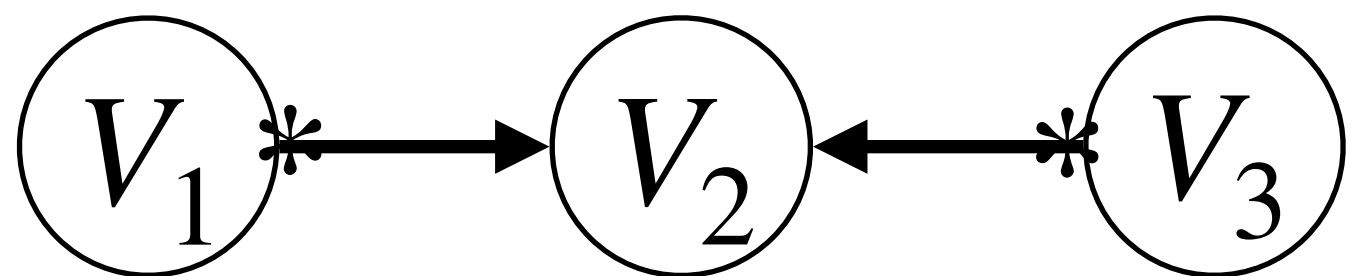
True causal diagram



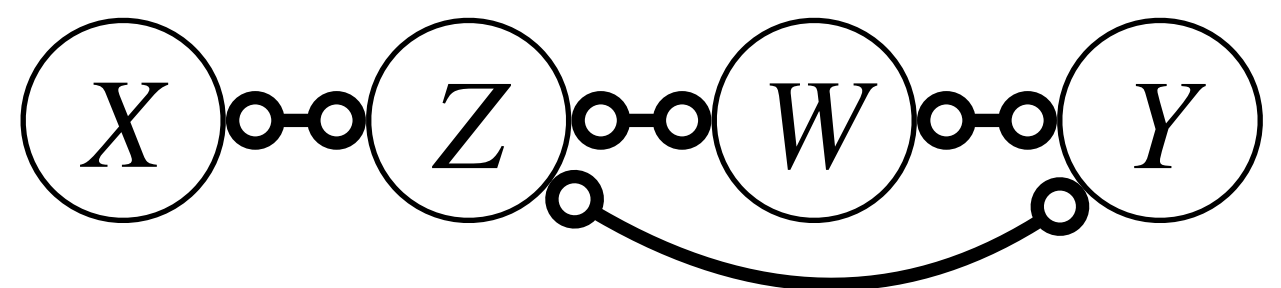
$X \perp\!\!\!\perp W$
 $X \perp\!\!\!\perp Y | Z, W$

Implied by the ADMG using d-separation

R0: If $\langle V_1, V_2, V_3 \rangle$ is unshielded and $V_2 \notin \text{Sepset}(V_1, V_3)$, then

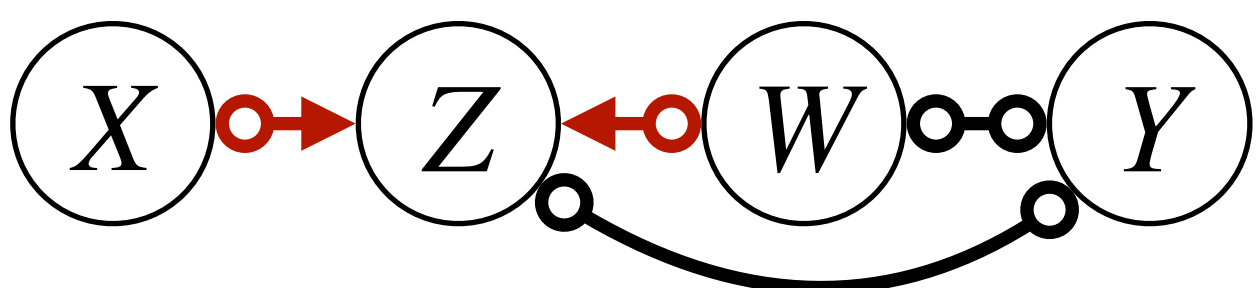
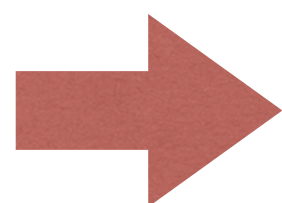


That is the only way for the path between V_1 and V_3 to be blocked when not conditioning on V_2



Skeleton

Applying R0:

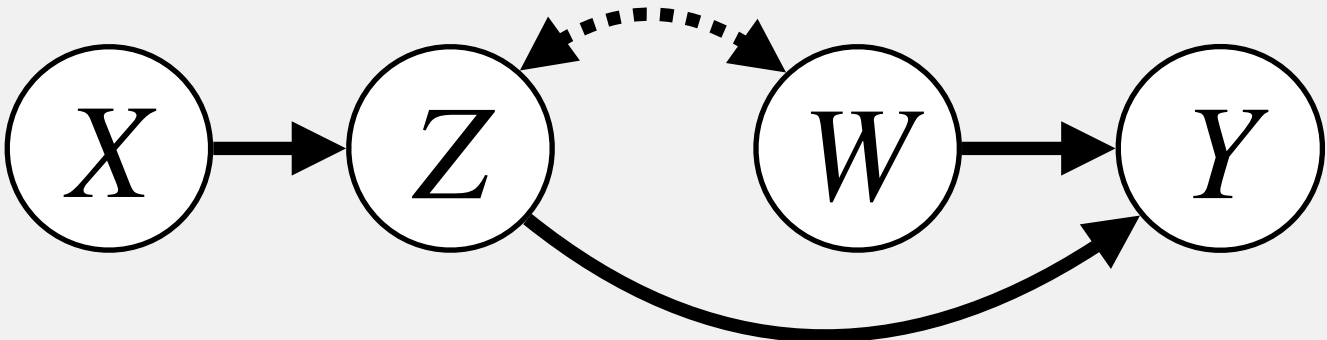


$X \perp\!\!\!\perp W$, where $Z \notin \text{Sepset}(X, W)$
 $X \perp\!\!\!\perp Y | Z, W$, where $Z \in \text{Sepset}(X, Y)$

Applying FCI's Rules

Unknown Reality

True causal diagram

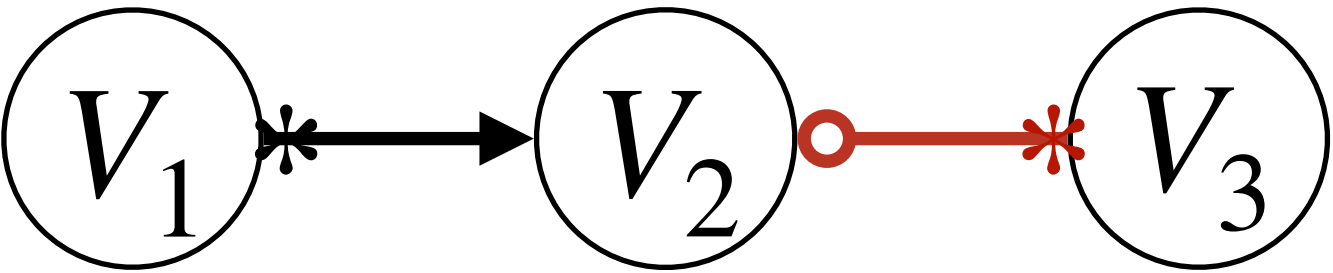


$$X \perp\!\!\!\perp W$$

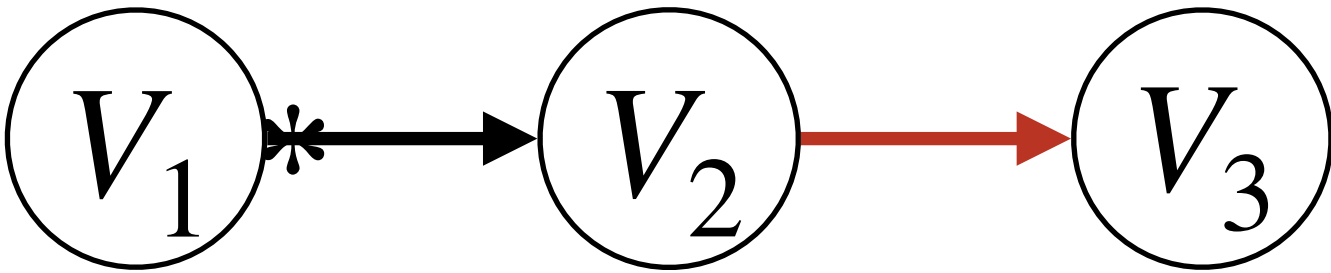
$$X \perp\!\!\!\perp Y \mid Z, W$$

Implied by the ADMG using d-separation

R1:

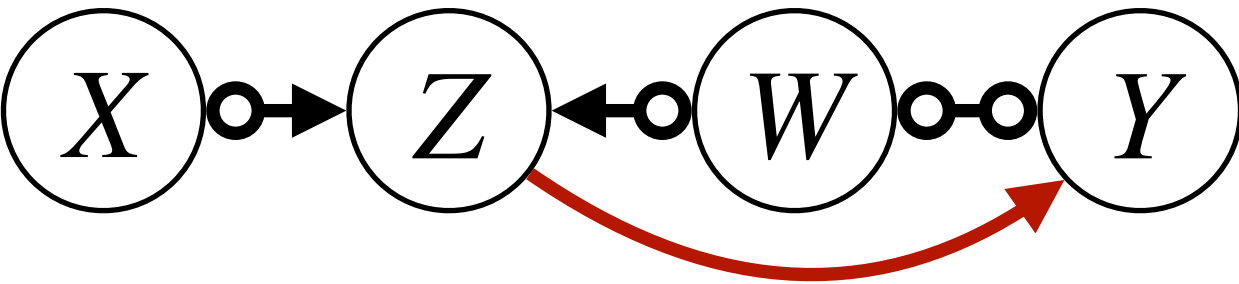
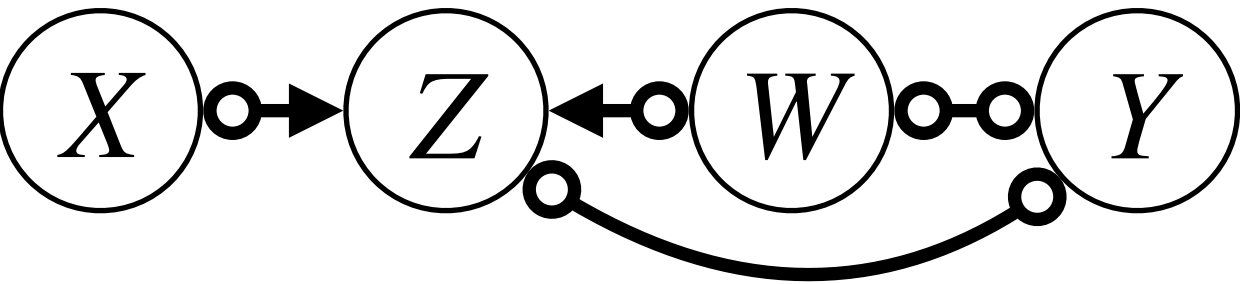
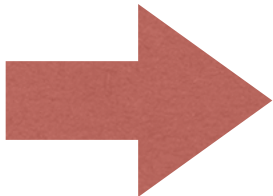


=>



where V_1 and V_3 are not adjacent

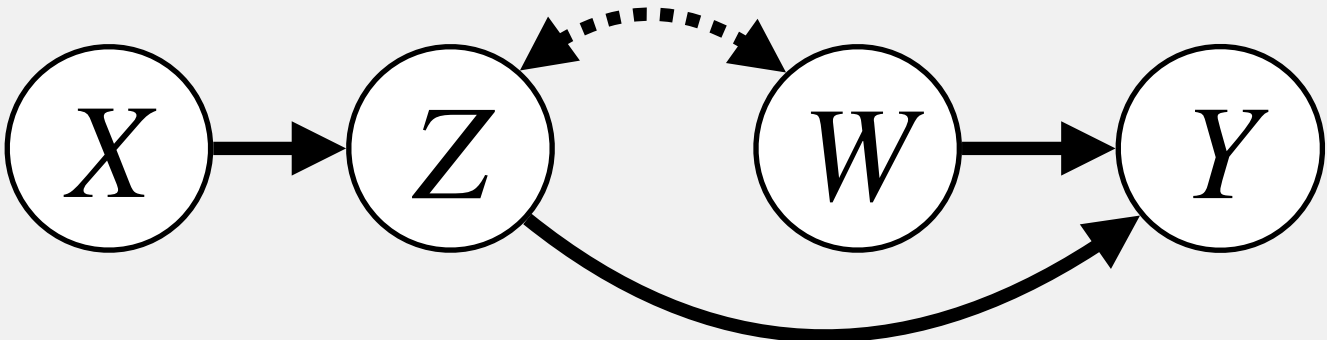
FCI's R1:



