Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Can one extract causal information from high-dimensional observational data?

Markus Kalisch

Seminar of Statistics, ETH Zurich, Switzerland

Joint work with

Peter Bühlmann

Marloes Maathuis

Diego Colombo

What is a causal effect?

What is a causal effect?

Drowning accidents

What is a causal effect?

Drowning accidents

Ice cream sales

What is a causal effect?

Drowning accidents

Ice cream sales

What is a causal effect?

Drowning accidents

Ice cream sales

What is a causal effect?

Drowning accidents

Ice cream sales

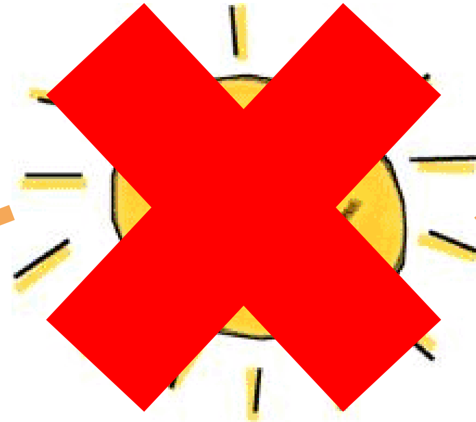
?

What is a causal effect?

Drowning accidents

Ice cream sales

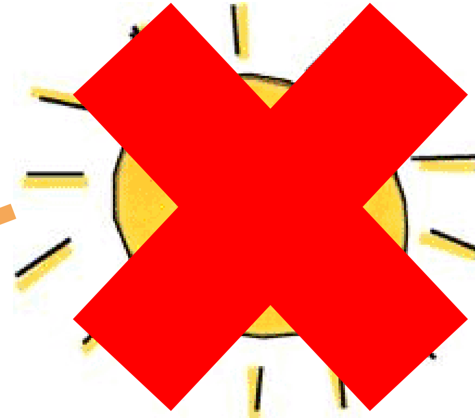
What is a causal effect?



Drowning accidents

Ice cream sales

What is a causal effect?



Drowning accidents

Ice cream sales

Another example: Smoking

Scenario 1: Observe 1000 smokers and count the incidence of lung cancer

Scenario 1: Observe 1000 smokers and count the incidence of lung cancer

Scenario 2: Make 1000 random people smoke and count the incidence of lung cancer

Scenario 1: Observe 1000 smokers and count the incidence of lung cancer

Scenario 2: Make 1000 random people smoke and count the incidence of lung cancer

are different.

What is a causal effect?

CHANGE

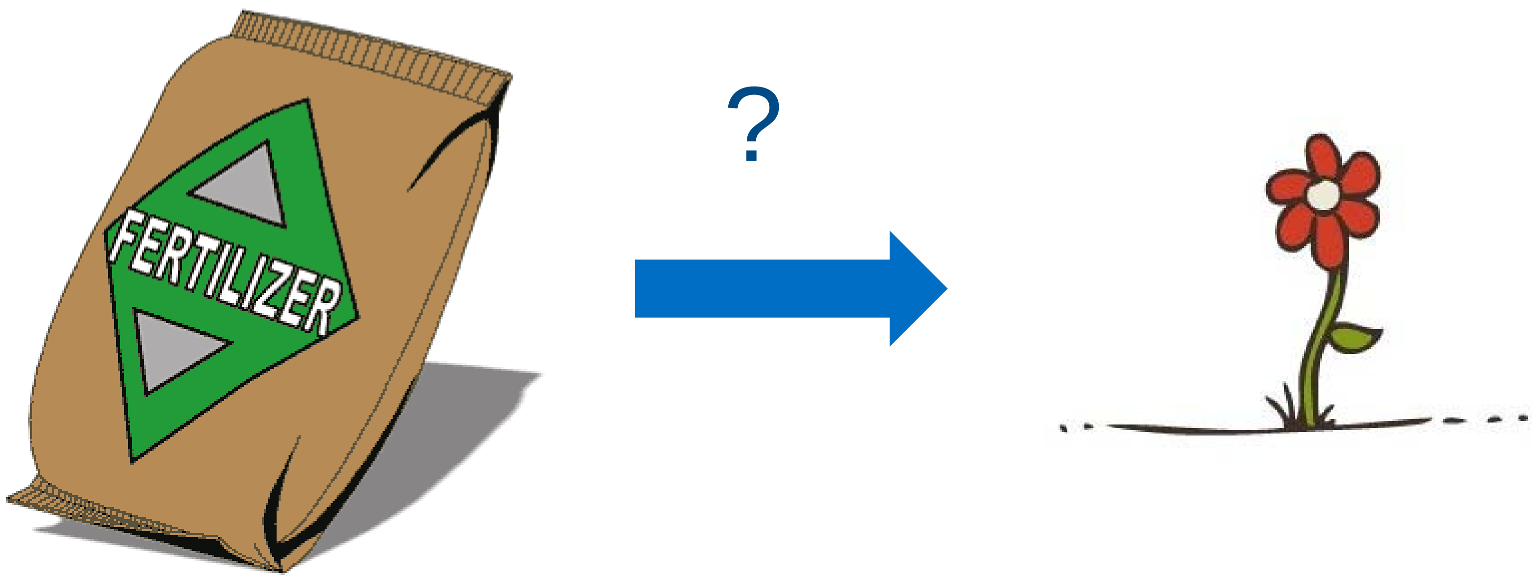
BY

INTERVENTION

How to find causal effects?

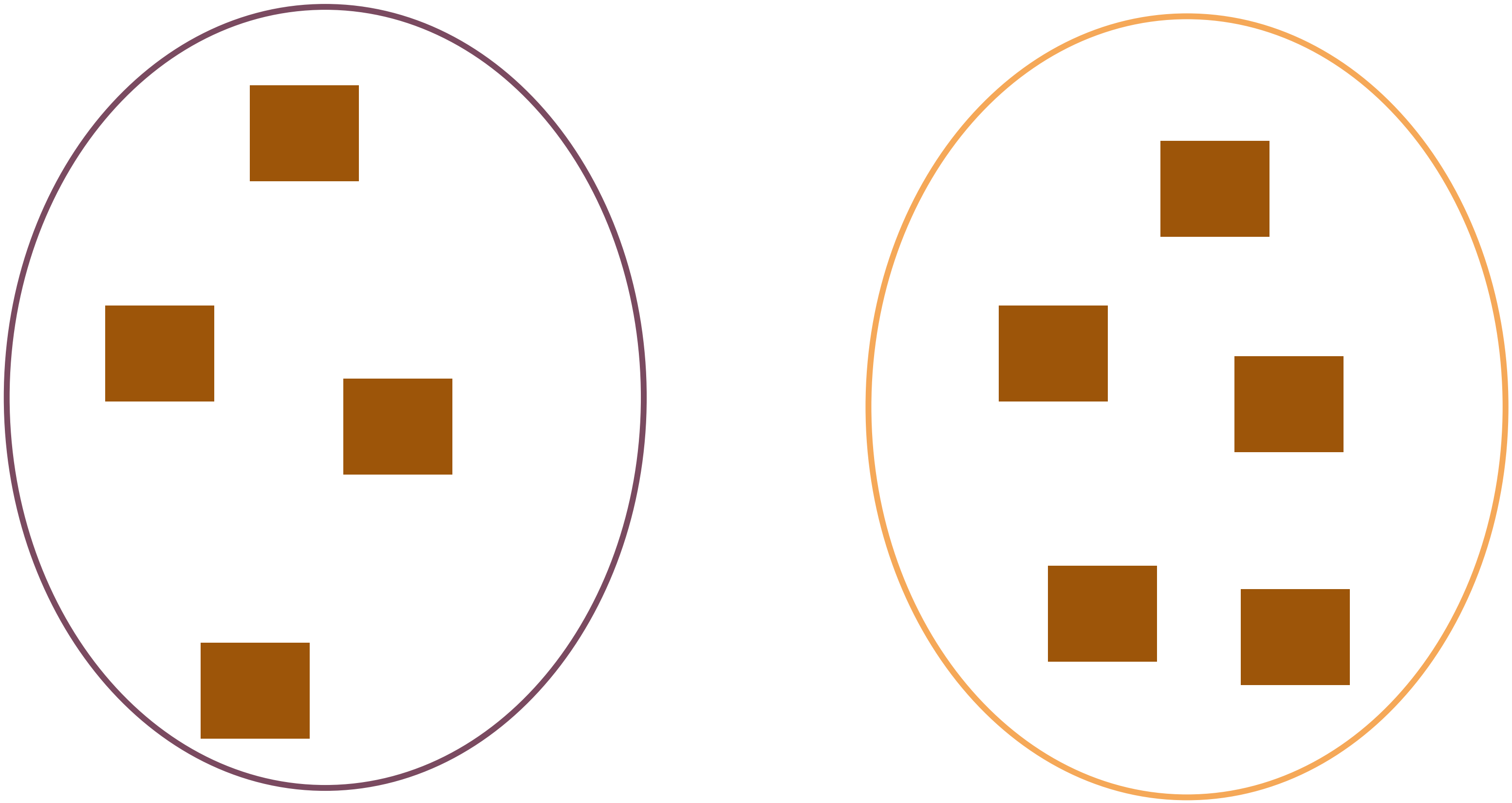
How to find causal effects?

Experimental Data



How to find causal effects?