Applying FCI's Rules

Unknown Reality

True causal diagram

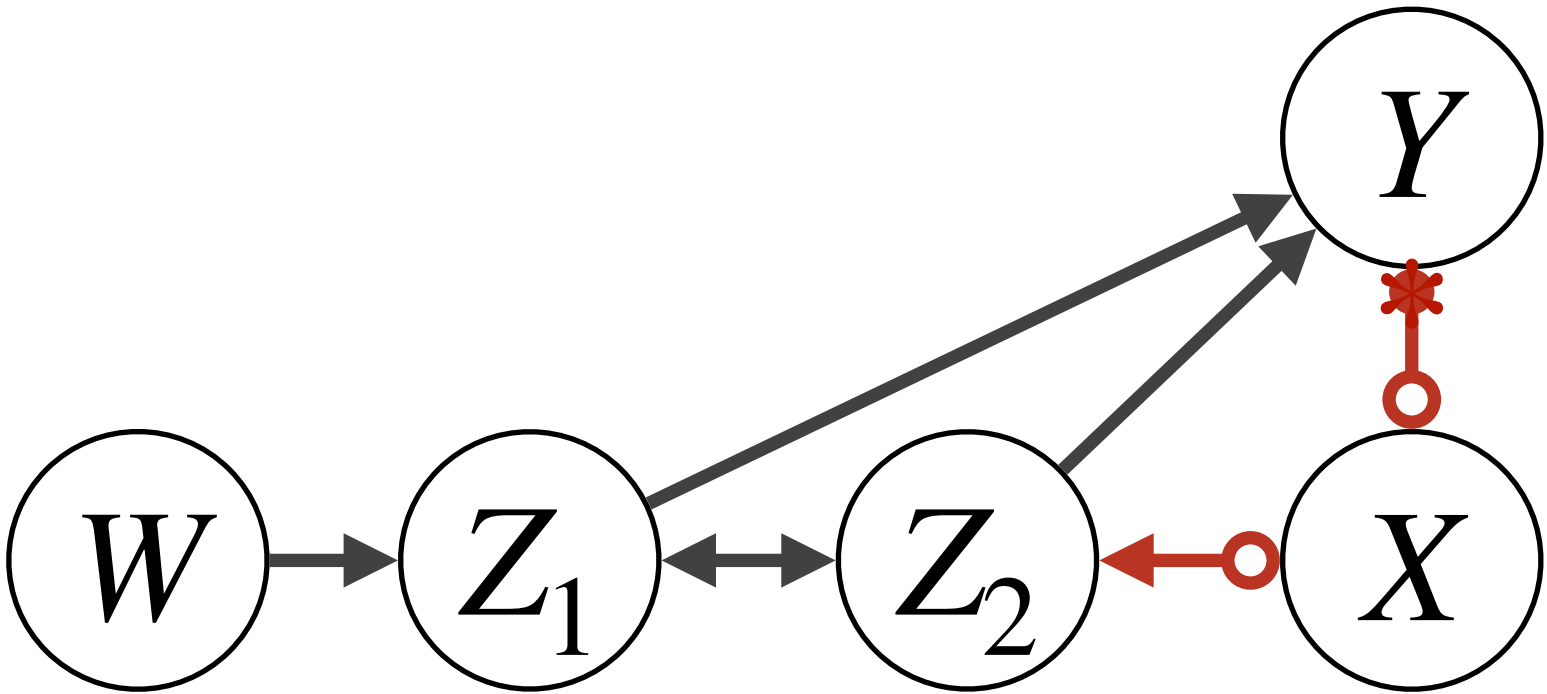


$$X \perp\!\!\!\perp W$$

$$X \perp\!\!\!\perp Y \mid Z, W$$

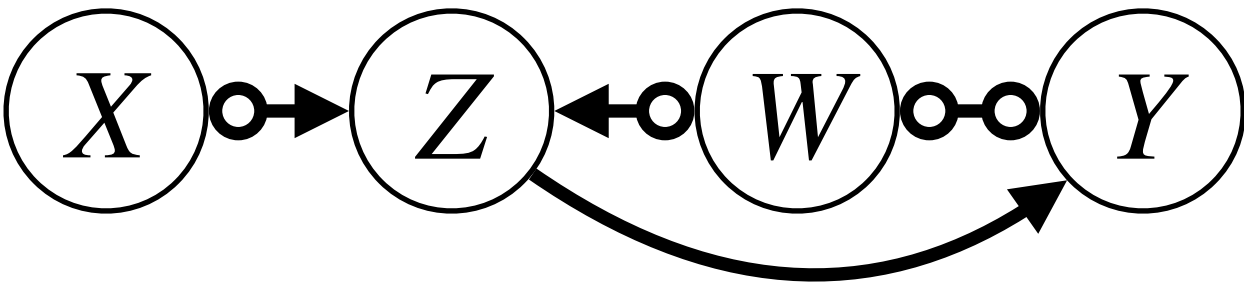
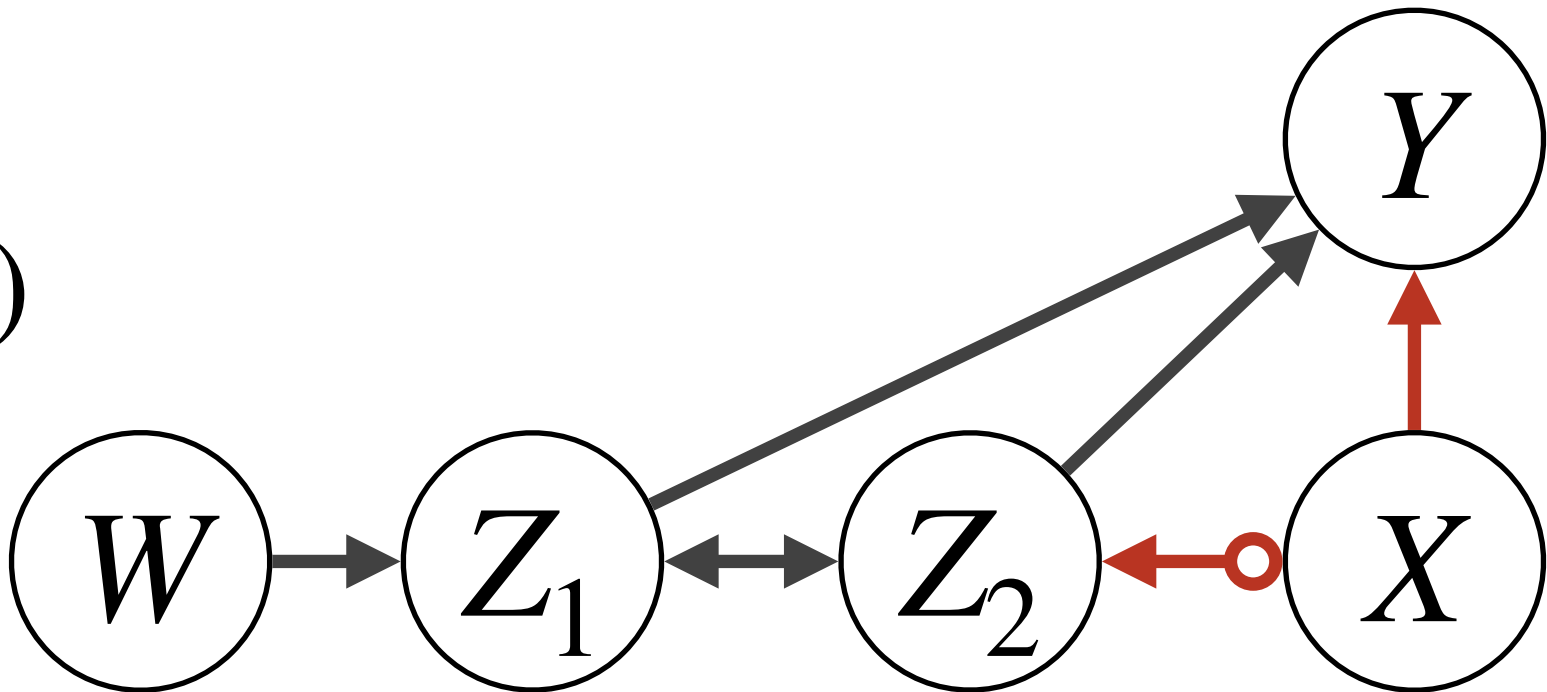
Implied by the ADMG using d-separation

R4:

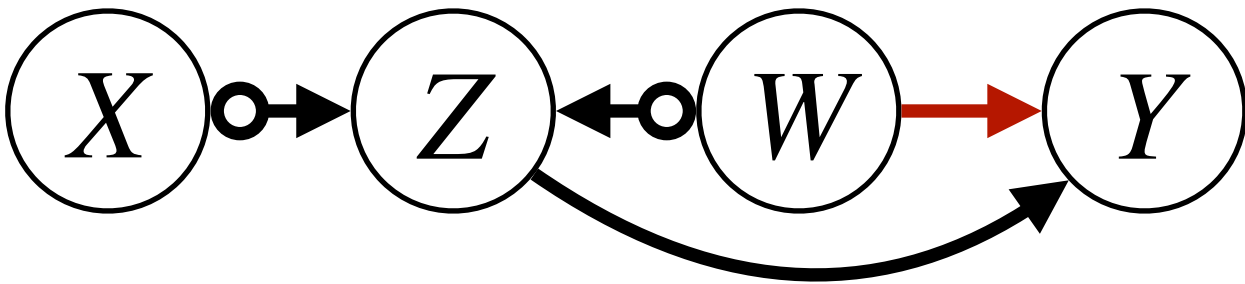
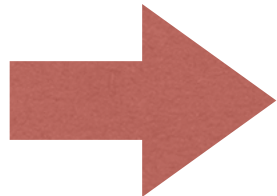


$\langle W, Z_1, Z_2, X, Y \rangle$ is a discriminating path for X

$$X \in \text{Sepset}(W, Y)$$



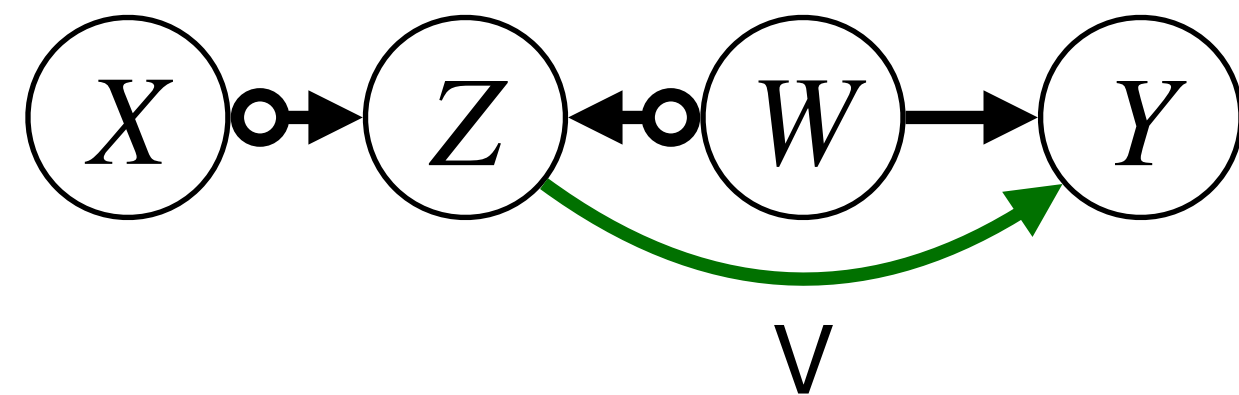
FCI's R4:



$$X \perp\!\!\!\perp Y \mid Z, W$$

$$W \in \text{Sepset}(X, Y)$$

PAG: Representation of the Markov Equivalence Class

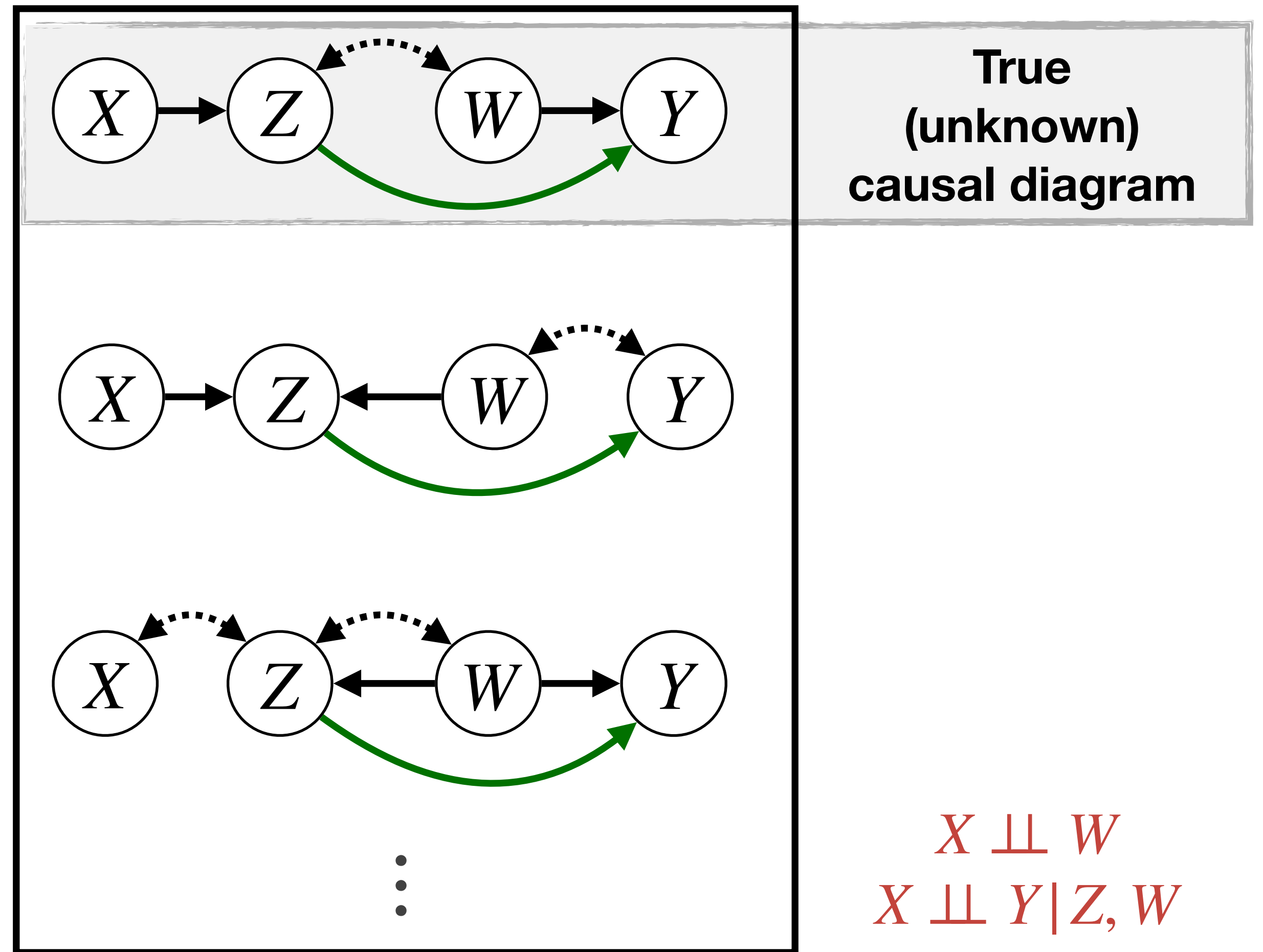


Partial Ancestral Graph
(PAG)

Z is not an ancestor of X or W .

Z and W are ancestors of Y .

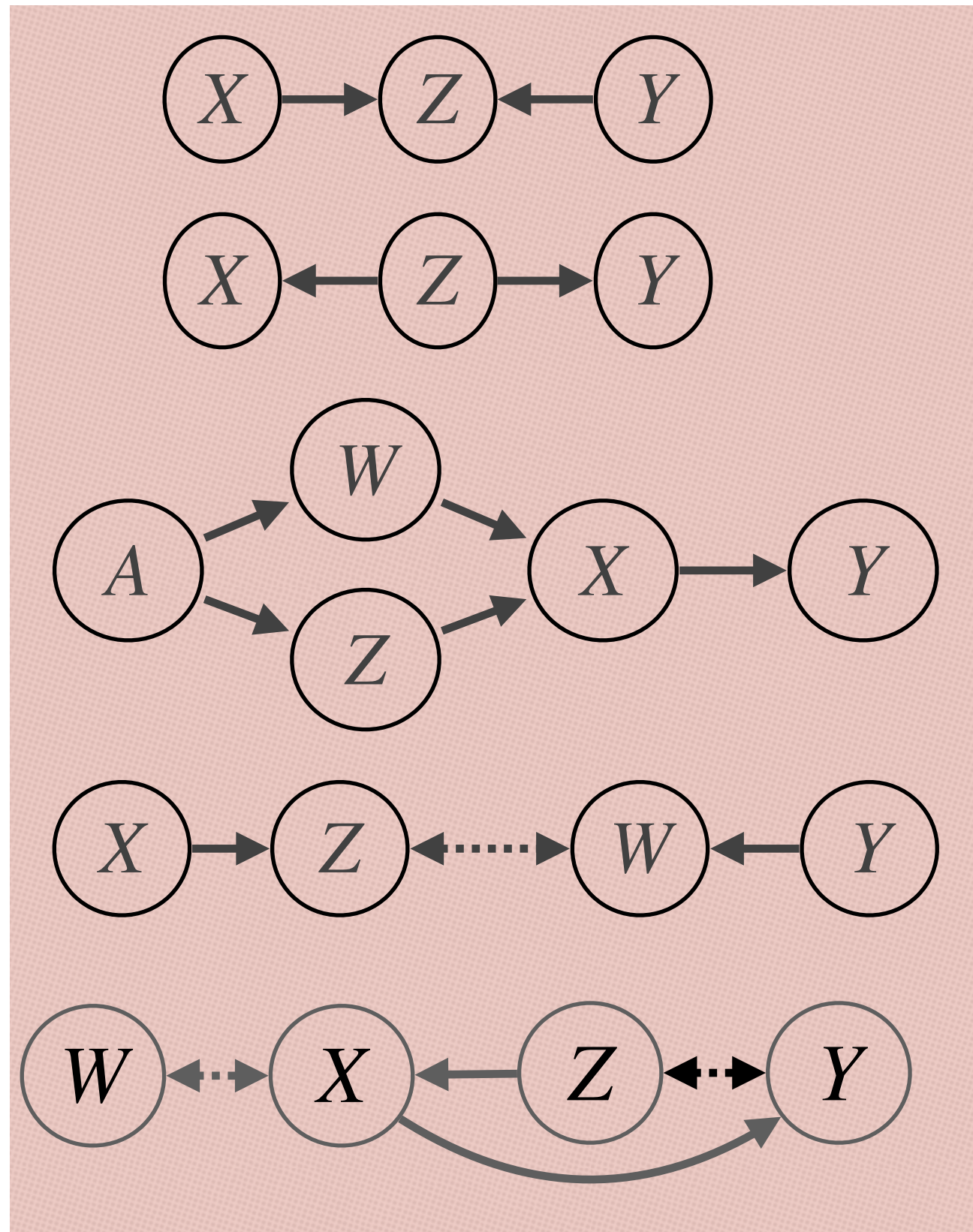
Z is not confounded with Y .