Experimental Data

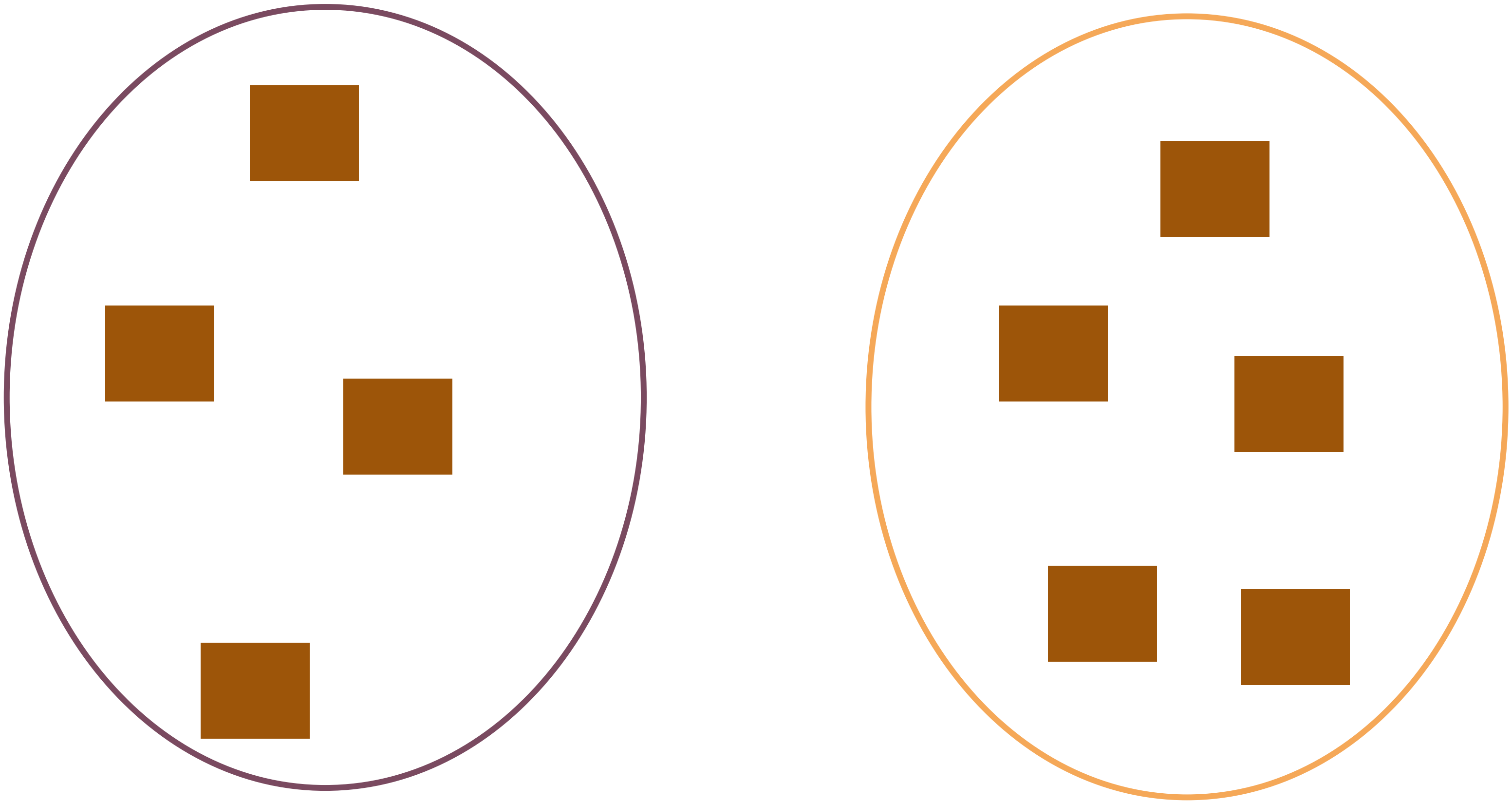


Two groups of plots: Identical in all aspects (sunlight, water, soil quality, …)

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

How to find causal effects?

Experimental Data

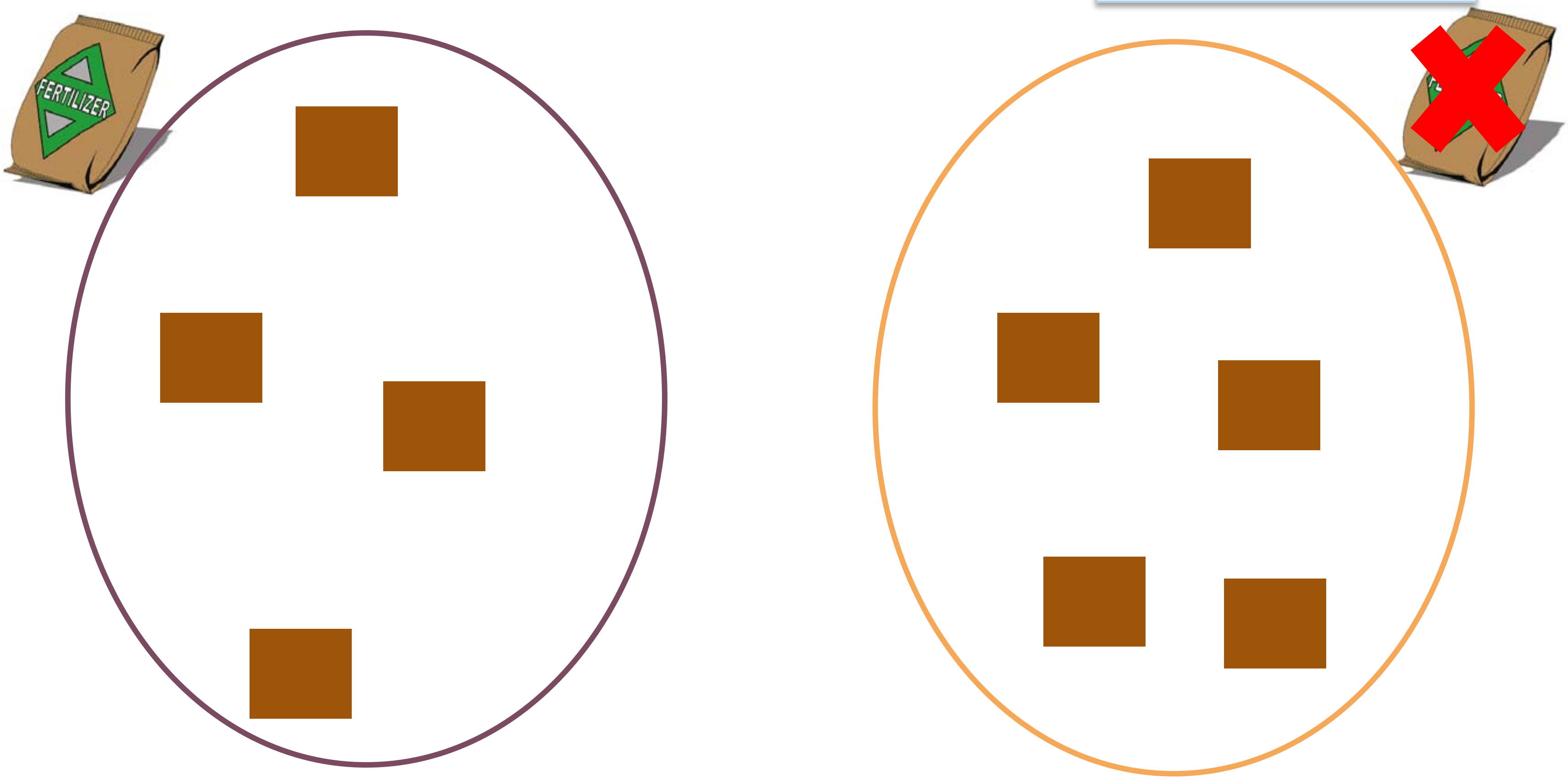


Two groups of plots: Identical in all aspects (sunlight, water, soil quality, …)

Practice: Randomized assignment

How to find causal effects?

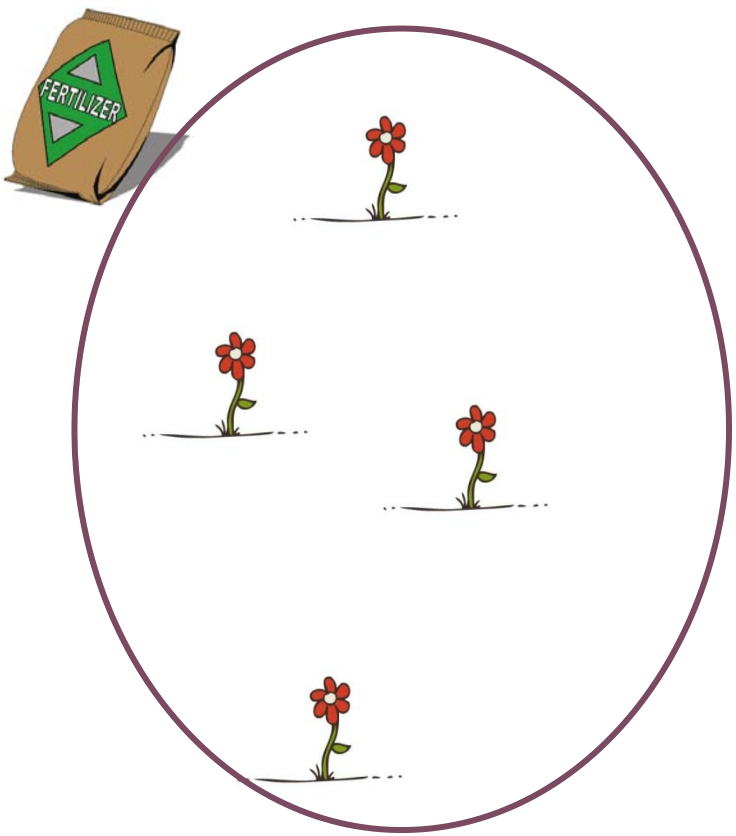
Experimental Data

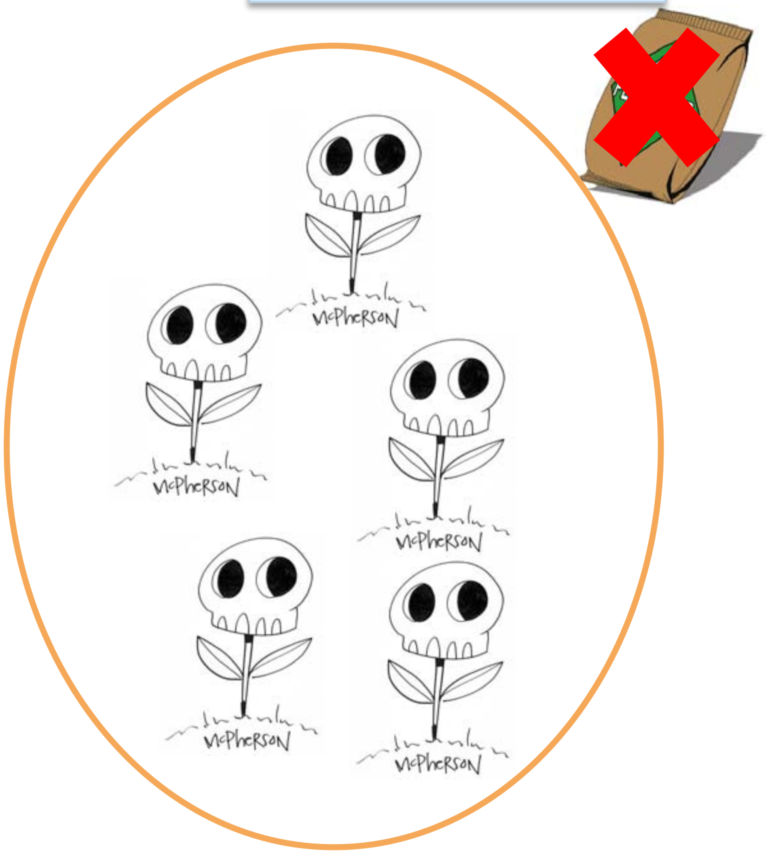


ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

How to find causal effects?