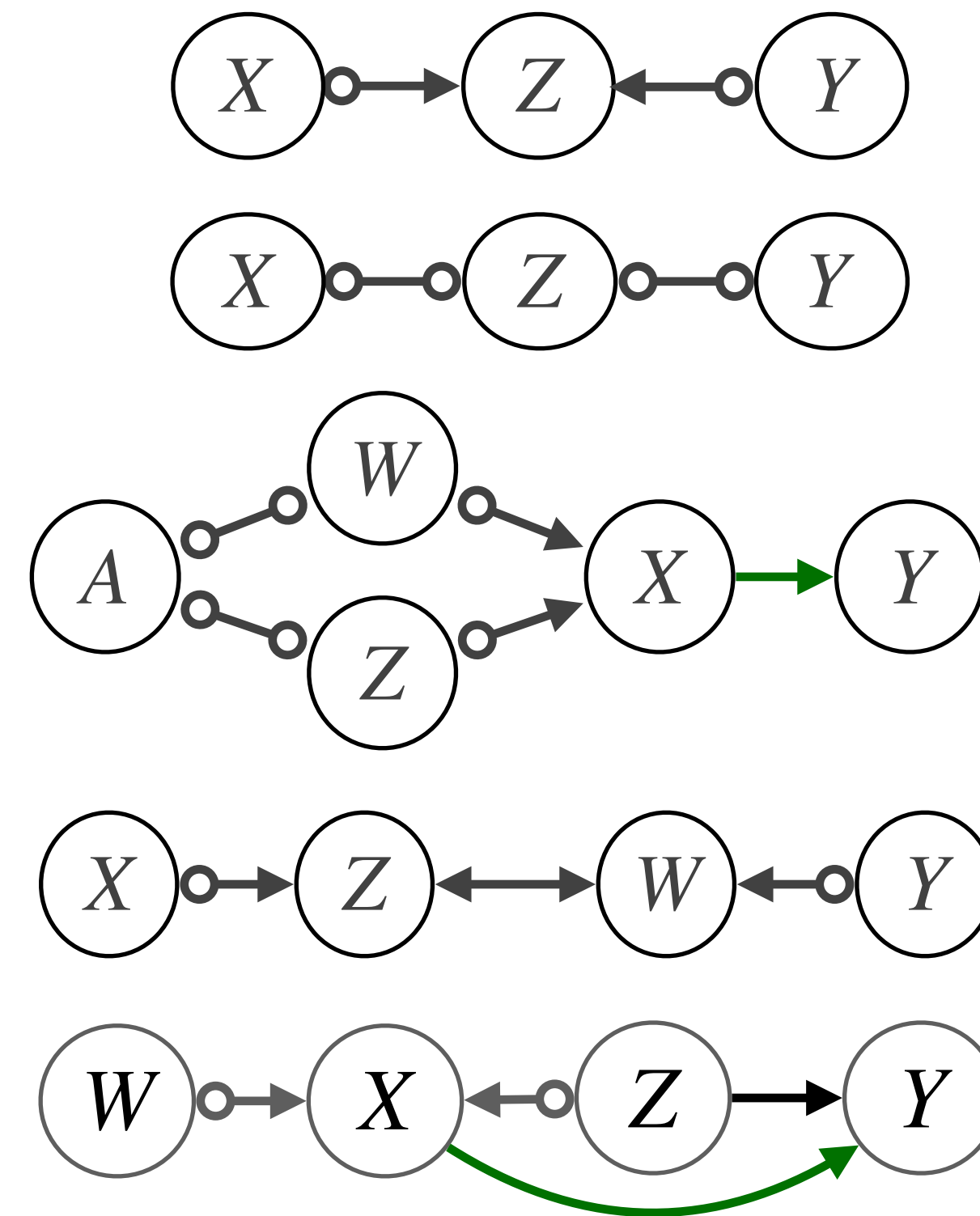
$X \perp\!\!\!\perp W$
 $X \perp\!\!\!\perp Y | Z, W$

Fast Causal Inference (FCI) Algorithm

Underlying Causal Diagram



Partial Ancestral Graph



Available Implementations of the FCI

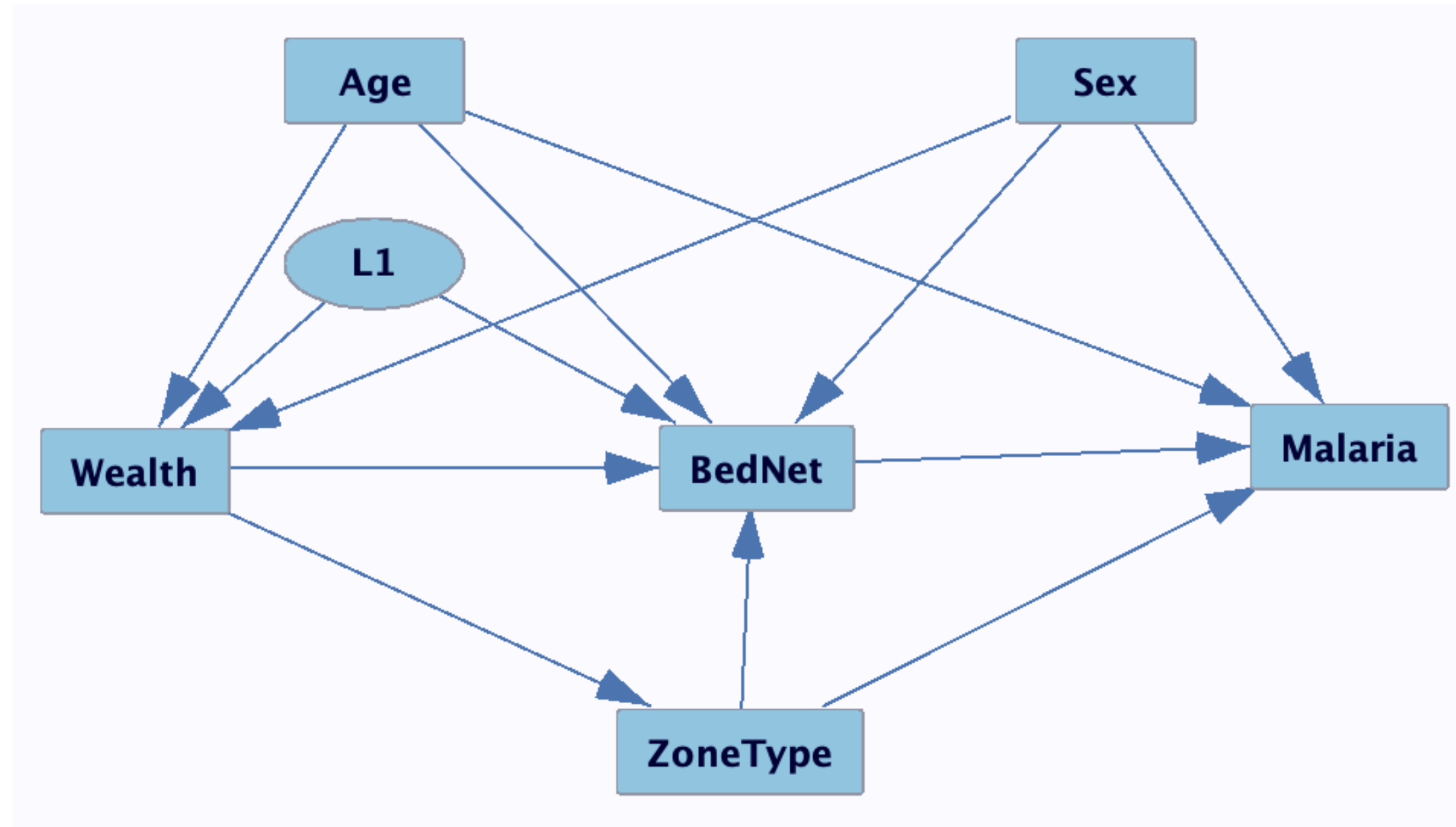
R Packages:

- pcalg R package:
 - <https://cran.r-project.org/web/packages/pcalg/>
 - <https://github.com/cran/pcalg/>
- RPy-Tetrad (Wrapper in R): <https://github.com/cmu-phil/py-tetrad/tree/main/pytetrad/R>

Python Packages:

- Do-discover in PyWhy: <https://github.com/py-why/dodiscover>
- Causal-Learn: <https://causal-learn.readthedocs.io/en/latest/index.html>
- Py-Tetrad (Wrapper in Python): <https://github.com/bd2kccd/py-causal>

Going Back to Our Simple Example...



True Causal Diagram
(unknown)

In this simplified hypothetical scenario, the observed variables are of **heterogenous** types:

Age — continuous

Sex — binary (female vs. male)

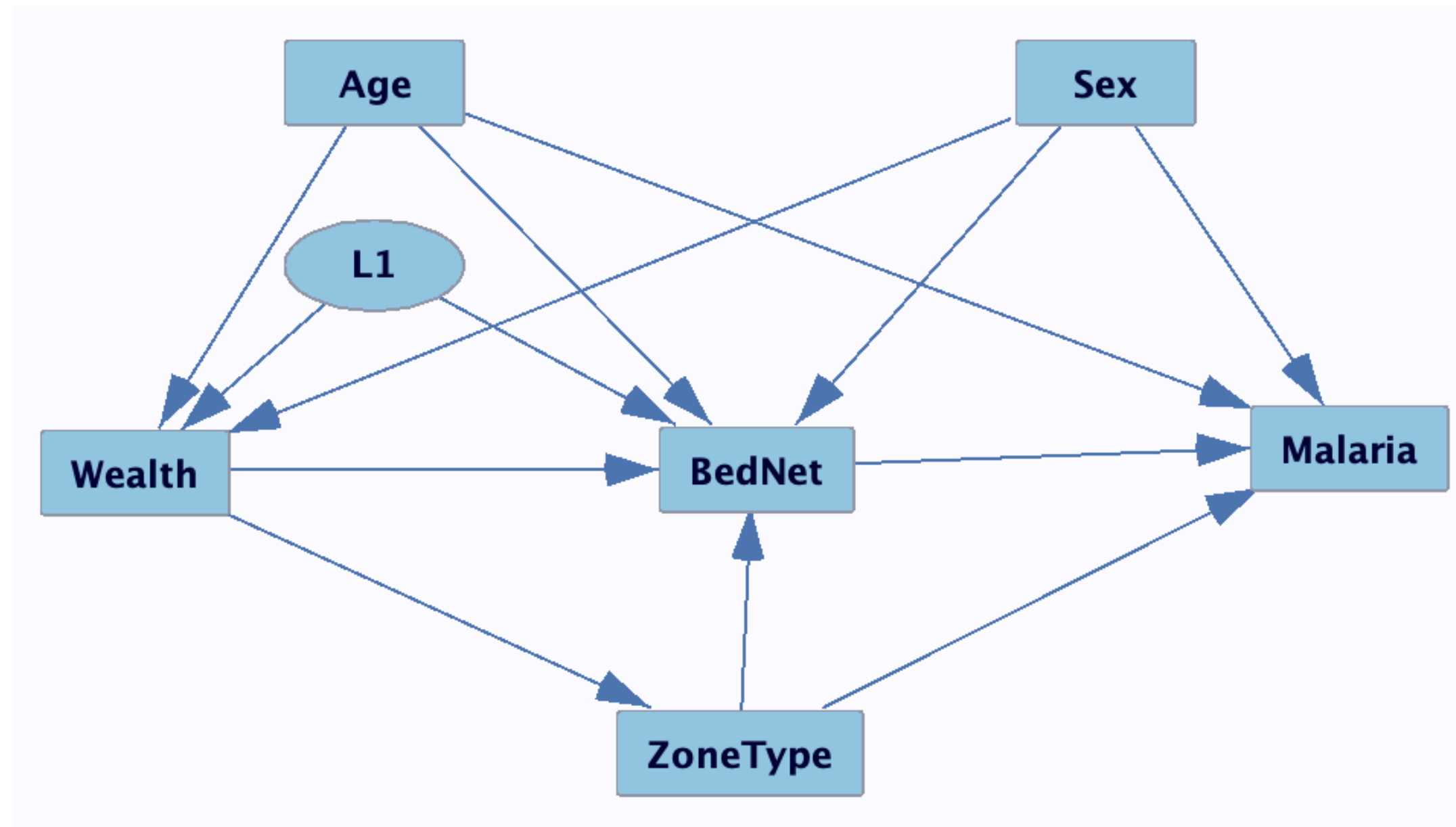
Wealth — continuous

Zone type — categorical (center vs. periphery)

Usual bed net use — binary (yes vs. no)

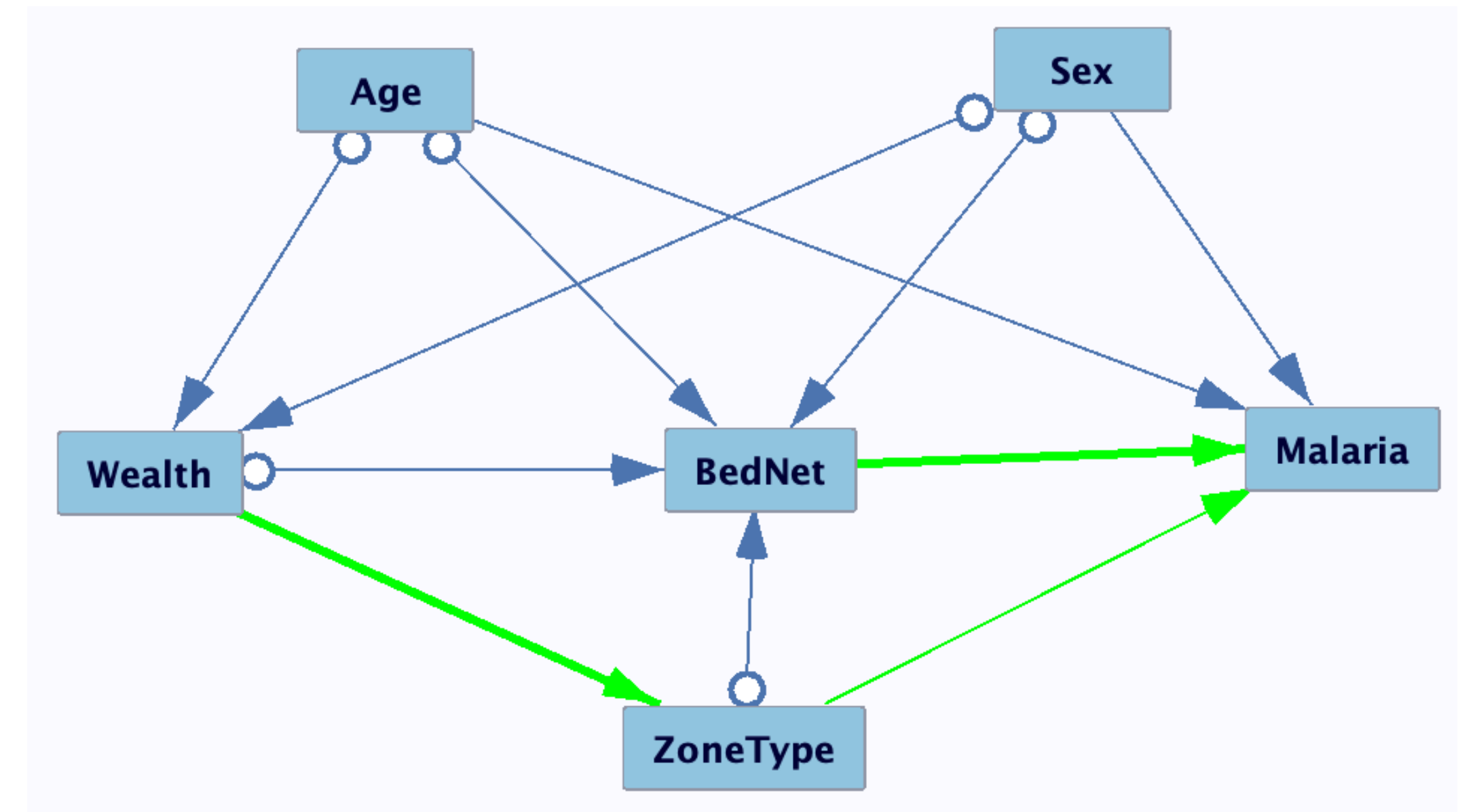
Number of malaria infections — count variable

Going Back to Our Simple Example...