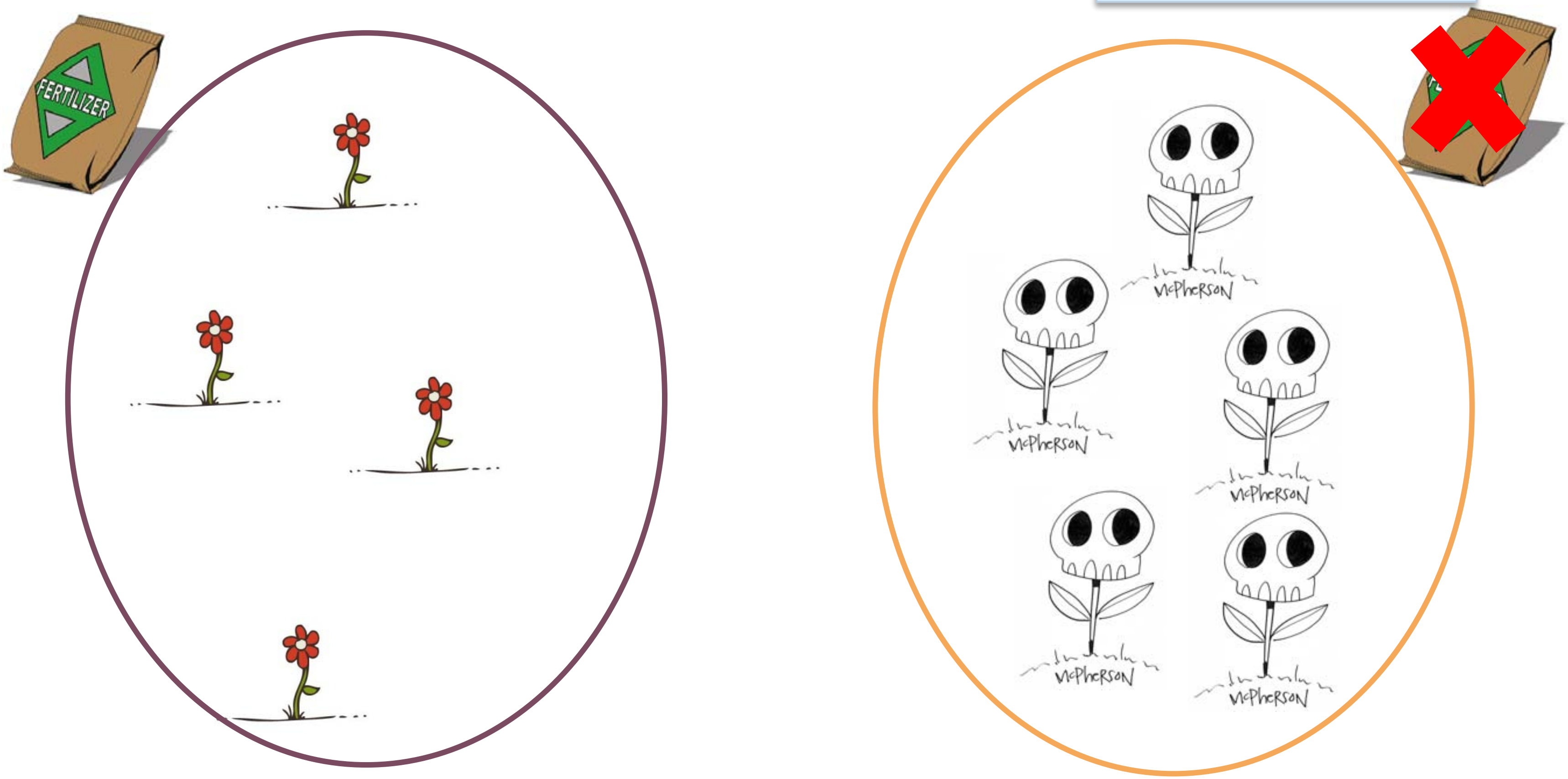
Experimental Data





How to find causal effects?

Experimental Data



Outcome due to fertilizer,
since everything else was equal

How to find causal effects?

Sometimes, randomized controlled experiments are

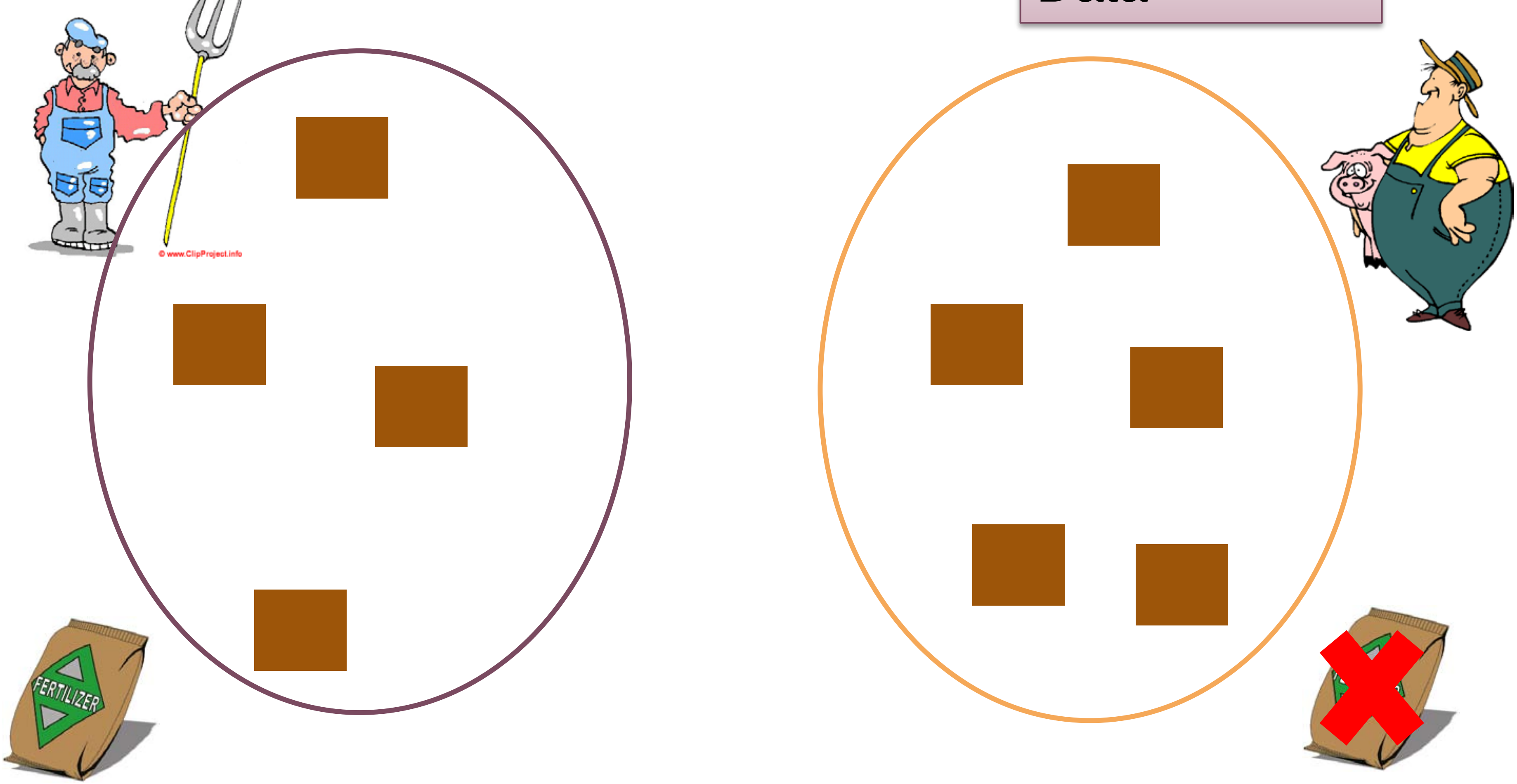
- too expensive (gene experiments)
- too time-consuming (gene experiments)
- unethical (HIV treatment)
- just not practical (smoking).

If experiment is impossible…

Observational Data

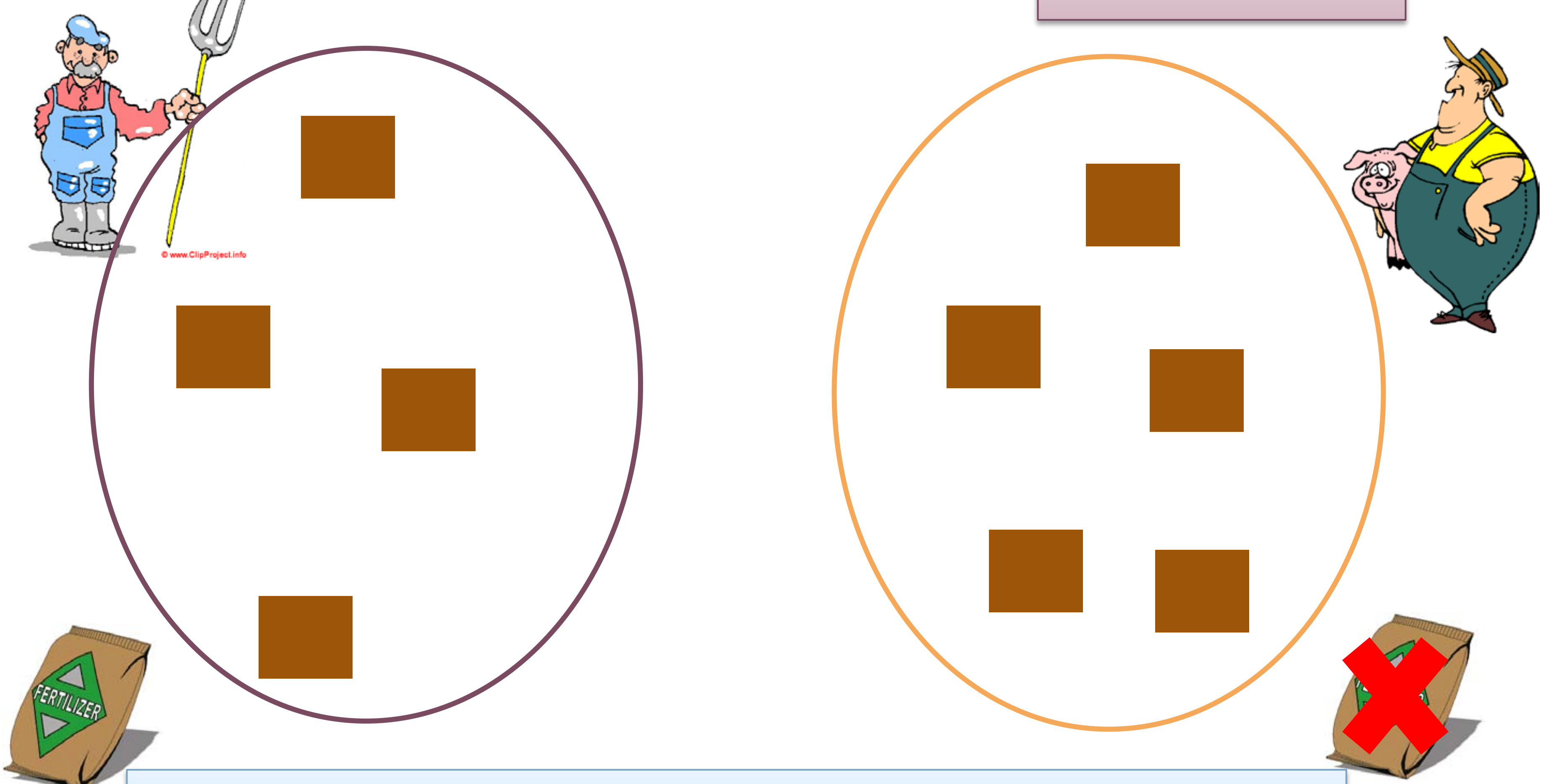
… observe fields of two farmers.

Observational Data



… observe fields of two farmers.