True Causal Diagram
(unknown)

FCI
➔



Partial Ancestral Graph
(learned from data)

Accurate PAG inference with FCI requires conditional independence tests that are statistically appropriate and **correctly specified** for the **variable types** and underlying **structural, distributional and functional** properties.

Conditional Independence Tests

With **independent observations**:

Linear models with **Gaussian errors**: partial correlation test

- **Fisher, R.A.** (1921). *On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample*.
- R package: <https://cran.r-project.org/web/packages/pcalg/>

Linear models with **heterogeneous data types**- likelihood ratio tests based on GLM

- **Tsagris, M., Borboudakis, G., Lagani, V. et al.** (2018) Constraint-based causal discovery with mixed data. *Int J Data Sci Anal* **6**, 19–30. ([Link](#))
- R package: <https://cran.r-project.org/web/packages/MXM/>

Linear / non-linear models with **heterogeneous data types** — kernel-based non-parametric test

- **Zhang, K., Peters, J., Janzing, D., & Schölkopf, B.** (2012). *Kernel-based conditional independence test and application in causal discovery*. In: Uncertainty in artificial intelligence. AUAI Press; 2011. p.804–13
- R package: <https://cran.r-project.org/web/packages/CondIndTests>

With **dependent observations**:

Linear models with **Gaussian errors** and **family data** :

- **Ribeiro A.H., Soler J.M.P.** (2020). *Learning Genetic and environmental graphical models from family data*, Statistics in Medicine.
- R package: <https://github.com/adele/FamilyBasedPGMs>

Fisher's Partial Correlation Test

Linear models with **Gaussian errors**: partial correlation test

- **Fisher, R.A.** (1921). *On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample*.
- R package: <https://cran.r-project.org/web/packages/pcalg/>

Testing $X \perp\!\!\!\perp Y | S$:

For multivariate Gaussian variables,
the **partial correlation test** is equivalent to a
conditional independence test.

1) Obtain the **partial correlation coefficient** via univariate **linear regression**:

Give i.i.d. data $D = \{x_i, y_i, s_{ki} : k = 1, \dots, |S| \}$

$$x_i = \alpha^X + \sum_{k=1}^{|S|} \beta_k^X s_{ki} + \varepsilon_i^X, \text{ where } \varepsilon_i^X \sim \mathcal{N}(0, \sigma_{X|S}^2)$$

$$y_i = \alpha^Y + \sum_{k=1}^{|S|} \beta_k^Y s_{ki} + \varepsilon_i^Y, \text{ where } \varepsilon_i^Y \sim \mathcal{N}(0, \sigma_{Y|S}^2)$$

Then, $\rho_{XY|S} = \text{corr}(\boldsymbol{\varepsilon}^X, \boldsymbol{\varepsilon}^Y)$

The partial correlation coefficient
is the Pearson correlation
between the residuals!

Fisher's Partial Correlation Test

2) Test the **significance** of the partial correlation coefficient:

Hypothesis test: $H_0 : \rho_{XY|S} = 0$ vs $H_1 : \rho_{XY|S} \neq 0$

Fisher's z-transformed partial correlation coefficient:

$$z(\rho_{XY|S}) = \frac{1}{2} \ln \left(\frac{1 + \rho_{XY|S}}{1 - \rho_{XY|S}} \right).$$

Under H_0 , the test statistic $T = \sqrt{N - |\mathbf{S}| - 3} |z(\rho_{XY|S})|$ asymptotically follows a standard Gaussian distribution (zero mean and unit variance).

We reject H_0 if p-value $< \alpha$, where p-value for the two-tailed test is given by

$$P(T > |t| | H_0) = 2 \times (1 - cdf(|t|)),$$

t is the observed (estimated) value of T , and

cdf is the cumulative distribution function of T under H_0 .

How can we accommodate heterogeneous data types?

Linear models with **heterogeneous data types**- likelihood ratio tests based on **GLM**

- Tsagris, M., Borboudakis, G., Lagani, V. *et al.* (2018) Constraint-based causal discovery with mixed data. *Int J Data Sci Anal* **6**, 19–30. ([Link](#))
- R package: <https://cran.r-project.org/web/packages/MXM/>

Likelihood Ratio Test (LRT) for testing $X \perp\!\!\!\perp Y | S$:

1) First, we test $H_{0_1} : P(Y | X, S) = P(Y | S)$

1.a) Fit the null and alternative models for Y using **generalized linear models (GLMs)**:

Give i.i.d. data $D = \{x_i, y_i, s_{ki} : k = 1, \dots, |S| \}$

$$\text{Null model } \mathcal{M}_0: \quad g(E[Y_i | S_i]) = \eta_i^{(0)} = \alpha + \sum_{k=1}^{|S|} \beta_{S_k} s_{ki}$$

$$\text{Alternative model } \mathcal{M}_1: \quad g(E[Y_i | X_i, S_i]) = \eta_i^{(1)} = \alpha + \beta_X X_i + \sum_{k=1}^{|S|} \beta_{S_k} s_{ki}$$

where $g(\cdot)$ is the link function appropriated to the data type of Y (e.g., identity for Gaussian, logit for binary, log for counts).

How can we accommodate heterogeneous data types?

1.b) Compare the models using a likelihood-ratio test (LRT):

Let:

- $\ell_0 = \mathbf{log} \mathcal{L}(\theta_{\mathcal{M}_0}) = \sum_{i=1}^n \log f(y_i | S_i^\top \beta_S)$, be the maximized log-likelihood under \mathcal{M}_0
- $\ell_1 = \mathbf{log} \mathcal{L}(\theta_{\mathcal{M}_1}) = \sum_{i=1}^n \log f(y_i | S_i^\top \beta_S + X_i^\top \beta_X)$ be the maximized log-likelihood under \mathcal{M}_1

Then the LRT statistic $\Lambda = -2 (\ell_0 - \ell_1)$ measures how much better the alternative model \mathcal{M}_1 fits the data compared to the null model \mathcal{M}_0 .

Under $H_{0_1} : \beta_x = 0$ and regularity conditions:

$$\Lambda \xrightarrow{d} \chi_k^2, \quad \text{with } k = \dim(\beta_X).$$

This enables computation of p-value p_1 corresponding to testing $H_{0_1} : P(Y | X, \mathbf{Z}) = P(Y | \mathbf{Z})$

How can we accommodate heterogeneous data types?

2) Repeat steps 1.a) and 1.b), switching the roles of Y and X , to test $H_{0_2} : P(X | Y, \mathbf{S}) = P(X | \mathbf{S})$

This enables computation of p-value p_2 .

3) Tests are not necessarily symmetric, so a final p-value can be computed as following:

$p = \min(p_1, p_2)$ Favors dependencies

$p = \max(p_1, p_2)$ Favors independencies

$p = \min\{2 \times \min(p_1, p_2), \max(p_1, p_2)\}$ Heuristic shown to have a good control of FP and FN

Priority: Continuous > Nominal > Ordinal

Given a significance level α (e.g., 0.05),

$p > \alpha$ suggests that X does not add
explanatory power for Y beyond \mathbf{S} .

$$\implies X \perp\!\!\!\perp Y | \mathbf{S}$$

LR-Based CI Test for Heterogenous Data Types

Implementations in R:

MXM R Package — functions `ci.mm2` and `ci.fast2`

- Handles **binary, nominal** and **ordinal** variables

Lagani V, Athineou G, Farcomeni A, Tsagris M, Tsamardinos I (2017). “Feature Selection with the R Package MXM: Discovering Statistically Equivalent Feature Subsets.” *Journal of Statistical Software*, 80(7). doi:10.18637/jss.v080.i07

CRAN: Check latest archived version: <https://cran.r-project.org/src/contrib/Archive/MXM/>

GitHub: <https://github.com/mensxmachina/MXM-R-Package>

FCI.Utills R Package — function `mixedCITest`

- Handles **binary, nominal, ordinal, proportions, and (zero inflated) count** variables

GitHub: <https://github.com/adele/FCI.Utills>

Equivalence of Residual, Likelihood, and Coefficient-Based Tests

Under **joint Gaussianity**, partial correlation, regression-based tests (e.g., t-tests), and likelihood ratio tests all provide **equivalent** ways to test for conditional independence:

Zero partial correlation between X and Y given \mathbf{S}

(zero correlation between the residuals of X and Y after regressing out \mathbf{S})



No improvement in model fit when adding X to a regression of Y on \mathbf{S}

(as tested by an LRT or F-test)



No statistical significance when testing whether $\beta_X = 0$ (e.g., via a t-test)

in the linear model $Y = \beta_X X + \mathbf{S}^T \beta_{\mathbf{S}} + \varepsilon$

Conditional Independence Tests in Finite Samples

Even when conditional independence tests are **correctly specified**, errors may arise due to:

- Limited sample size
- Weak effect sizes
- Multicollinearity or measurement noise

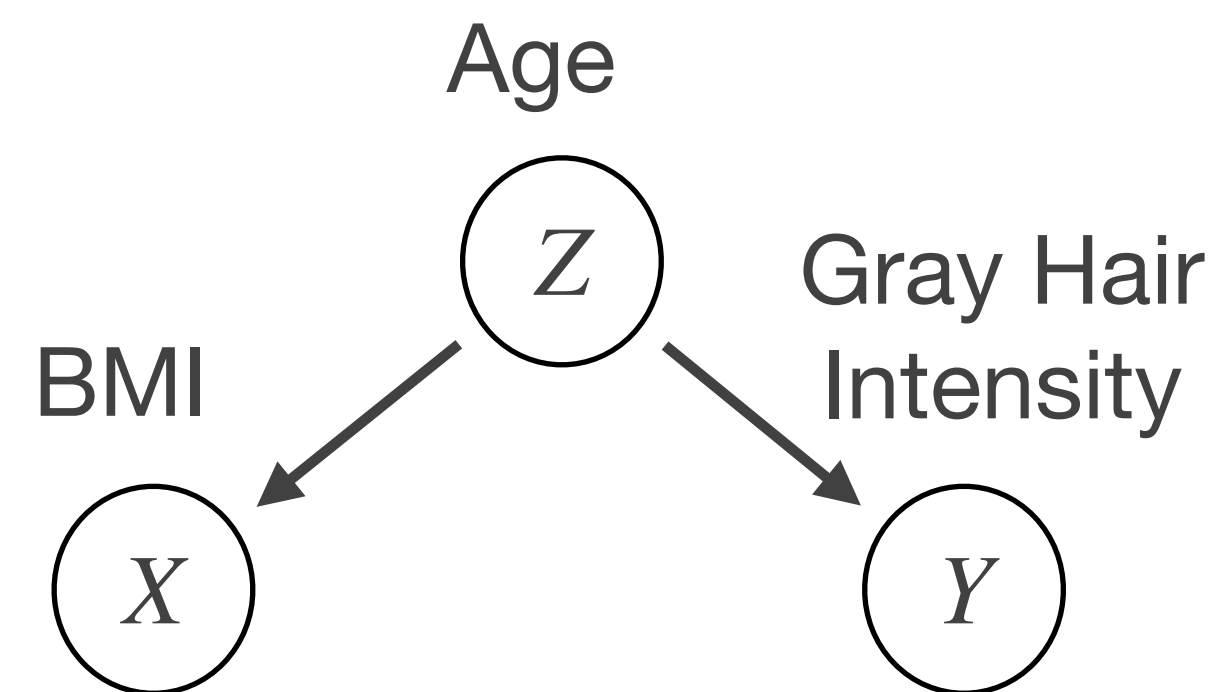
Consequences include:

- Failure to reject H_0 — **false independencies**
- Incorrect rejection of H_0 - **false dependencies**

Example - Fork

Fork

Z as a common cause



~~$X \perp\!\!\!\perp Y$~~

$X \perp\!\!\!\perp Y | Z$

```
n <- 1000
age <- rnorm(n, 30, 5)
bmi <- 15 + 0.5*age + rnorm(n, 0, 2)
gray_hair <- 0.02*age + rnorm(n, 0, 0.01)
```

```
library("psych")
xyS <- data.frame(age, bmi, gray_hair)
cort <- cor.test(xyS)
```

"Pearson correlation matrix via psych:"

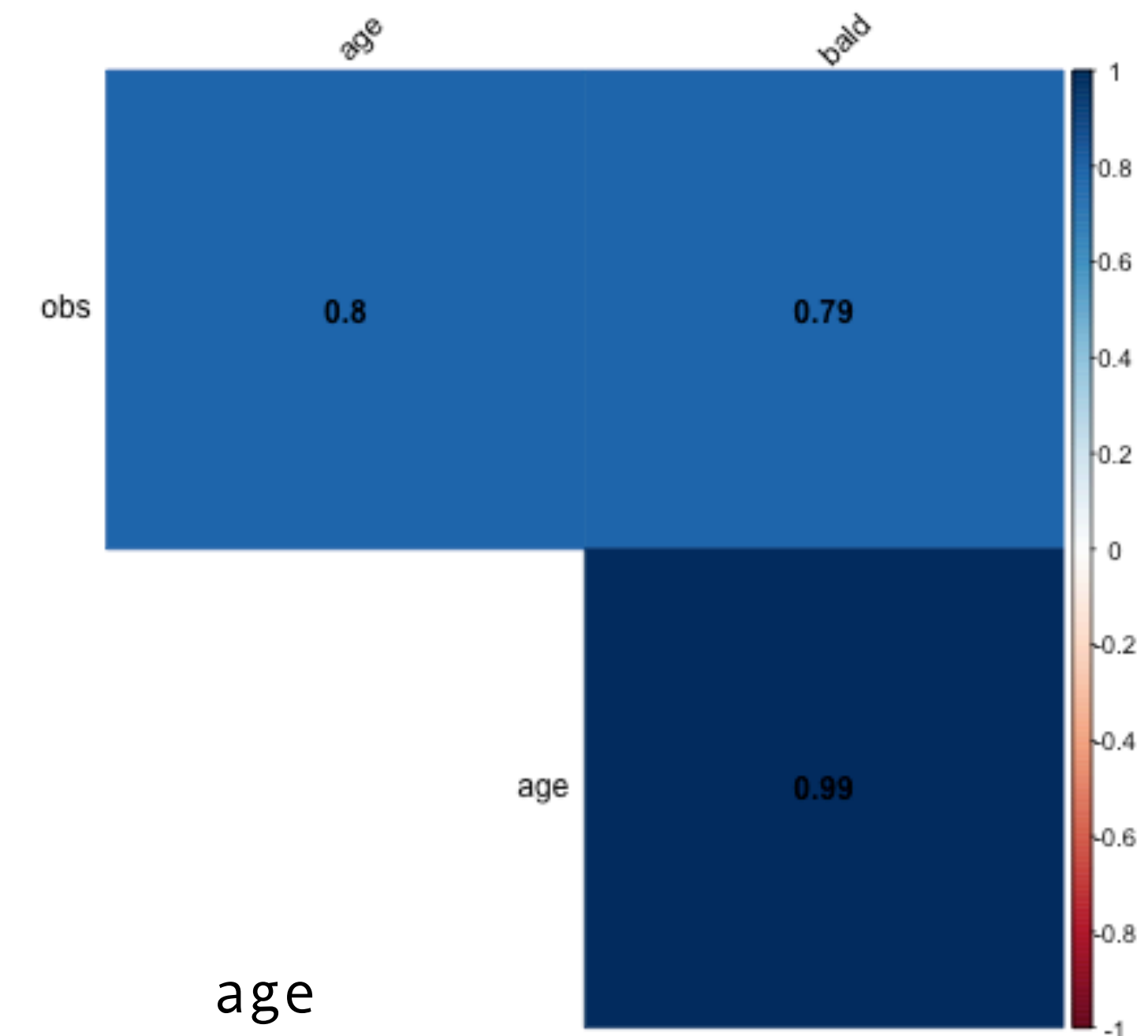
```
> cort$r
```

	Bmi	gray_hair	age
obs	1.0000000	0.7928973	0.7991331
gray_hair	0.7928973	1.0000000	0.9949050
age	0.7991331	0.9949050	1.0000000