Observational Data

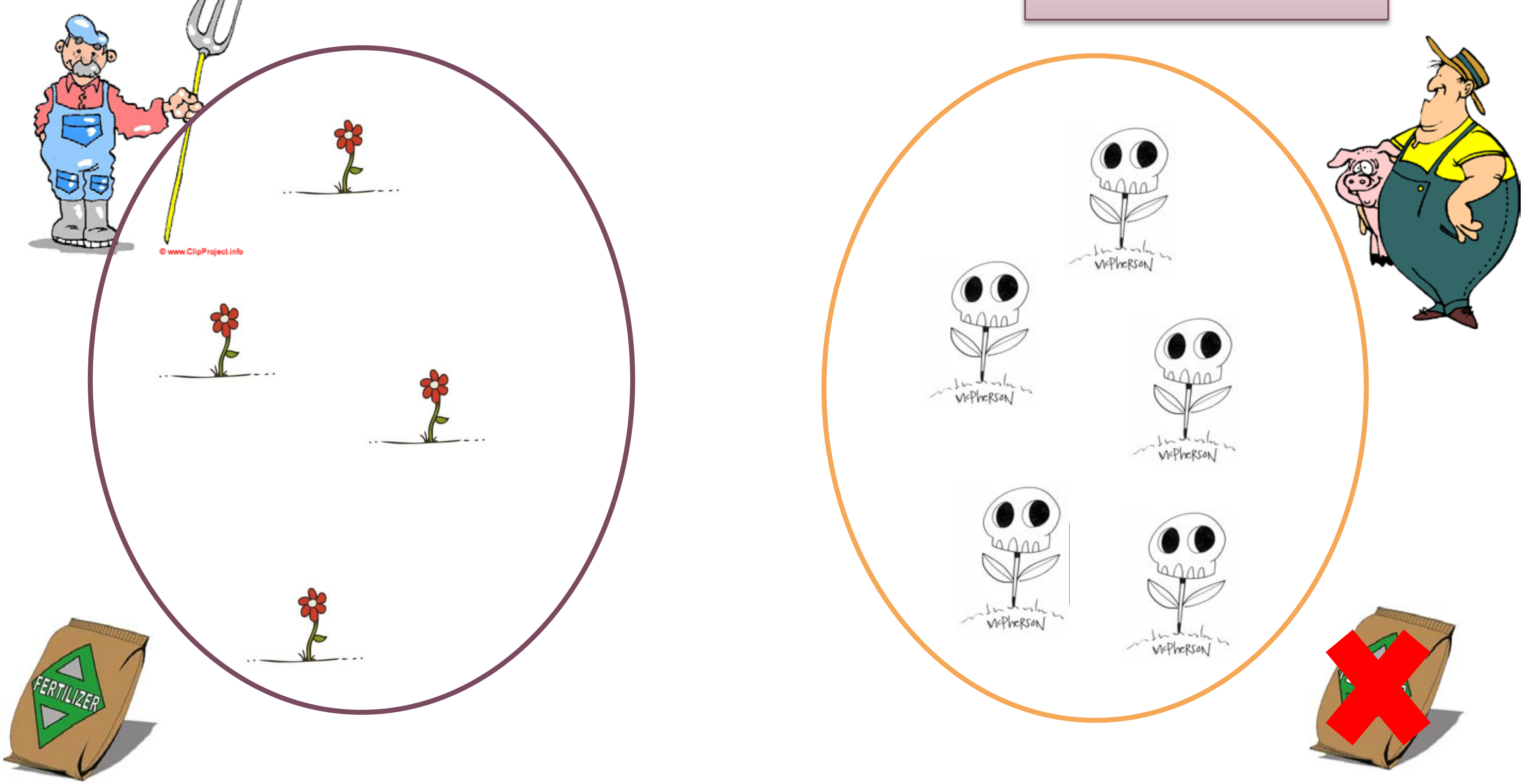


Groups not guaranteed

to be identical in all aspects (sunlight, water, soil quality, …)

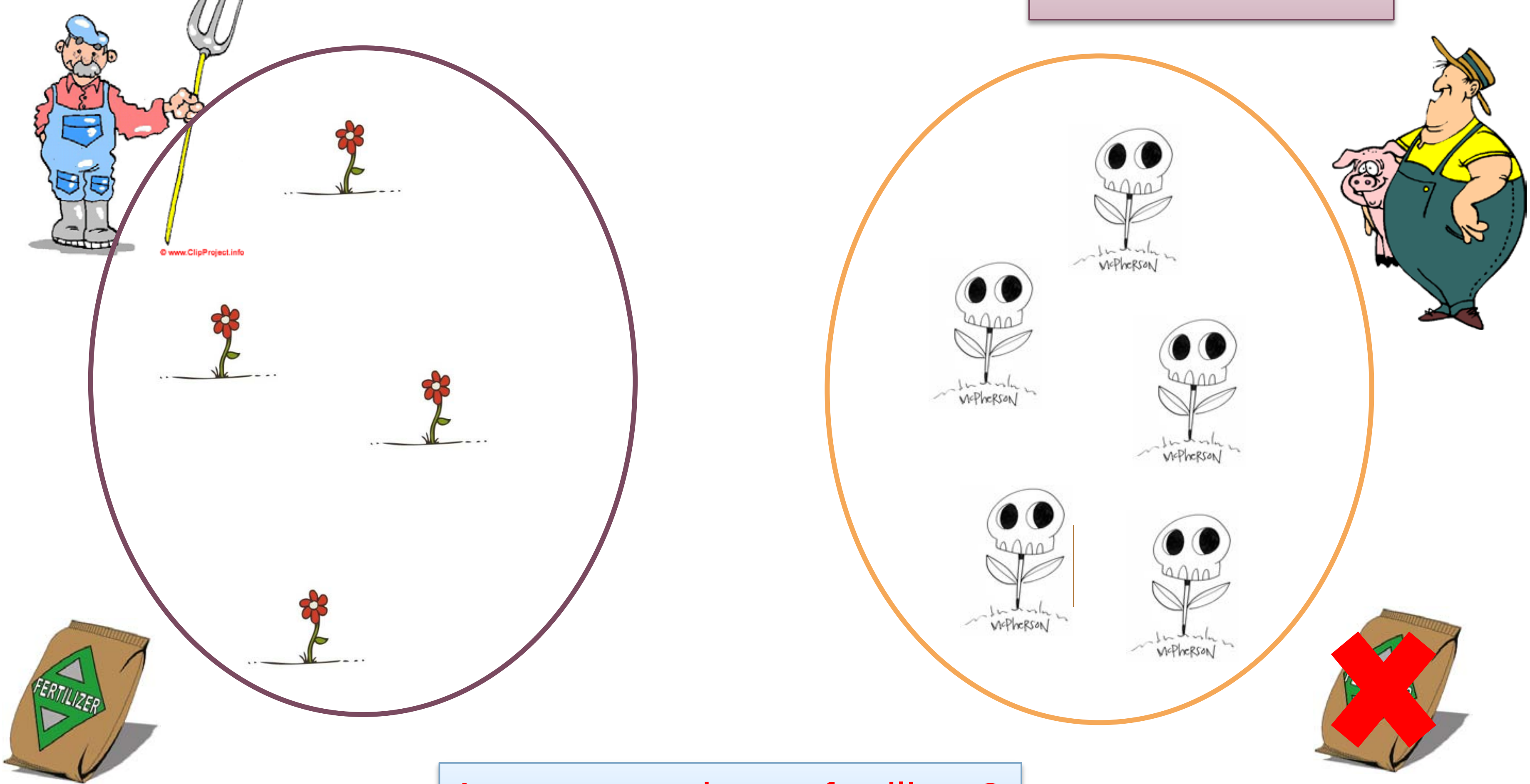
Observational Data

… observe fields of two farmers.



… observe fields of two farmers.

Observational Data

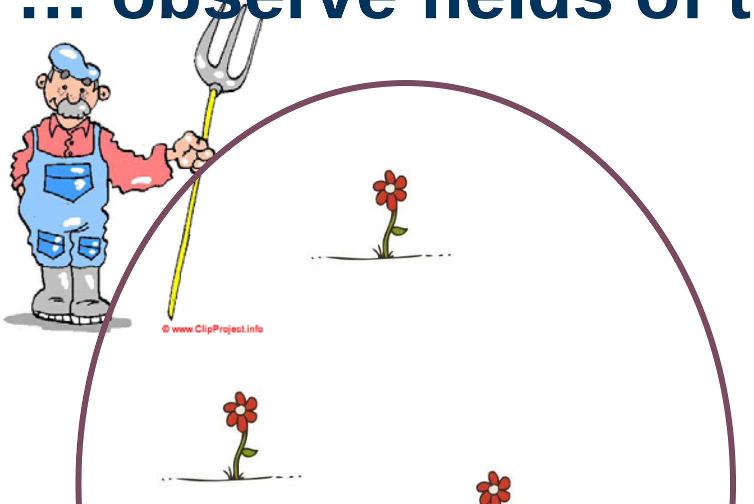


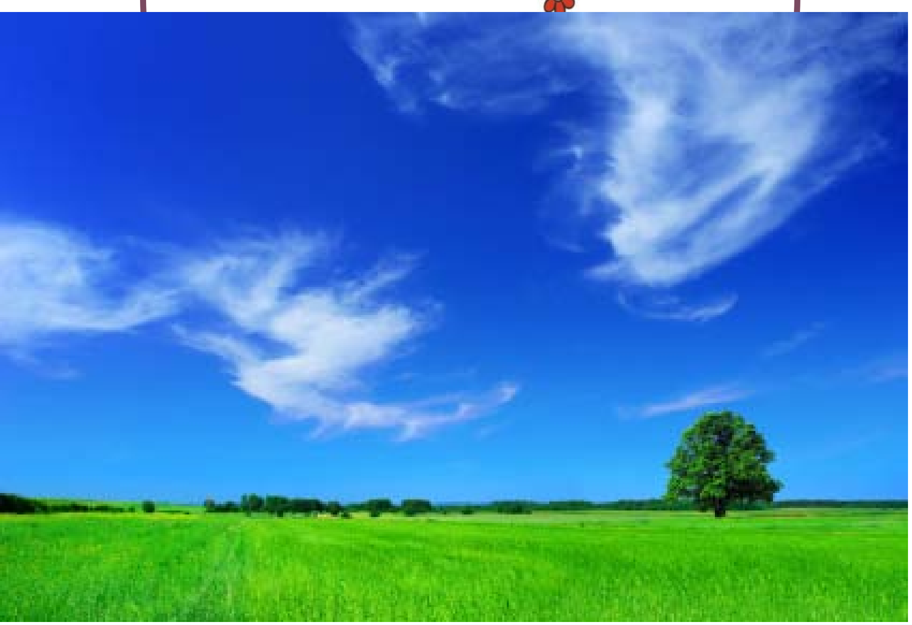
Is outcome due to fertilizer?