"p-values:"

```
> cort$p
```

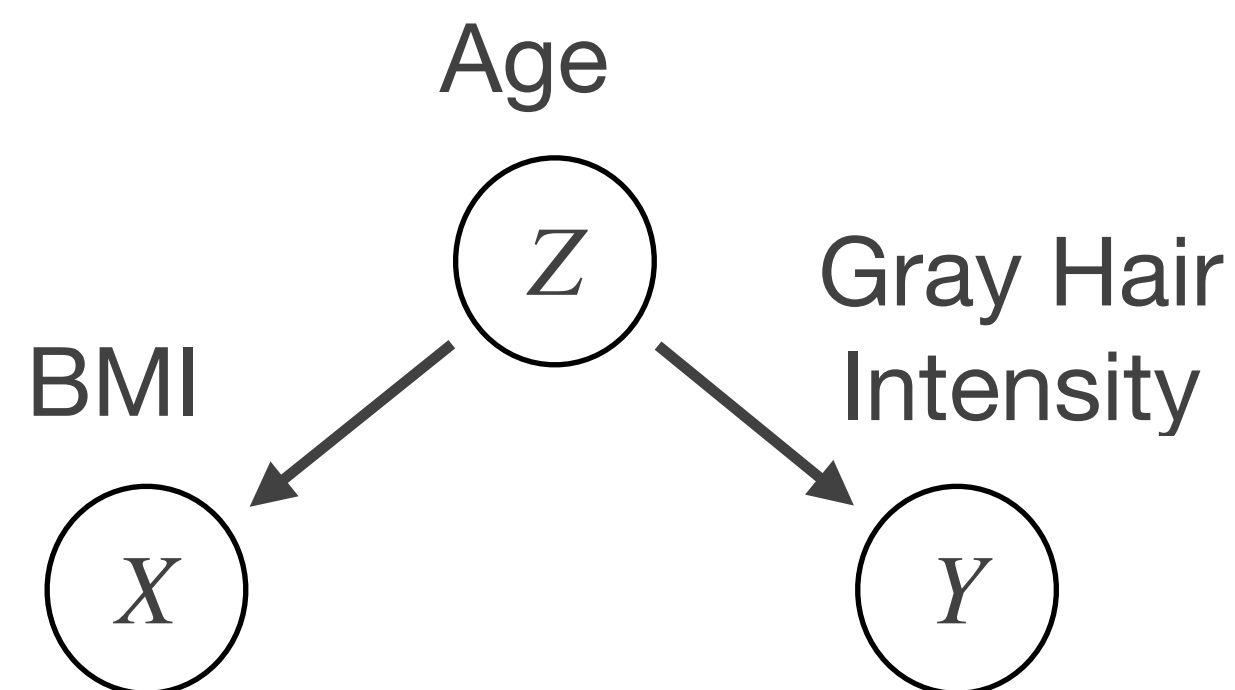
	Bmi	gray_hair	age
bmi	0.0000000e+00	6.363175e-217	1.691675e-222
gray_hair	6.363175e-217	0.0000000e+00	0.0000000e+00
age	8.458374e-223	0.0000000e+00	0.0000000e+00



Example - Fork

Fork

Z as a common cause



~~$X \perp\!\!\!\perp Y$~~

$X \perp\!\!\!\perp Y | Z$

```
n <- 1000
age <- rnorm(n, 30, 5)
Bmi <- 15 + 0.5*age + rnorm(n, 0, 2)
gray_hair <- 0.02*age + rnorm(n, 0, 0.01)
```

```
# Estimation via linear regression
```

```
fit_X.S <- lm(bmi ~ age)
fit_Y.S <- lm(gray_hair ~ age)
cor(residuals(fit_X.S), residuals(fit_Y.S))
```

```
> -0.03570906
```

```
# Alternative with a test
```

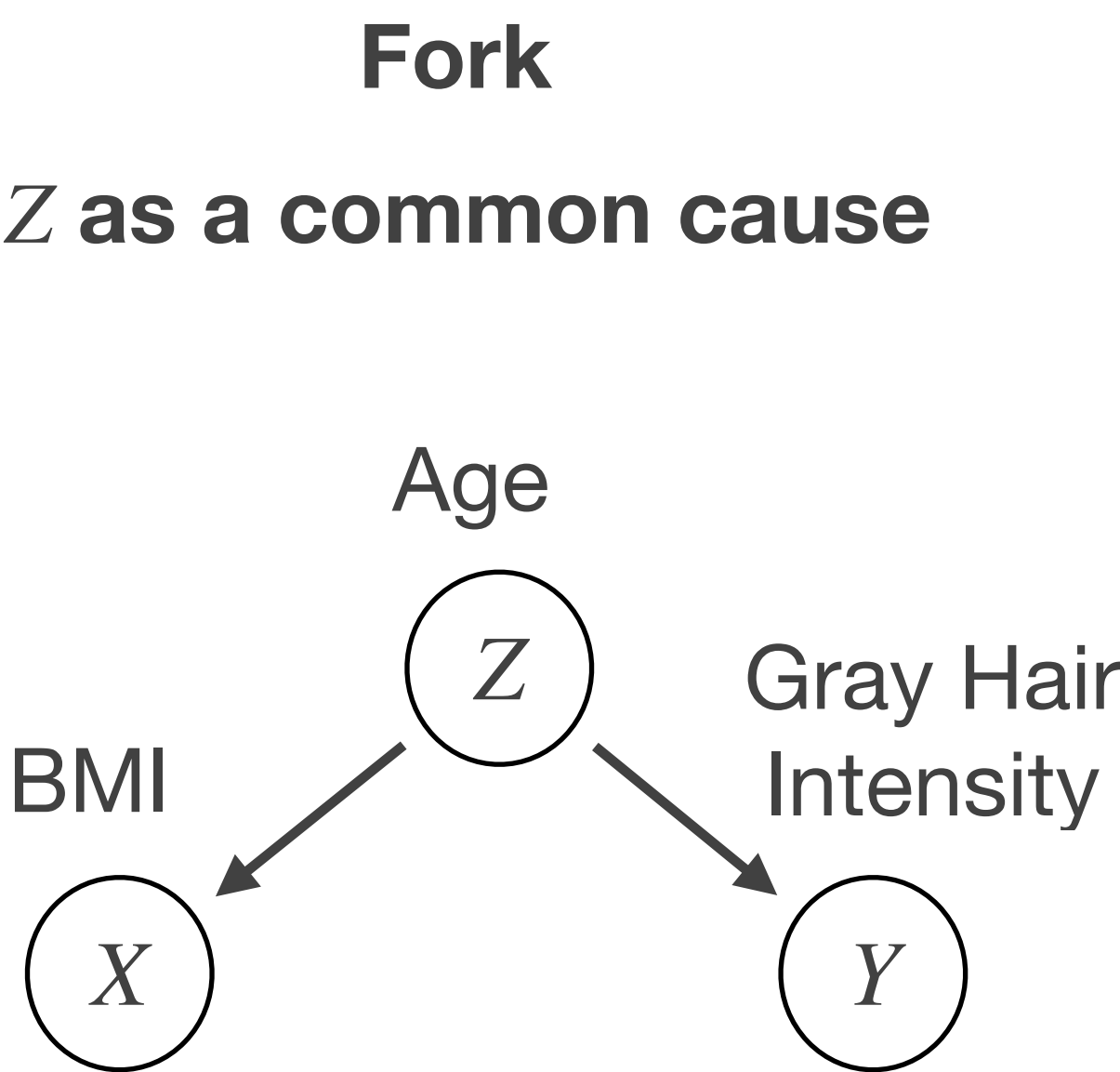
```
cor.test(residuals(fit_X.S), residuals(fit_Y.S))
```

Pearson's product-moment correlation

```
data: residuals(fit_X.S) and residuals(fit_Y.S)
t = -1.1288, df = 998, p-value = 0.2592
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.09748633  0.02634234
sample estimates:
cor
-0.03570906
```

Empirical Unfaithfulness with Limited Data

N=500. Results from 1000 runs:



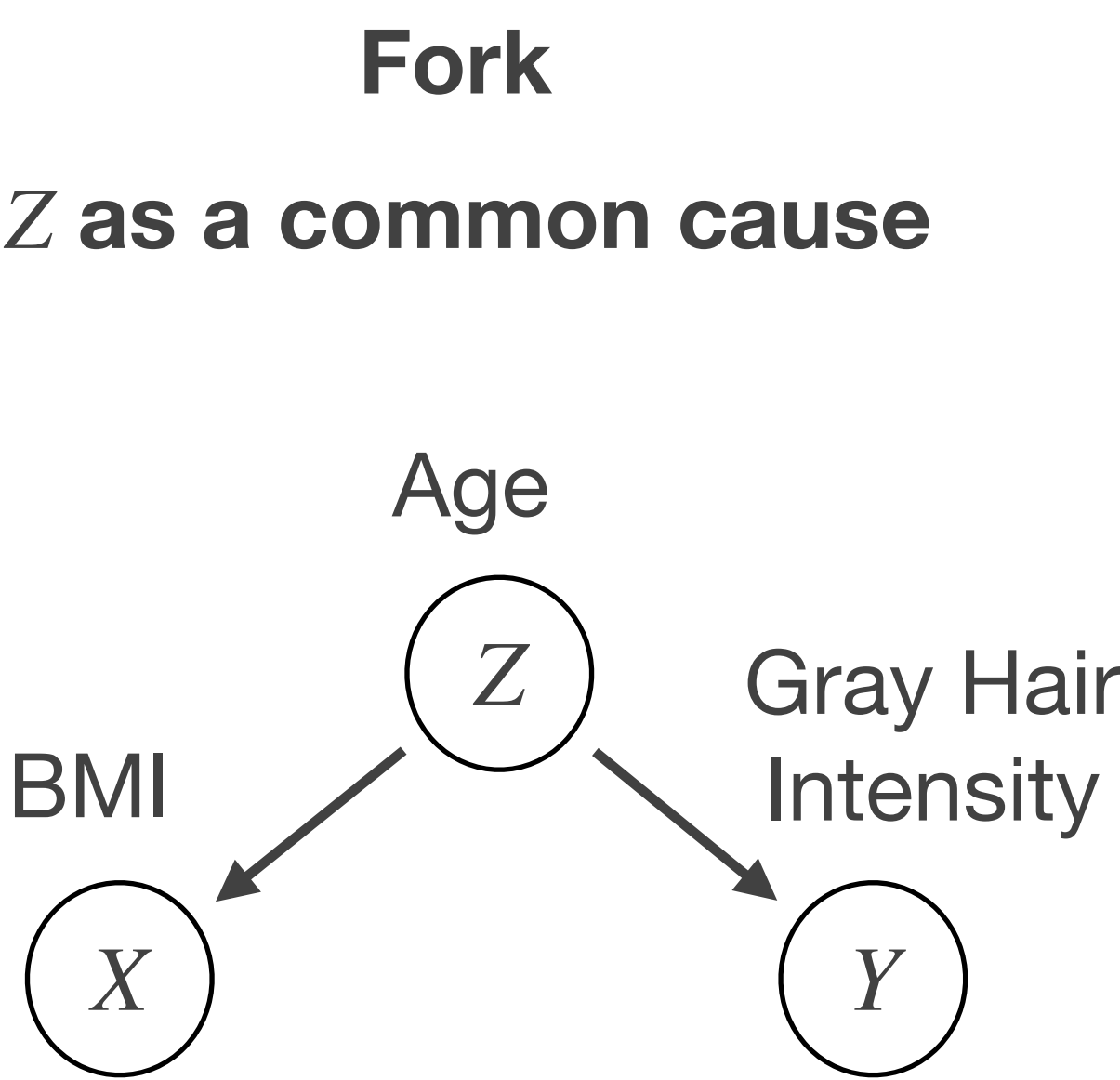
~~$X \perp\!\!\!\perp Y$~~
 $X \perp\!\!\!\perp Y | Z$

Partial Correlation Coefficients Mean +- SD			P-Values Mean +- SD		
	Gray Hair	Age		Gray Hair	Age
BMI	0 +- 0.04 min: -0.12 max: 0.14	0.13 +- 0.04 min: -0.01 max: 0.24	BMI	0.51 +- 0.29 min: 0 max: 1	0.05 +- 0.12 min: 0 max: 0.99
Gray Hair		0.99 +- 0 min: 0.98 max: 0.99	Gray Hair		0 +- 0 min: 0 max: 0

False dependencies :
 $\rho_{GrayHair,BMI|Age}$ can be incorrectly estimated as non-zero!

Empirical Unfaithfulness with Limited Data

N=500. Results from 1000 runs



~~$X \perp\!\!\!\perp Y$~~
 $X \perp\!\!\!\perp Y | Z$

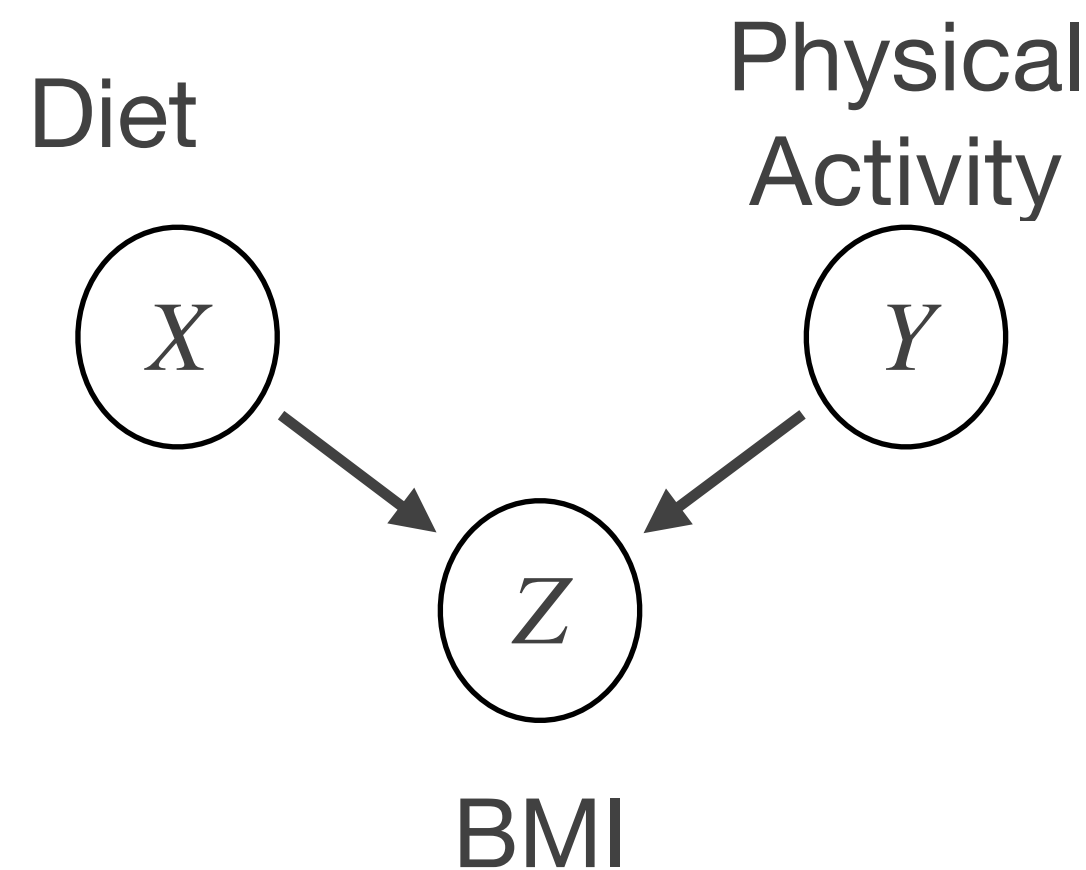
Partial Correlation Coefficients Mean +- SD			P-Values Mean +- SD		
	Gray Hair	Age		Gray Hair	Age
BMI	0 +- 0.04 min: -0.12 max: 0.14	0.13 +- 0.04 min: -0.01 max: 0.24	BMI	0.51 +- 0.29 min: 0 max: 1	0.05 +- 0.12 min: 0 max: 0.99
Gray Hair		0.99 +- 0 min: 0.98 max: 0.99	Gray Hair		0 +- 0 min: 0 max: 0

False independences:
 $\rho_{Age,BMI|GrayHair}$ can be incorrectly
estimated as zero!