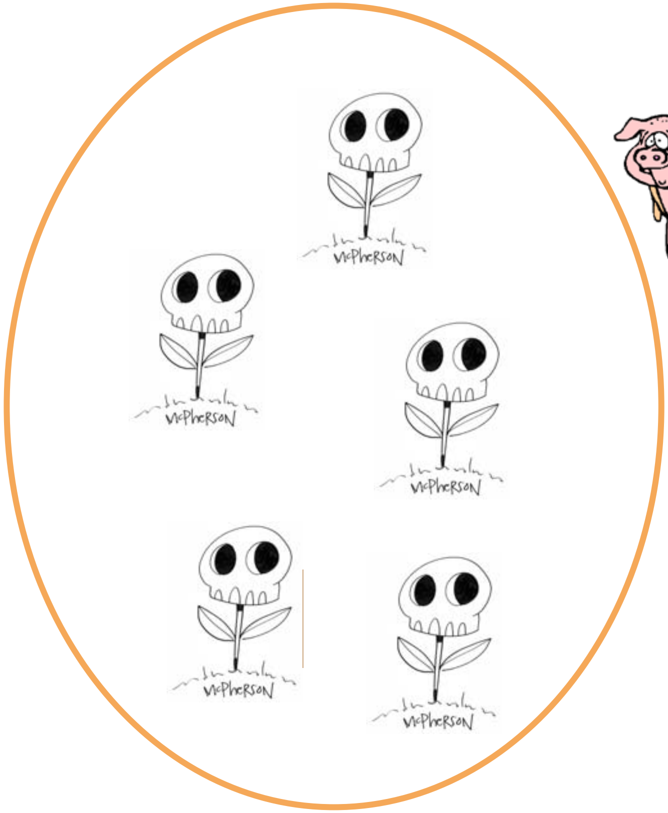
We can’t tell !

… observe fields of two farmers.

Observational Data

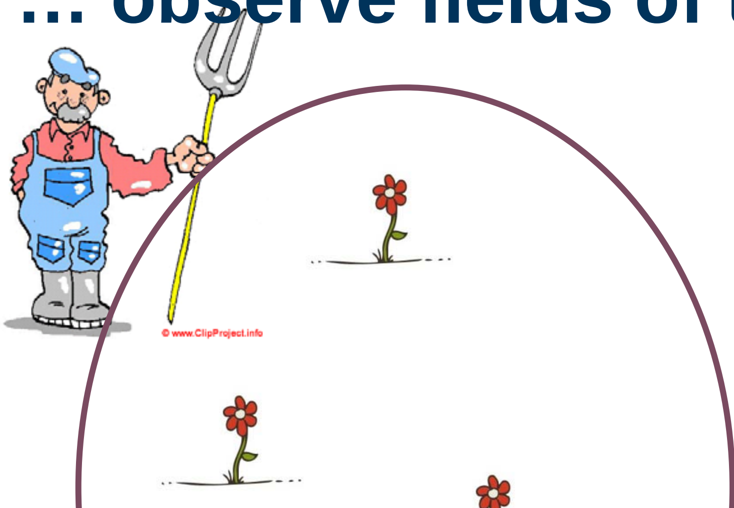


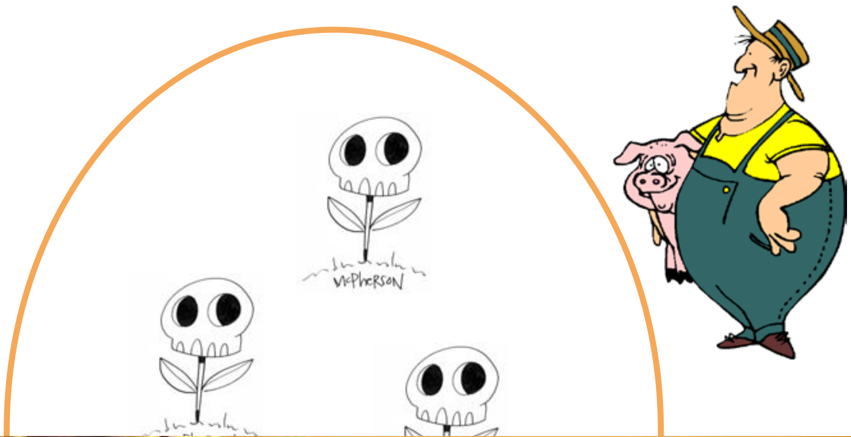


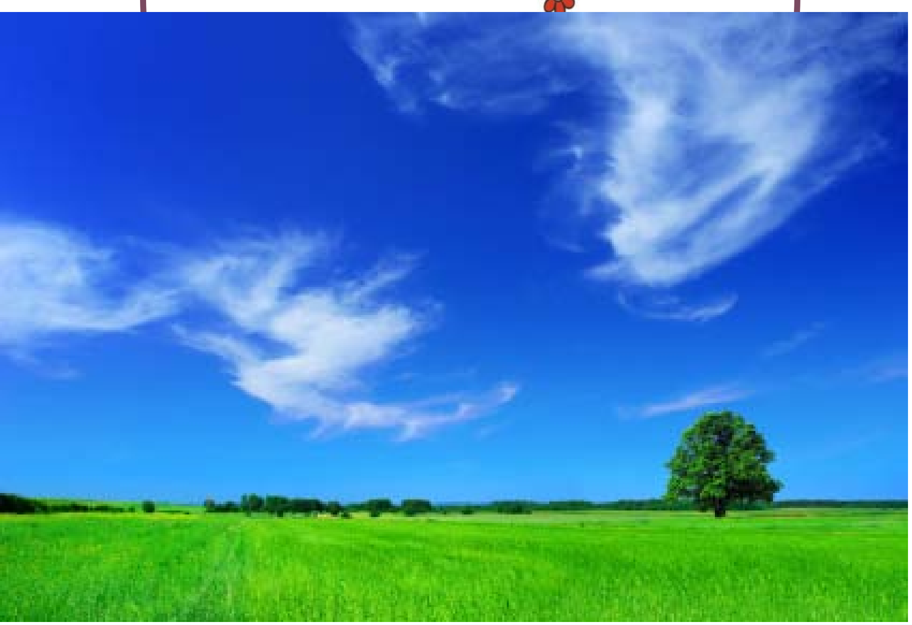


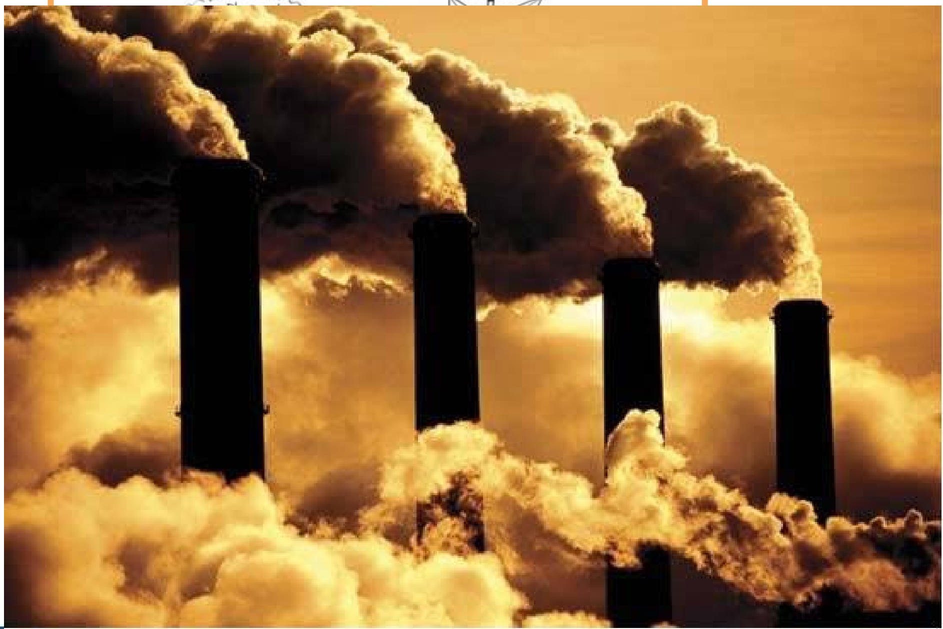
… observe fields of two farmers.

Observational Data









How to find causal effects?

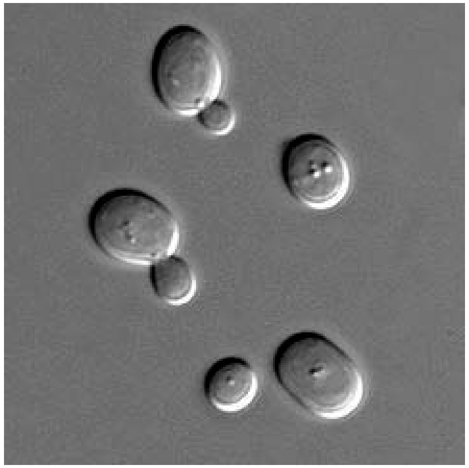
Can one extract causal information from observational data alone?

Goal of this talk

IDA

- **IDA** finds a **set of possible causal effects given observational data** consistently even in high dimensions.
- One element of the set is the true causal effect; **bounds** on set are useful
- Does not replace randomized experiments
- Helps **prioritizing and designing random experiments**

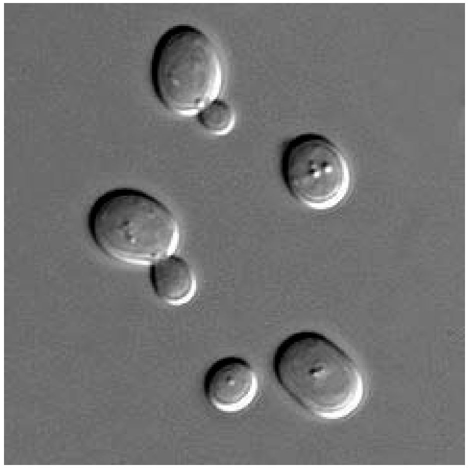
Example



- Yeast: Saccharomyces cerevisiae

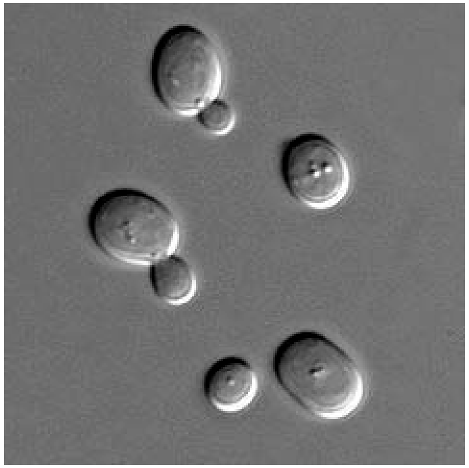
Example

- Yeast: Saccharomyces cerevisiae

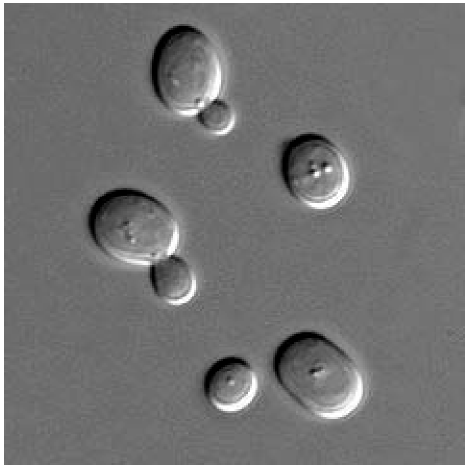


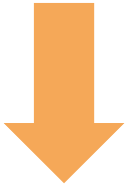
Example

- Yeast: Saccharomyces cerevisiae
- What are the causal effects among the thousands of genes?