Example - Collider

V-Structure

Z as a collider or common effect



$X \perp\!\!\!\perp Y$

$X \not\perp\!\!\!\perp Y | Z$

```
n <- 1000
diet <- rnorm(n, 3, 1)
physAct <- rnorm(n, 1, 0.3)
bmi <- 25 + 1.8*diet - 2.5*physAct + rnorm(n, 0, 2)
```

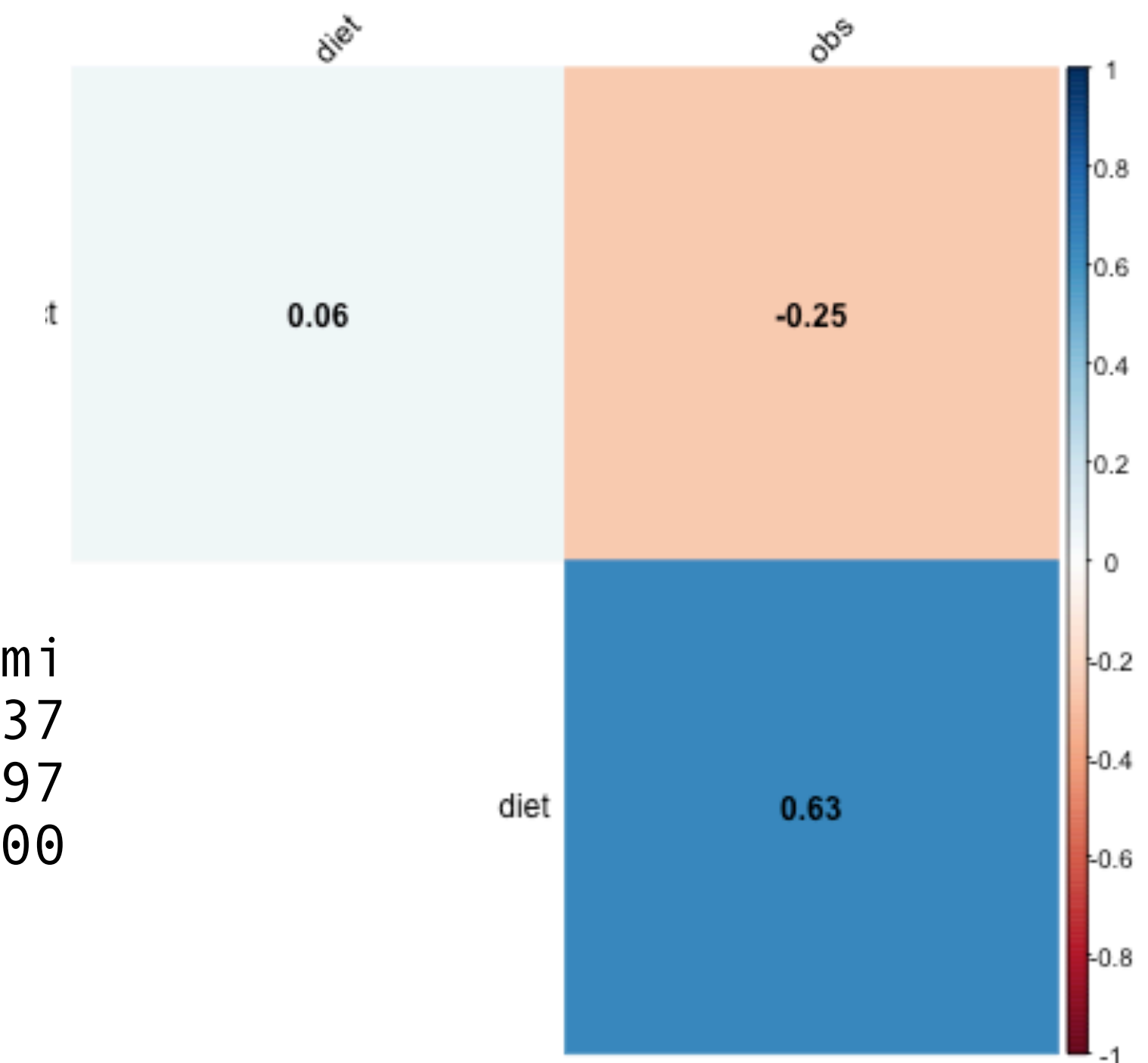
```
library("psych")
xyS <- data.frame(diet, physAct, bmi)
cort <- cor.test(xyS)
```

"Pearson correlation matrix via psych:"
> cort\$r

	diet	physAct	bmi
diet	1.00000000	0.05663004	0.6337937
physAct	0.05663004	1.00000000	-0.2502297
Bmi	0.63379369	-0.25022971	1.00000000

"p-values:"
> cort\$p

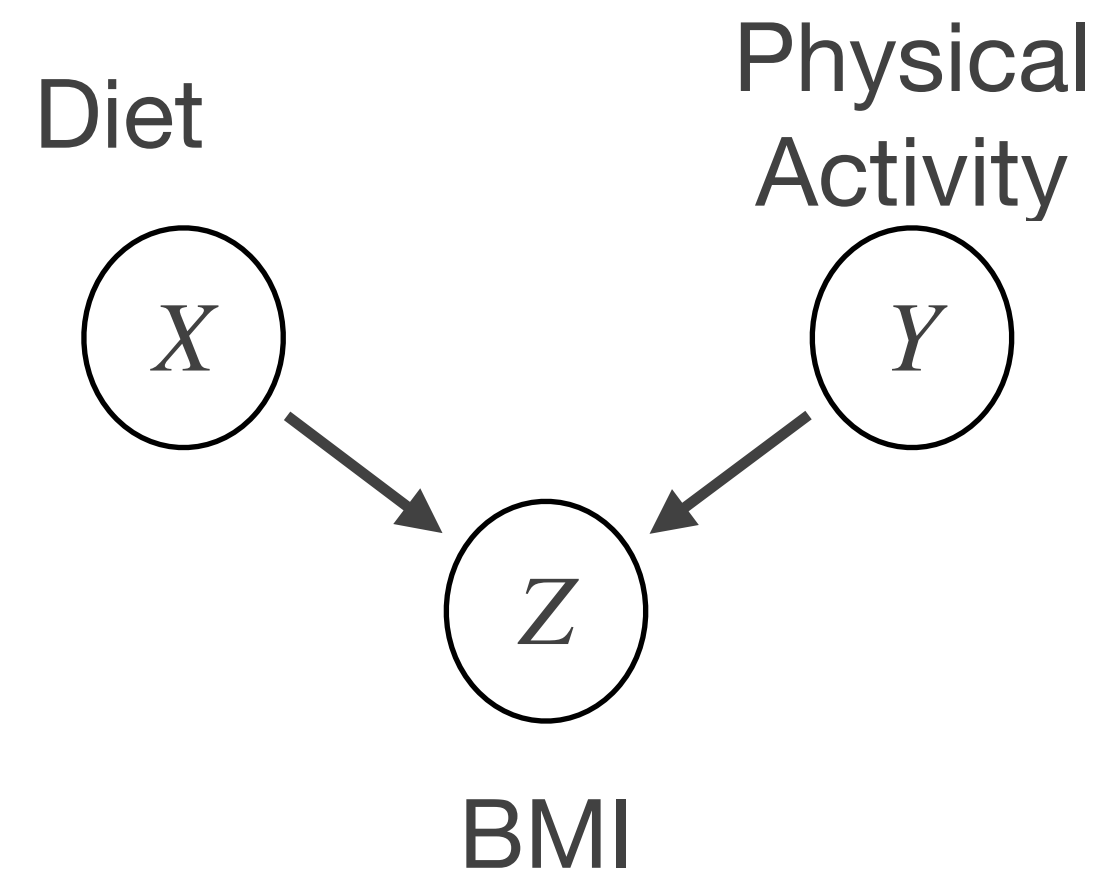
	diet	physAct	bmi
diet	0.0000000e+00	7.345456e-02	5.772847e-113
physAct	7.345456e-02	0.0000000e+00	1.931014e-15
Bmi	1.924282e-113	9.655072e-16	0.0000000e+00



Example - Collider

V-Structure

Z as a collider or common effect



$$X \perp\!\!\!\perp Y$$

$$\cancel{X \perp\!\!\!\perp Y | Z}$$

```
n <- 1000
diet <- rnorm(n, 3, 1)
physAct <- rnorm(n, 1, 0.3)
Bmi <- 25 + 1.8*diet -2.5*physAct + rnorm(n, 0, 2)
```

```
# Estimation via linear regression
```

```
fit_X.S <- lm(diet ~ bmi)
fit_Y.S <- lm(physAct ~ bmi)
cor(residuals(fit_X.S), residuals(fit_Y.S))
```

```
> 0.2873891
```

```
# Alternative with a test
```

```
cor.test(residuals(fit_X.S), residuals(fit_Y.S))
```

Pearson's product-moment correlation

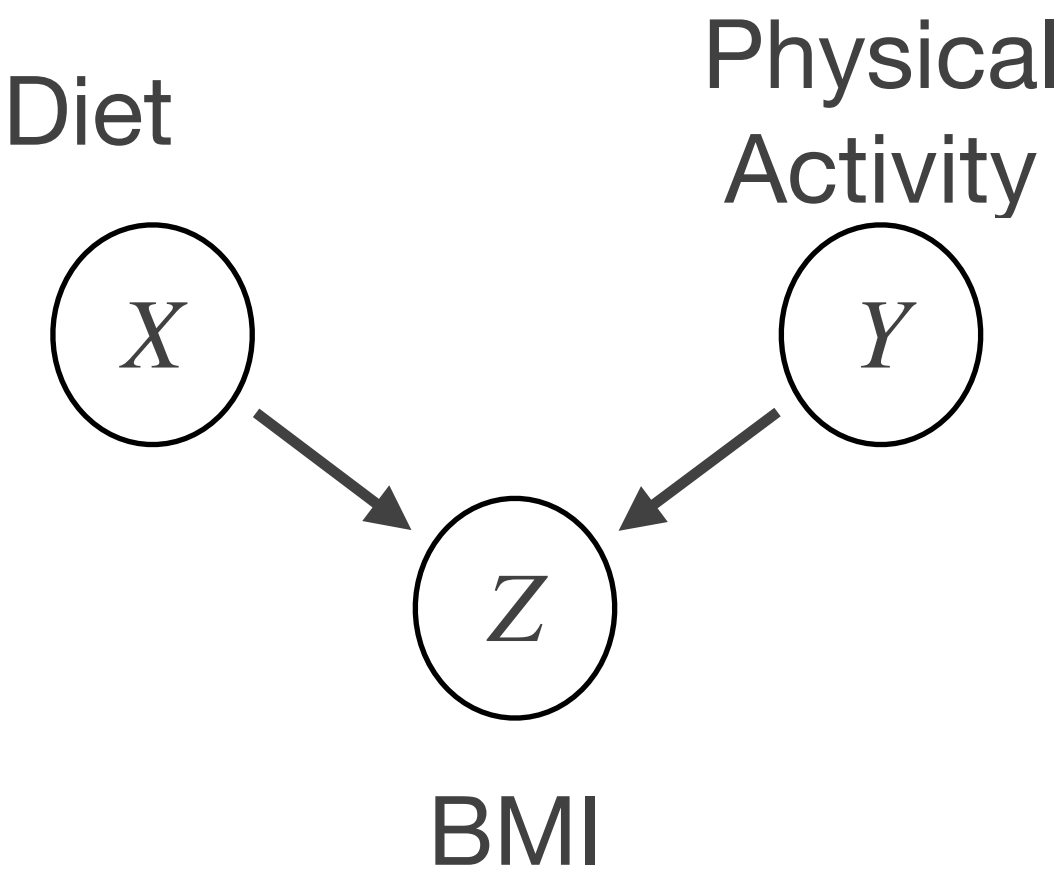
```
data: residuals(fit_X.S) and residuals(fit_Y.S)
t = 9.4788, df = 998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2294845 0.3432665
sample estimates:
      cor
0.2873891
```


Empirical Unfaithfulness with Limited Data

N=500. Results from 1000 runs

V-Structure

Z as a collider or common effect



$X \perp\!\!\!\perp Y$
 ~~$X \perp\!\!\!\perp Y | Z$~~

Partial Correlation Coefficients Mean +- SD			P-Values Mean +- SD		
	Diet	BMI		Diet	BMI
Physical Activity	0.23 +- 0.05 min: 0.03 max: 0.36	0 +- 0 min: 0 max: 0	Physical Activity	0 +- 0.02 min: 0 max: 0.55	0 +- 0 min: 0 max: 0
BMI		0 +- 0 min: 0 max: 0	BMI		0 +- 0 min: 0 max: 0

False independencies:

$\rho_{Diet, PhysAct | Obesity}$ can be incorrectly estimated as zero!

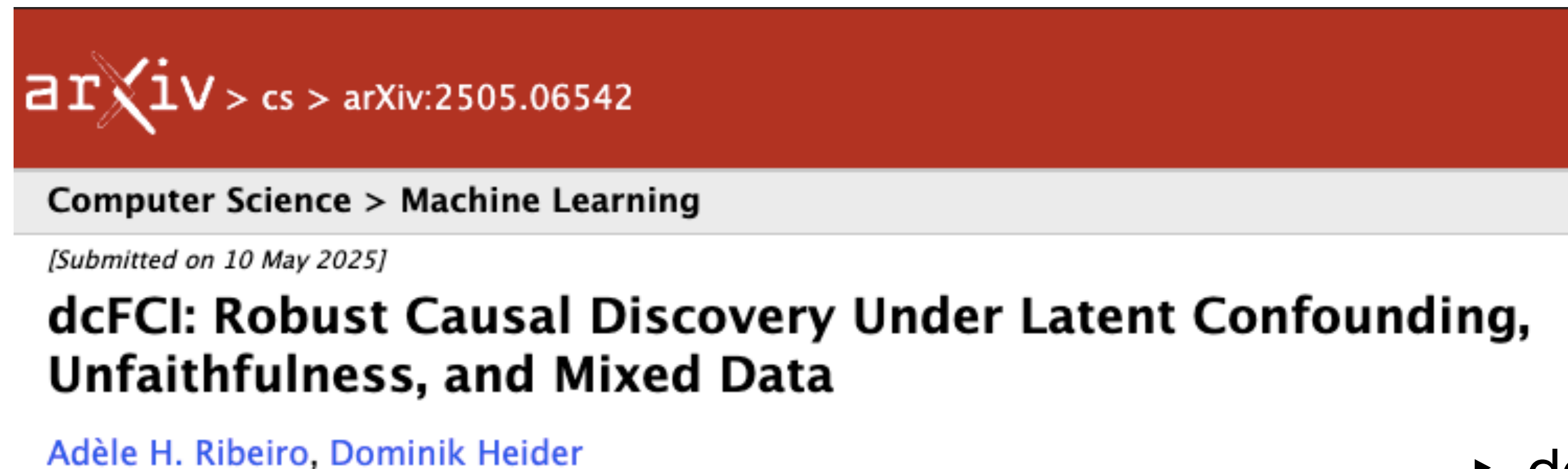
~~$Y \perp\!\!\!\perp Z$~~ $X \perp\!\!\!\perp Y$
 ~~$X \perp\!\!\!\perp Z$~~ ~~$X \perp\!\!\!\perp Y | Z$~~

Inconsistencies \Rightarrow Ambiguous Triplet

Robust Causal Discovery under Empirical Unfaithfulness

These errors can ultimately lead to misleading causal or structural inferences.

We need **uncertainty-aware** approaches to ensure reliable causal inference!



- dcFCI R package:
GitHub repository: [@adele/dcFCI](https://github.com/adele/dcFCI)

Ribeiro, A. H., & Heider, D. (2025). dcFCI: Robust Causal Discovery Under Latent Confounding, Unfaithfulness, and Mixed Data. *arXiv preprint arXiv:2505.06542 (under review)*.

Handling Missing Data in Conditional Independence Tests

Test-Wise Deletion

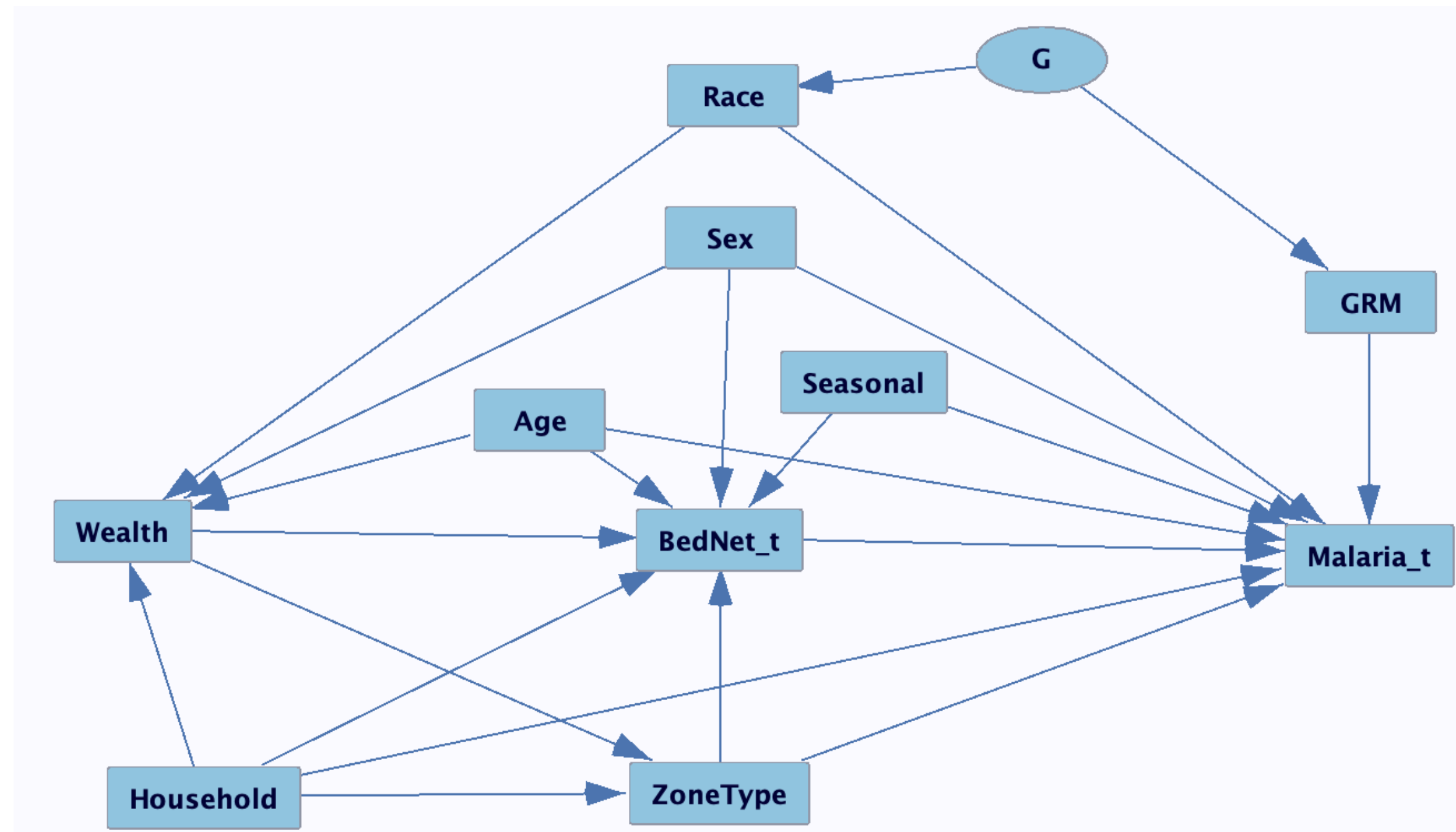
- Selectively removes only rows with missing values in variables involved in the current test
- Unlike list-wise deletion, it **preserves more data** by not discarding entire rows
- Assumes data are **Missing Completely At Random (MCAR)**
- More efficient in large datasets with low missingness (<5%)
- Balances simplicity and statistical power

Handling Missing Data in Conditional Independence Tests

Multiple Imputation (MI)

- Designed for data **Missing At Random (MAR)** — where missingness depends only on **observed variables**
 - Uses a predictive model to generate **multiple plausible values** for missing entries
 - Produces several **complete datasets**, which are analyzed separately
 - Final results are **pooled**, accounting for uncertainty due to imputation
 - Offers greater flexibility and validity than deletion-based methods, especially in **moderate to high** missingness scenarios
- micd R package: Multiple Imputation for Causal Graph Discovery — ([Link](#))

Going Back to Our Simple Example...



True Causal Diagram
(unknown)

In many real-world settings, observations are **not independent**:

- **Longitudinal data:** repeated measures per individual
- **Genetic / family data:** related individuals
- **Household data:** shared environment

Classical conditional independence tests assume **independent observations**.

How can we test $X \perp\!\!\!\perp Y | \mathbf{S}$ accounting for dependencies due to household, genetic, and longitudinal factors? e.g. X (BedNet) and Y (Malaria)?

How can we handle unit-level dependencies?

- **Genetic Relatedness:** Restrict analyses to a subset of genetically unrelated individuals (e.g., 808 pseudo-independent individuals in the Mâncio Lima cohort – $GRM_{ij} < 0.125$).
- **Temporal Autocorrelation:** Transform longitudinal measures into count outcomes and model them using **zero-inflated negative binomial regression** to account for overdispersion and excess zeros.
- **Household Clustering:** Not explicitly modeled due to weak or negligible effect observed in preliminary analyses.

These are limited by reduced sample size, loss of temporal granularity, and potential residual biases from unmodeled dependencies.

Our proposal: to formulate likelihood ratio-based conditional independence tests using **generalized linear mixed models (GLMMs)**, explicitly modeling within-group dependence through random effects.

How can we handle unit-level dependencies?

Classical conditional independence tests assume **independent observations**.

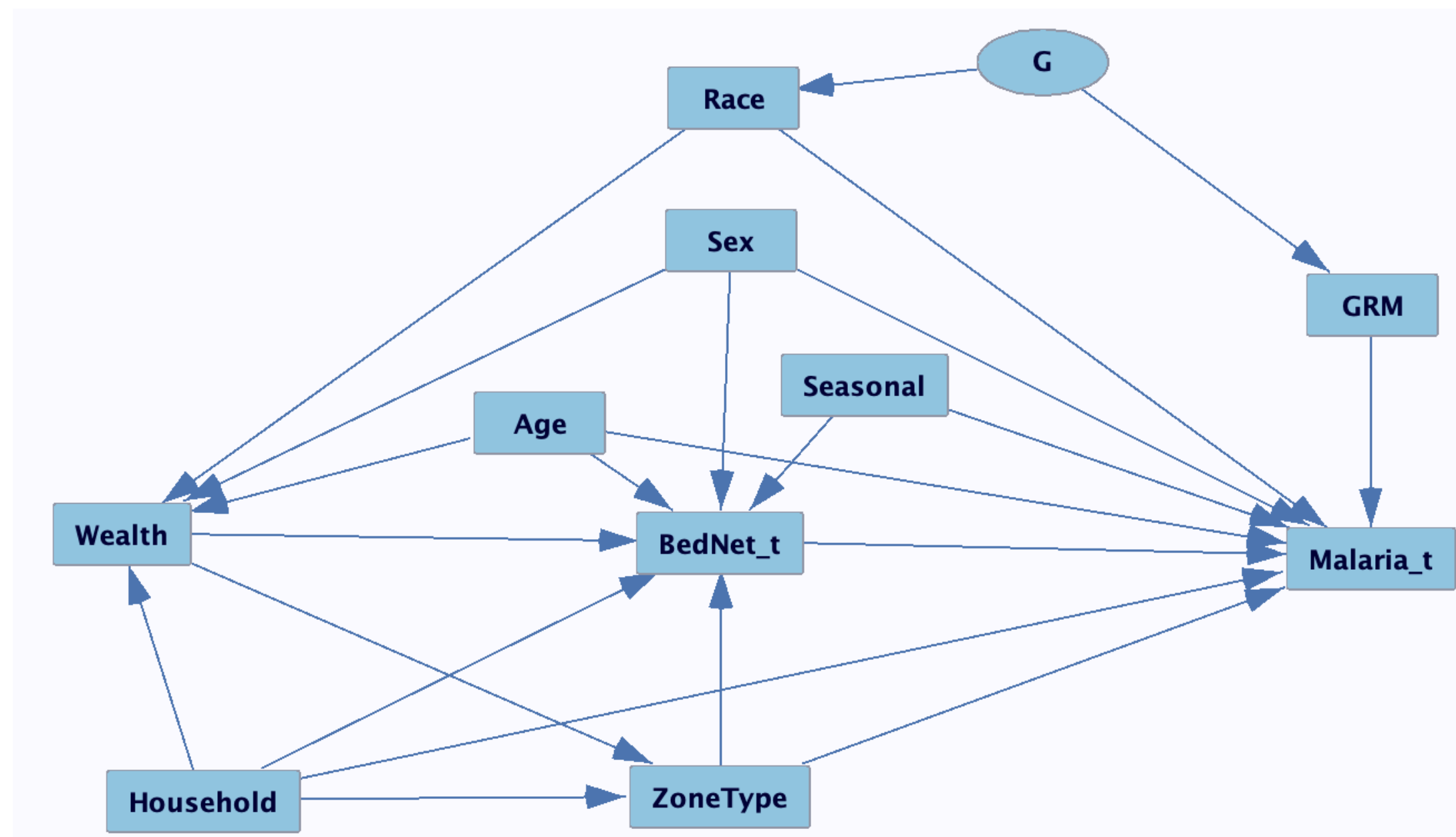
In many real-world settings, observations are **not independent**:

- **Longitudinal data:** repeated measures per individual
- **Genetic / family data:** related individuals
- **Household data:** shared environment

The literature remains limited when it comes to modeling complex data structures!

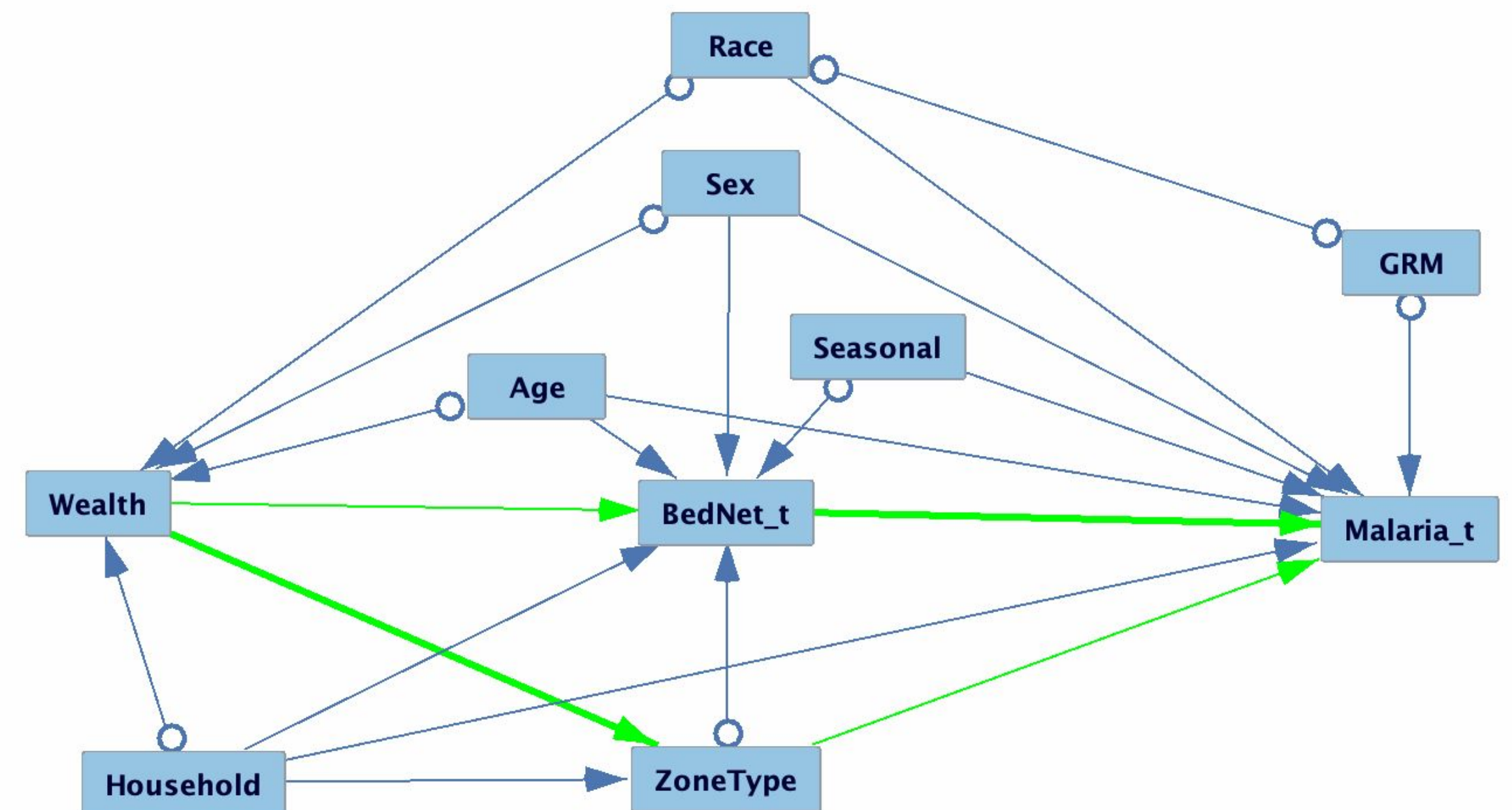
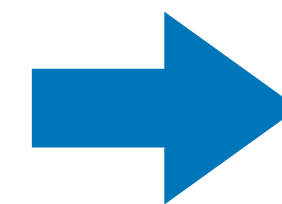
Modeling unit-level dependencies via GLMMs

When applied with tests that appropriately account for multiple dependence structures in the data, FCI can infer the correct:



True Causal Diagram
(unknown)

FCI



Partial Ancestral Graph
(learned from data)

Modeling unit-level dependencies via GLMMs

Likelihood ratio test for $X \perp\!\!\!\perp Y | \mathbf{S}$ with dependent data: $D = \{x_{it}, y_{it}, s_{kit} : k = 1, \dots, |\mathbf{S}|; i = 1, \dots, n\}$

Null model \mathcal{M}_0 :

$$g(E[Y_{it} | \mathbf{S}_i]) = \eta_{it}^{(0)} = \alpha + \sum_{k=1}^{|\mathbf{S}|} \beta_{S_k} s_{kit} + \beta_t \cdot t_i + \underbrace{u_{0_i} + Z_{h(i)} u_{0_h} + Z_{g(i)} u_{0_g}}_{\text{Basal, household, and genetic random intercepts}} + \underbrace{(u_{1_i} + Z_{h(i)} u_{1_h} + Z_{g(i)} u_{1_g}) \cdot t_i}_{\text{Basal, household, and genetic random slopes for time}}$$

Alternative model \mathcal{M}_1 :

$$g(E[Y_{it} | X_{it}, \mathbf{S}_{it}]) = \eta_{it}^{(1)} = \alpha + \beta_X x_{it} + \sum_{k=1}^{|\mathbf{S}|} \beta_{S_k} s_{kit} + \beta_t \cdot t + \underbrace{u_{0_i} + Z_{h(i)} u_{0_h} + Z_{g(i)} u_{0_g}}_{\text{Basal, household, and genetic random intercepts}} + \underbrace{(u_{1_i} + Z_{h(i)} u_{1_h} + Z_{g(i)} u_{1_g}) \cdot t}_{\text{Basal, household, and genetic random slopes for time}}$$

Where:

- $Z_{h(i)}$ is the i -th row of the $n \times D$ incidence matrix, with $Z_{h(i,j)} = 1$ if individual i is in household j , and 0 otherwise.
- $Z_{g(i)}$ is the i -th row of the $n \times n$ matrix Z_g , such that $Z_g Z_g^T = \text{GRM}$

$$u_{0_i}, u_{1_i} \sim \mathcal{N} \left(\mathbf{0}, \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} \\ \sigma_{0,1} & \sigma_1^2 \end{pmatrix} \right); \quad u_{0_h}, u_{1_h} \sim \mathcal{N}_{2D} \left(\mathbf{0}, \mathbf{I}_D \otimes \begin{pmatrix} \sigma_{h_0}^2 & \sigma_{h_{0,1}} \\ \sigma_{h_{0,1}} & \sigma_{h_1}^2 \end{pmatrix} \right); \quad u_{0_g}, u_{1_g} \sim \mathcal{N}_{2n} \left(\mathbf{0}, \mathbf{I}_n \otimes \begin{pmatrix} \sigma_{g_0}^2 & \sigma_{g_{0,1}} \\ \sigma_{g_{0,1}} & \sigma_{g_1}^2 \end{pmatrix} \right)$$

LRT for Fixed Effects as Conditional Independence Test

For **tests involving a fixed effects** X for Y , it follows as before:

Under $H_{0_1} : P(Y|X, \mathbf{Z}) = P(Y|\mathbf{Z})$ — case in which $\beta_x = 0$ — and regularity conditions:

$$\Lambda = -2(\ell_0 - \ell_1) \xrightarrow{d} \chi_k^2, \quad \text{with } k = \dim(\beta_X).$$

Compute p-value p_1 corresponding to H_{0_1} .