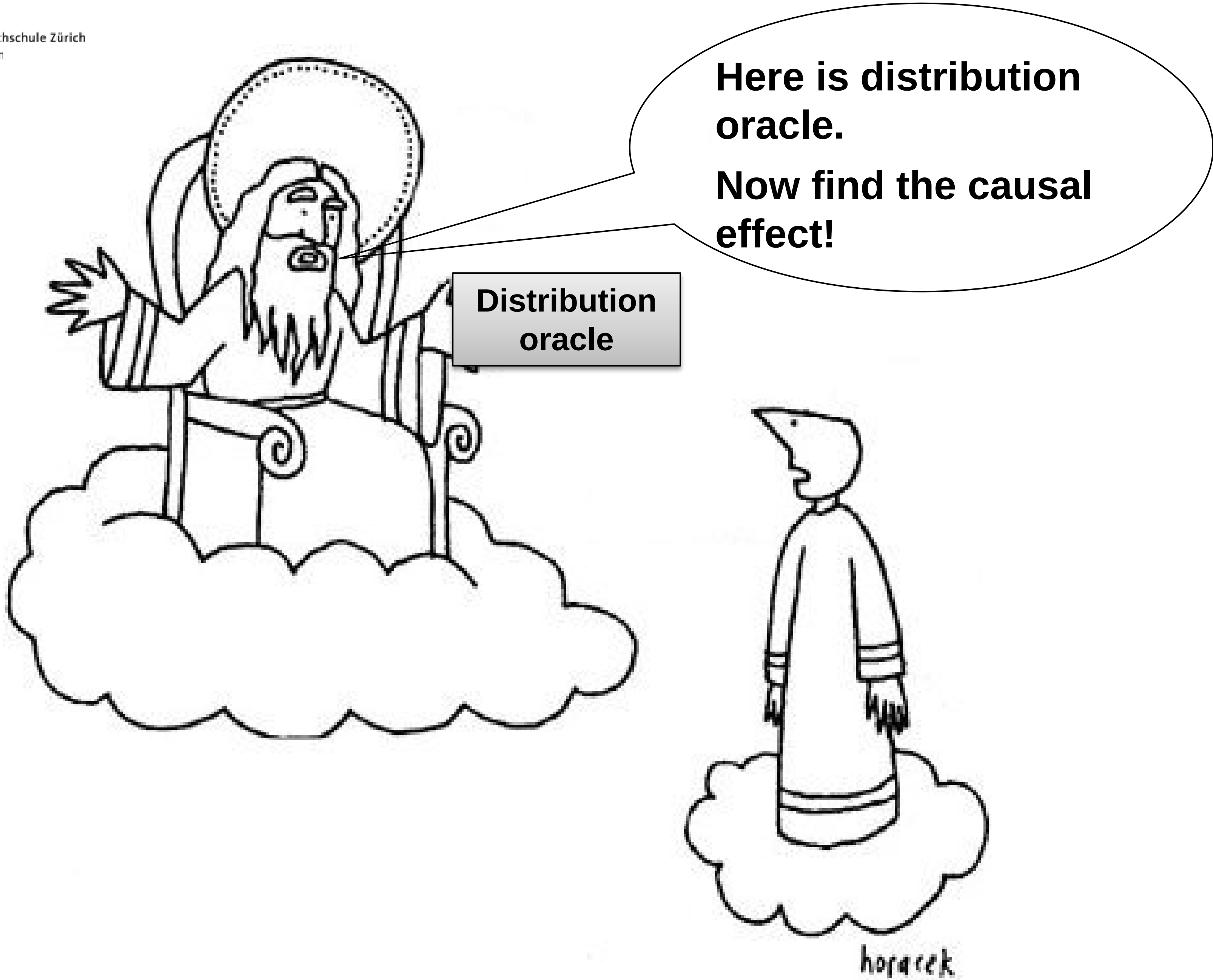
Example



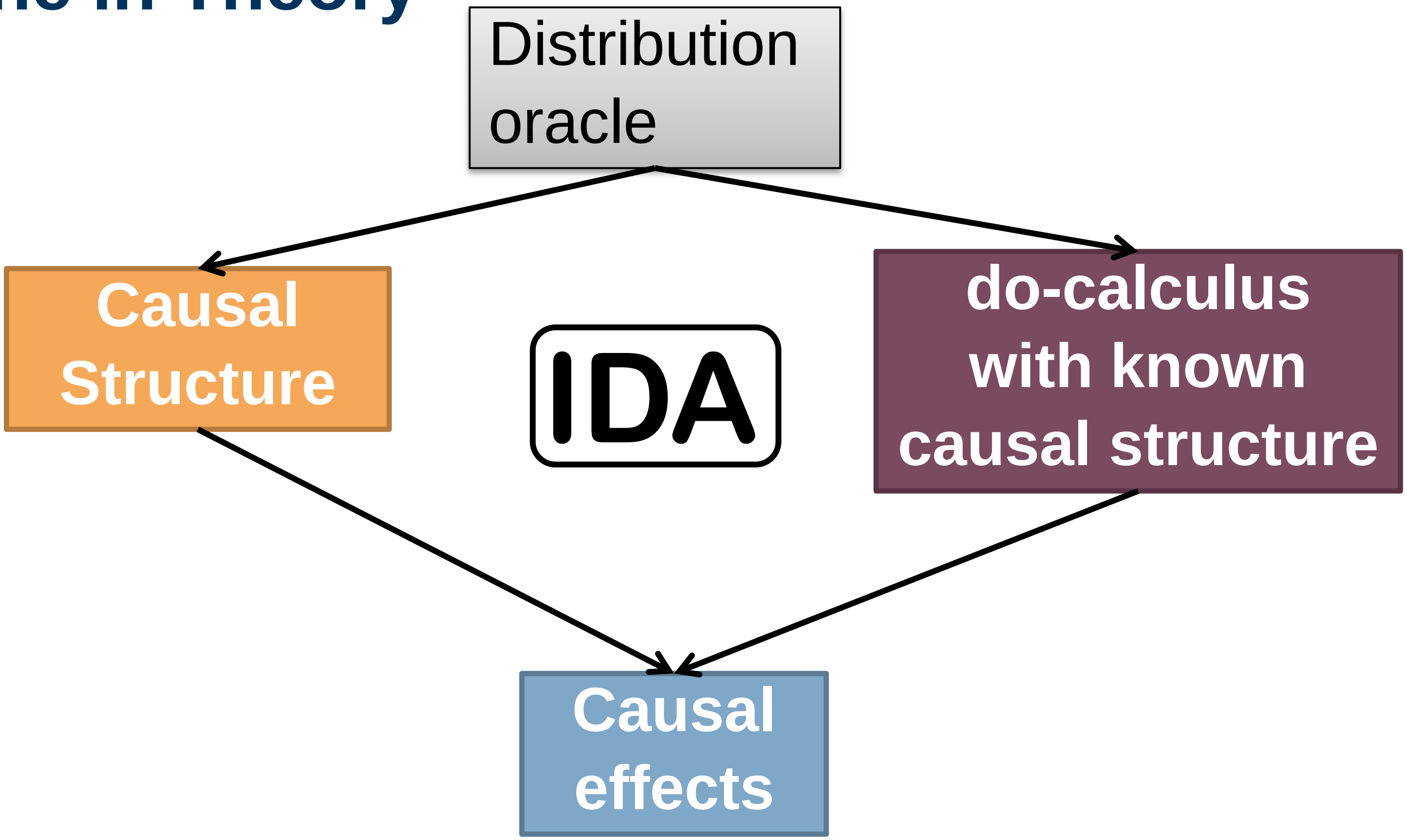


- Yeast: Saccharomyces cerevisiae
- What are the causal effects among the thousands of genes?
- **Approach:**
  Model gene expression of each gene as a random variable.

  Can we use the
  joint distribution of gene expression
  to extract
  causal information?


Here is distribution oracle.
Now find the causal effect!
Distribution oracle
horacek

Outline in Theory



do-calculus with known causal structure

Pearl's do-operator

- Notation for causal intervention

  $\boxed{P(Y=y \mid do(X=x))}$

  "distribution of Y, if there is an intervention in variable X"
- Causal effect

  $\boxed{C(x') = d/dx\, E[Y=y \mid do(X=x)]|_{x=x'}}$

  "change in expected value of Y, if there is an intervention in variable X"

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

$P(Y=y \mid X=x) \neq P(Y=y \mid do(X=x))$

do-calculus with known causal structure

Pick a random day:

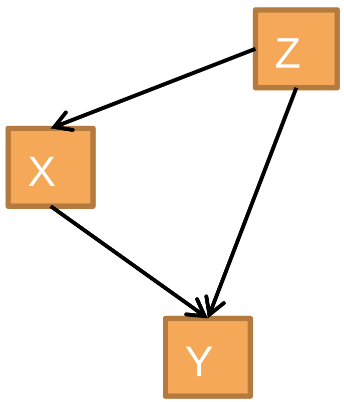
P(rain | wet) = high

P(rain | do(wet)) =
= P(rain) =
= low

Pearl's do-calculus

do-calculus with known causal structure

Causal structure



+

Rules:
Expression with "do"

Expression without "do"

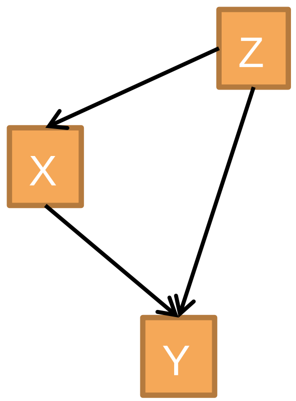
Judea Pearl, "*Causality*", 2010, **Cambridge University Press**

do-calculus with known causal structure

Example: Back-door Adjustment

Assume Z is binary (0/1)

Causal structure



+ **Rules** ⇒

P(Y=y | do(X=x))

=

P(Y=y | X=x, Z=0) * P(Z=0) +

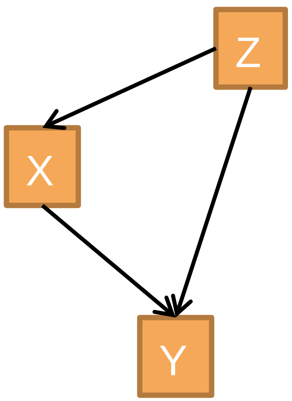
P(Y=y | X=x, Z=1) * P(Z=1)

Example: Back-door Adjustment

do-calculus with known causal structure

Assume Z is binary (0/1)

Causal structure



\+ **Rules** →

“do”

P(Y=y | do(X=x))

=

P(Y=y | X=x, Z=0) * P(Z=0) +

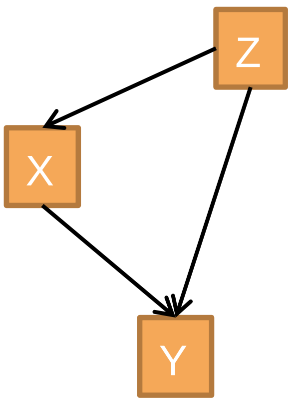
P(Y=y | X=x, Z=1) * P(Z=1)

Example: Back-door Adjustment

do-calculus with known causal structure

Assume Z is binary (0/1)

Causal structure



\+ **Rules** →

P(Y=y | do(X=x))

=

P(Y=y | X=x, Z=0) * P(Z=0) +
P(Y=y | X=x, Z=1) * P(Z=1)

No "do"