For symmetric testing, test $H_{0_2} : P(X|Y, \mathbf{S}) = P(X|\mathbf{S})$ and obtain p_2

Compute the combined p-value: $p = \min\{2 \times \min(p_1, p_2), \max(p_1, p_2)\}$

Interpretation: $p < \alpha \rightarrow$ evidence that $X \not\perp\!\!\!\perp Y | \mathbf{S}$

LRT for Random Effects as Conditional Independence Test

For **tests involving a latent group variable** $U_g \sim \mathcal{N}(0, \sigma_g^2)$ (e.g., genetic, household) for Y :

Under $H_{0_1} : P(Y | U_g, \mathbf{S}) = P(Y | \mathbf{S})$ — case in which $\sigma_g^2 = 0$:

$$\Lambda = -2(\ell_0 - \ell_1) \xrightarrow{d} \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2 \text{ when testing a single random effect (*).}$$

Compare the likelihood ratio statistic or use **parametric bootstrap** to obtain p-value p_1 .

No additional test is required to enforce symmetry.

Interpretation: $p_1 < \alpha \implies$ evidence that $X \not\perp\!\!\!\perp U_g | \mathbf{S}$

(*) The **asymptotic distribution** under the null is often a **mixture of chi-squares** (Self & Liang, 1987).

Modeling Assumptions and Interpretation

- **Assumptions:**
 - Random effects are assumed to be **exogenous**, i.e., not caused by any observed variables in the model.
 - Random effects are assumed to be independent of each other if not modeled otherwise.
- **Interpretation:** A **significant random effect** for outcome variable Y implies the presence of a **statistical dependency** between the random effect and Y .
- **Implication:** An **edge is inferred** between the random effect and Y only if the random effect remains significant when including all adjacent variables (i.e., all significant fixed and random effects of Y).

Implementing Conditional Independence Tests with Complex Dependencies

We can use established **R packages** in both Bayesian and frequentist frameworks:

- Bayesian approach: `brms`
 - Estimation via Markov Chain Monte Carlo (MCMC)
 - Highly flexible formula syntax
 - Supports a wide range of response distributions (e.g., binomial, zero-inflated models)
 - Computationally intensive; scales less efficiently with large samples or random effects
- Frequentist approach: `GMMAT`
 - Estimation via maximum likelihood (with penalized quasi-likelihood for GLMMs)
 - Specifically designed for genetic studies
 - More efficient for large datasets

Causal Identification from Equivalence Classes



Can we identify causal effects from an equivalence class?

Effect Identification from PAGs:

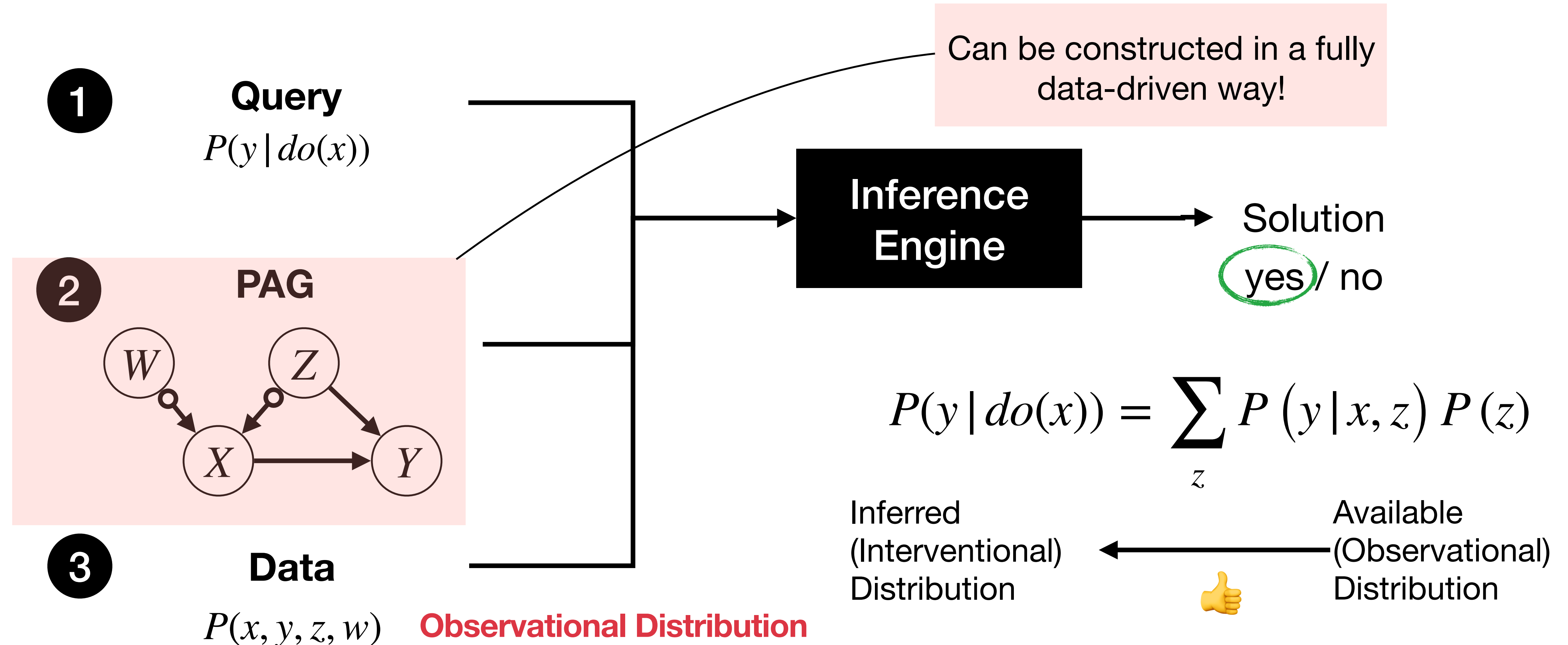
For Covariate Adjustment, we can use the Generalized Adjustment Criterion.

Recently, we proposed complete calculus and algorithms for the identification of marginal and conditional causal effect in PAGs!

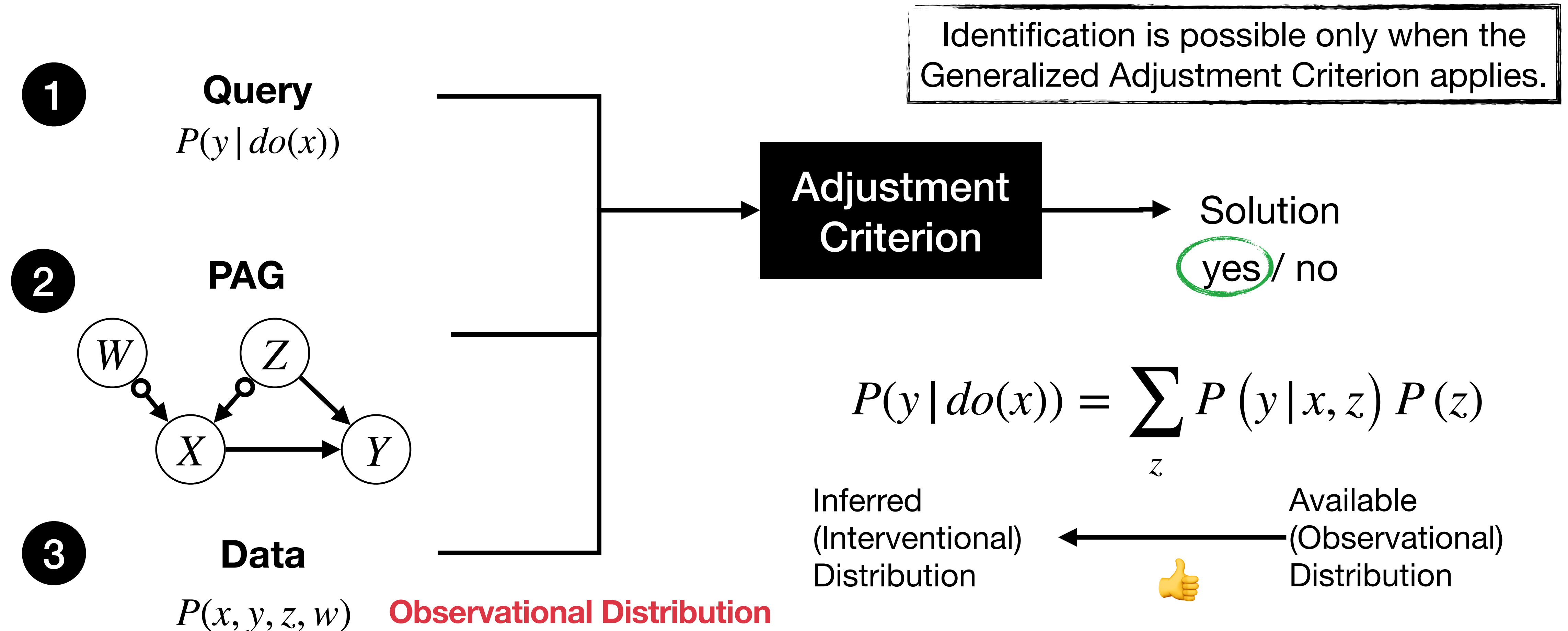
Perkovic, E., Textor, J. C., Kalisch, M., & Maathuis, M. H. (2018). Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs. Journal of Machine Learning Research 18 (2018) 1-62

Jaber A., **Ribeiro A. H.**, Zhang, J., Bareinboim, E. (2022) Causal Identification under Markov Equivalence - Calculus, Algorithm, and Completeness. In Proceedings of the 36th Annual Conference on Neural Information Processing Systems, NeurIPS. ([Link](#))

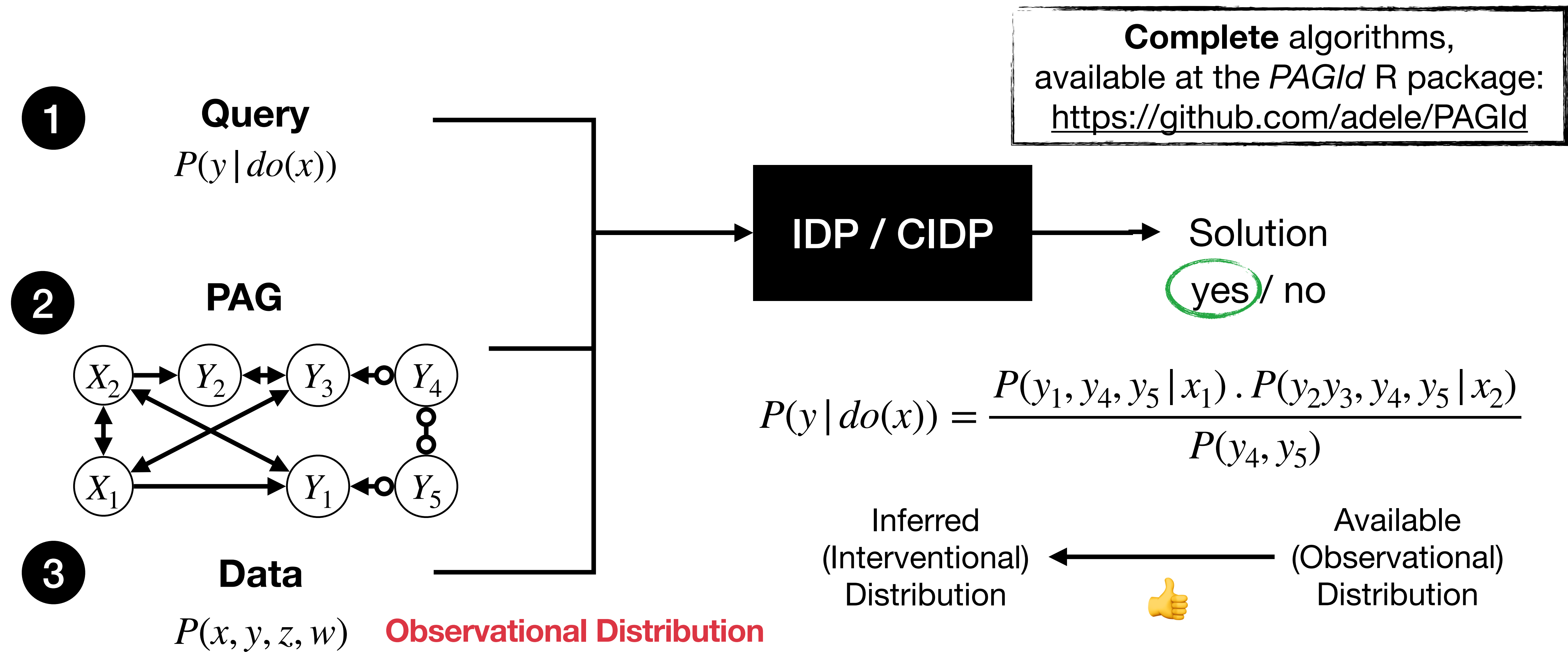
Effect Identification in Markov Equivalence Classes



Identification via Adjustment in Markov Equivalence Classes



General Identification in Markov Equivalence Classes



Coding Exercises

Causality Tutorial:



- **Source Code in R** and slides available on GitHub: [RBras 2025](#)

Check Part III: Causal Discovery using FCI

Check Part IV: Causal Effect Identification from the Markov Equivalence Class

1. Generalized Backdoor / Adjustment Criterion -- pcalg R package
2. CIDP Algorithm -- pagID R package

Educational Resources

- Causality Lectures on YouTube  : [@adelehelena](#)
 - Complete course (13 lectures at HHU): [Playlist](#)
 - 3-hour tutorials at summer schools:
 - Lisbon Machine Learning School (LxMLS): [Playlist \(2021-2025\)](#)
 - Nordic Probabilistic AI School (ProbAI): [Playlist \(2023-2024\)](#)
 - European Summer School on Artificial Intelligence (ESSAI 2024): [Playlist](#)
- Tutorial on GitHub  : [@adele](#) → [Causality-Tutorial](#)
 - Causal Discovery — Google Colab Notebook: [Link](#)
 - Causal Effect Identification — Google Colab Notebook: [Link](#)

Thank you! :)

Feel free to reach out to me if you have any questions:

adele.ribeiro@uni-muenster.de | pavan@ime.usp.br