Conclusion 1

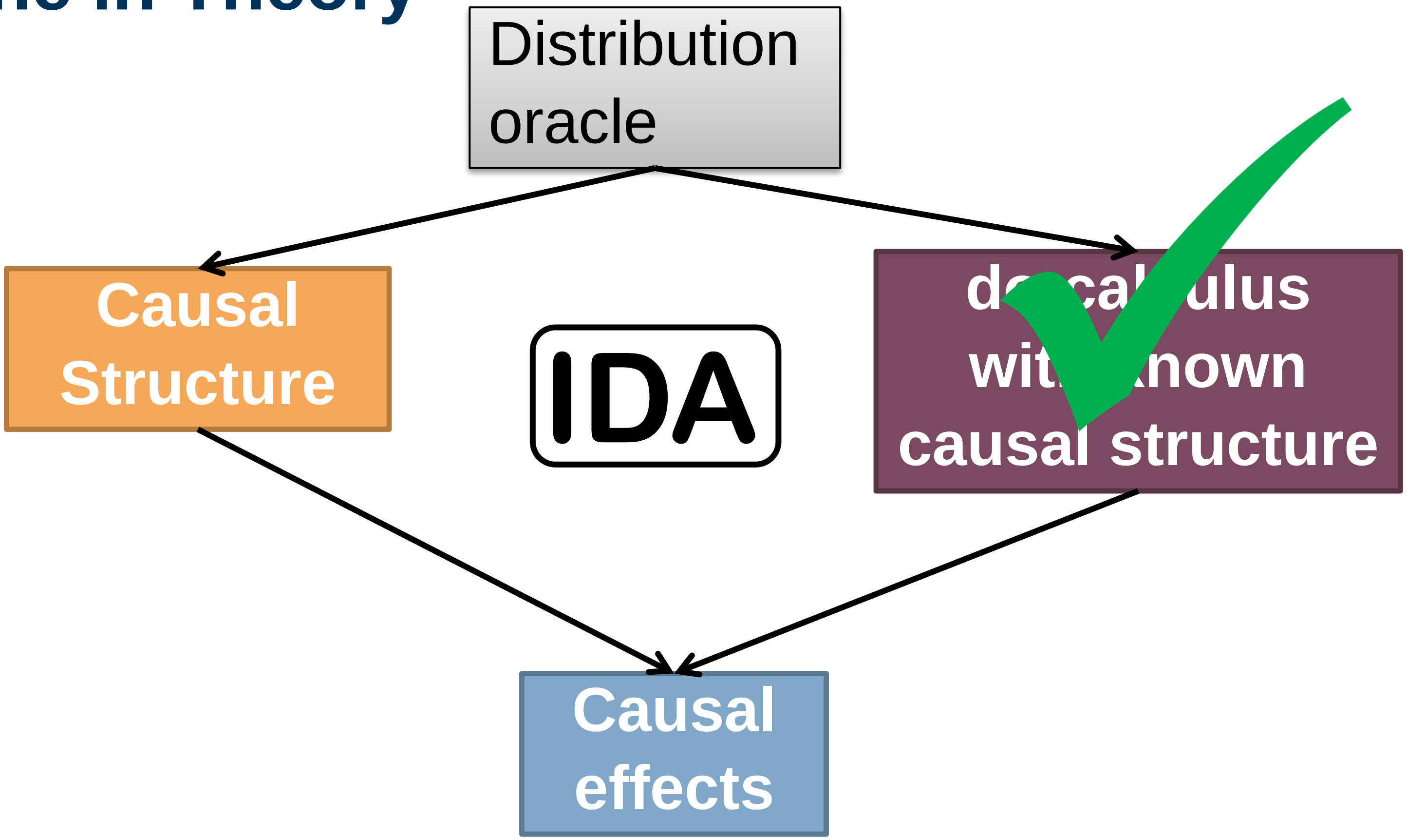
do-calculus
with known
causal structure

If causal structure is known,

we can infer causal effects

from observations

Outline in Theory



Estimate Causal Structure

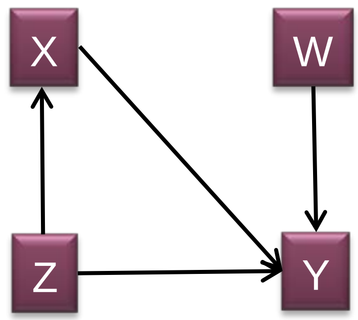
Oftentimes, causal structure is unknown

↓

Estimate causal structure

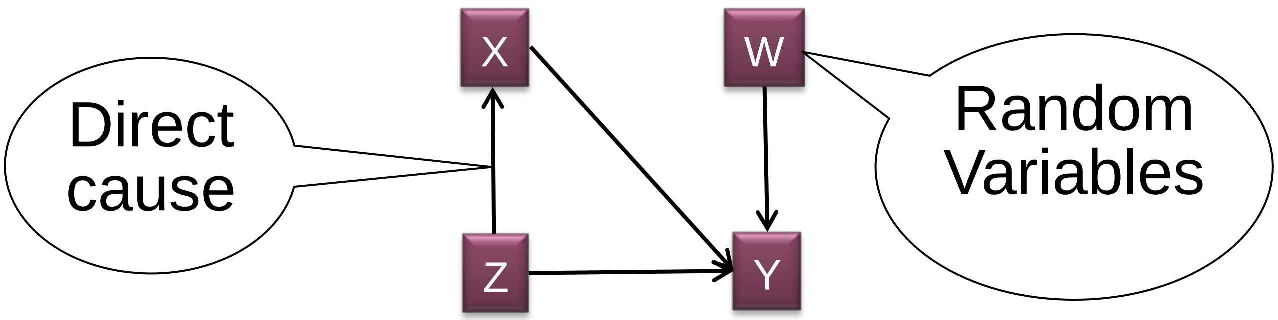
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Causal Directed Acyclic Graph (DAG)

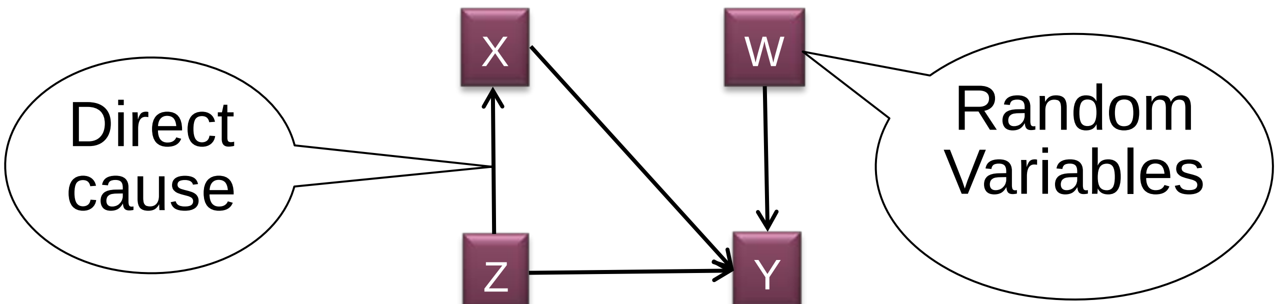


Causal Structure

Causal Directed Acyclic Graph (DAG)



Causal Directed Acyclic Graph (DAG)

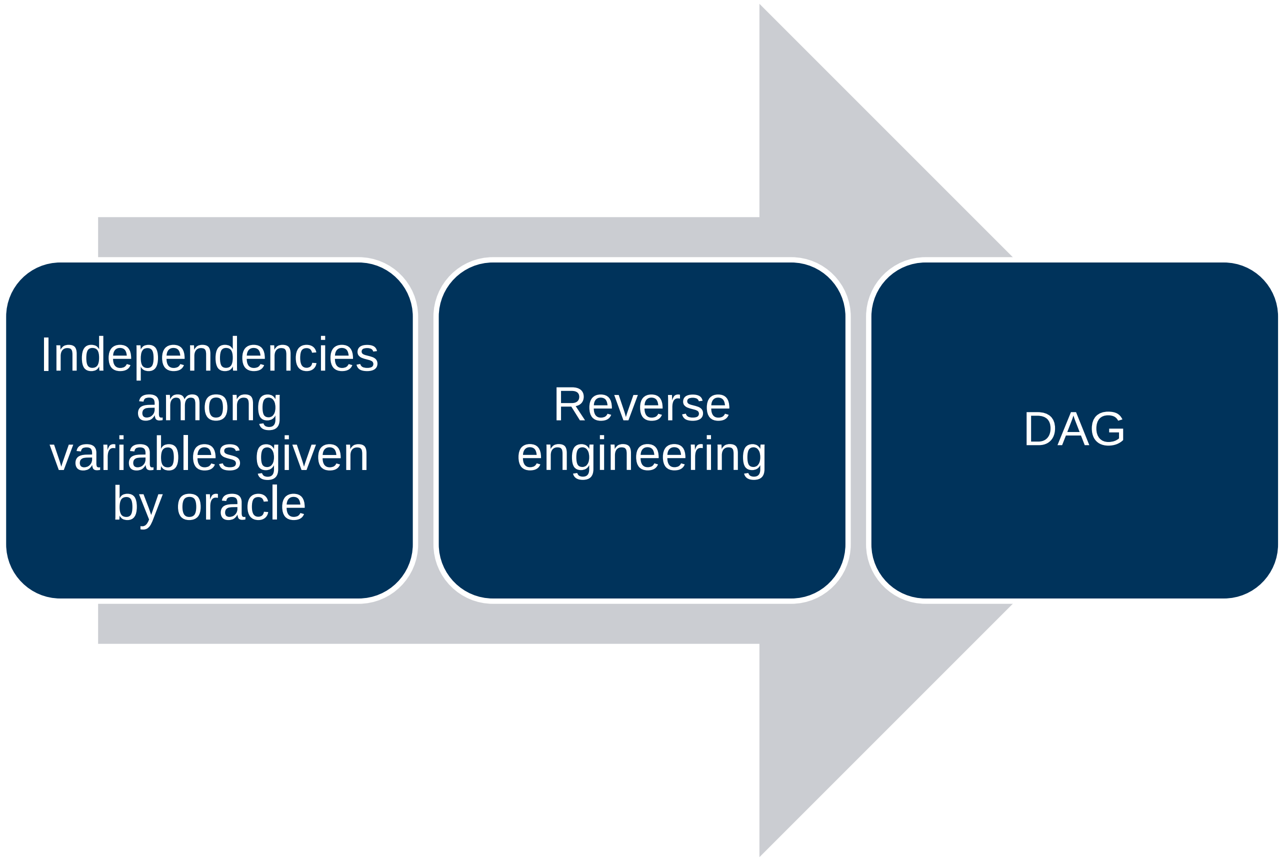


implies

Conditional independence relations among variables

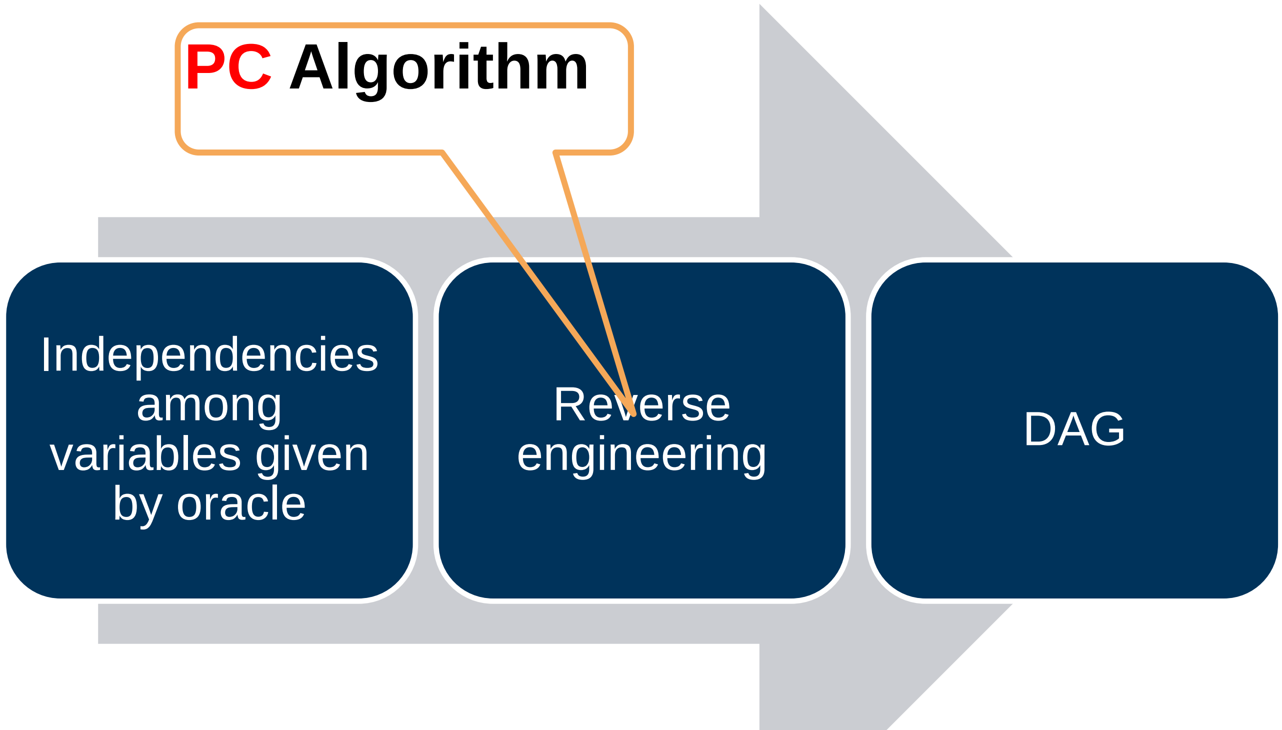
Estimate a DAG model

DAG encodes independence information



Estimate a DAG model

DAG encodes independence information



P. Spirtes, C. Glymour, R. Scheines, "*Causation, Prediction, and Search*", 2000, **MIT Press**

Ambiguity: Equivalence class

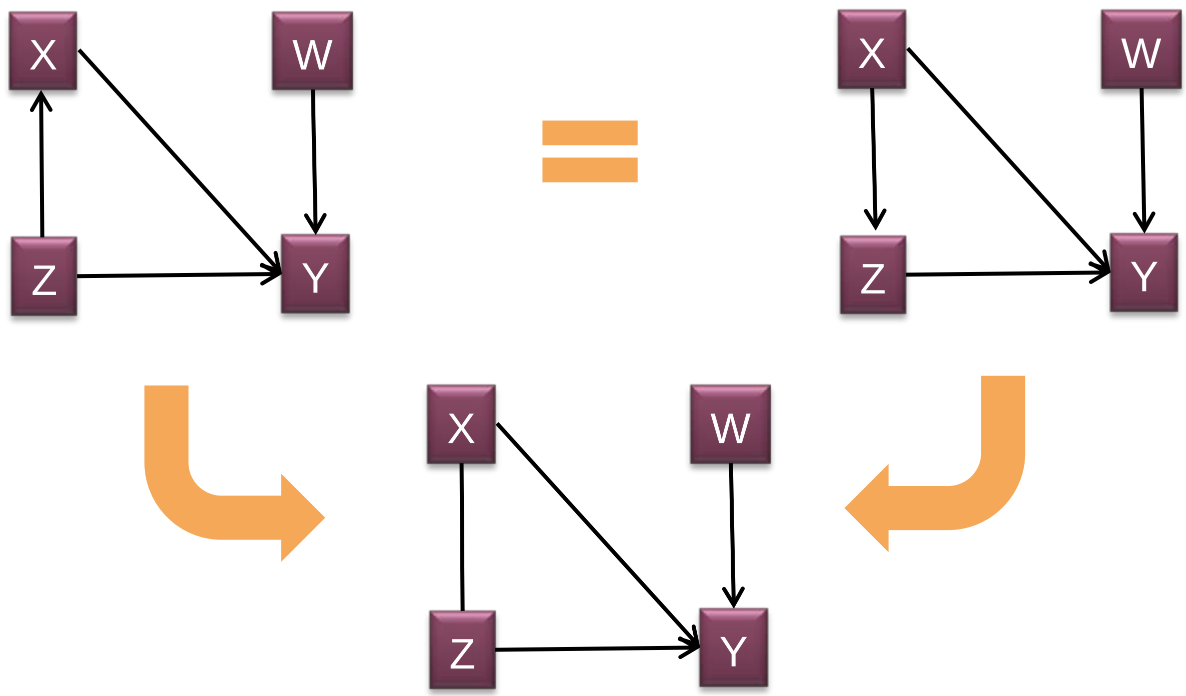
Several DAGs describe exactly the same list of independence relations

Ambiguity: Equivalence class

Several DAGs describe exactly the same list of independence relations

Ambiguity: Equivalence class

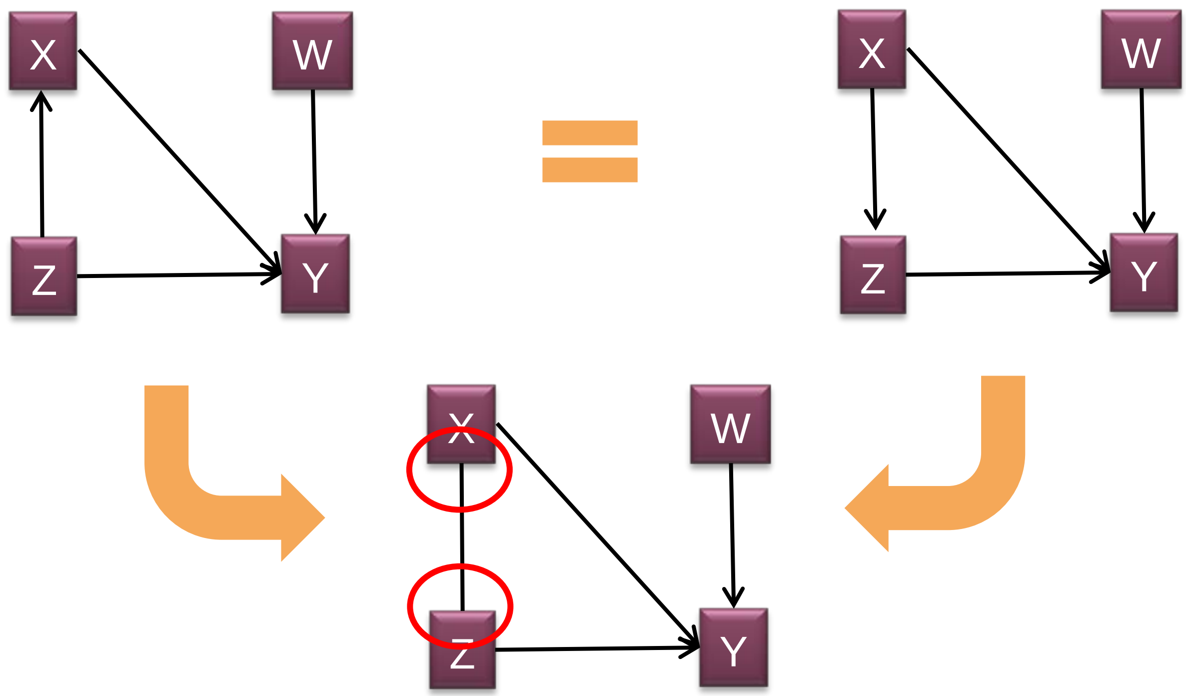
Several DAGs describe exactly the same list of independence relations



Equivalence class: PARTIALLY Directed Acyclic Graph (PDAG)

Ambiguity: Equivalence class

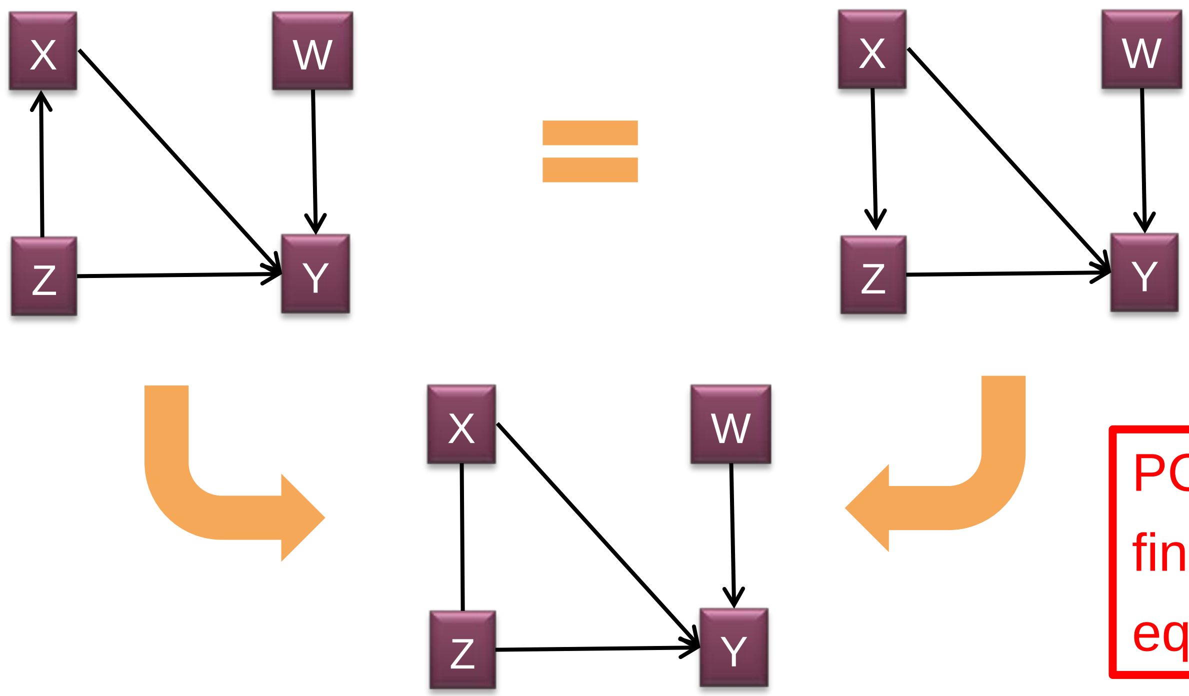
Several DAGs describe exactly the same list of independence relations



Equivalence class: PARTIALLY Directed Acyclic Graph (PDAG)

Ambiguity: Equivalence class

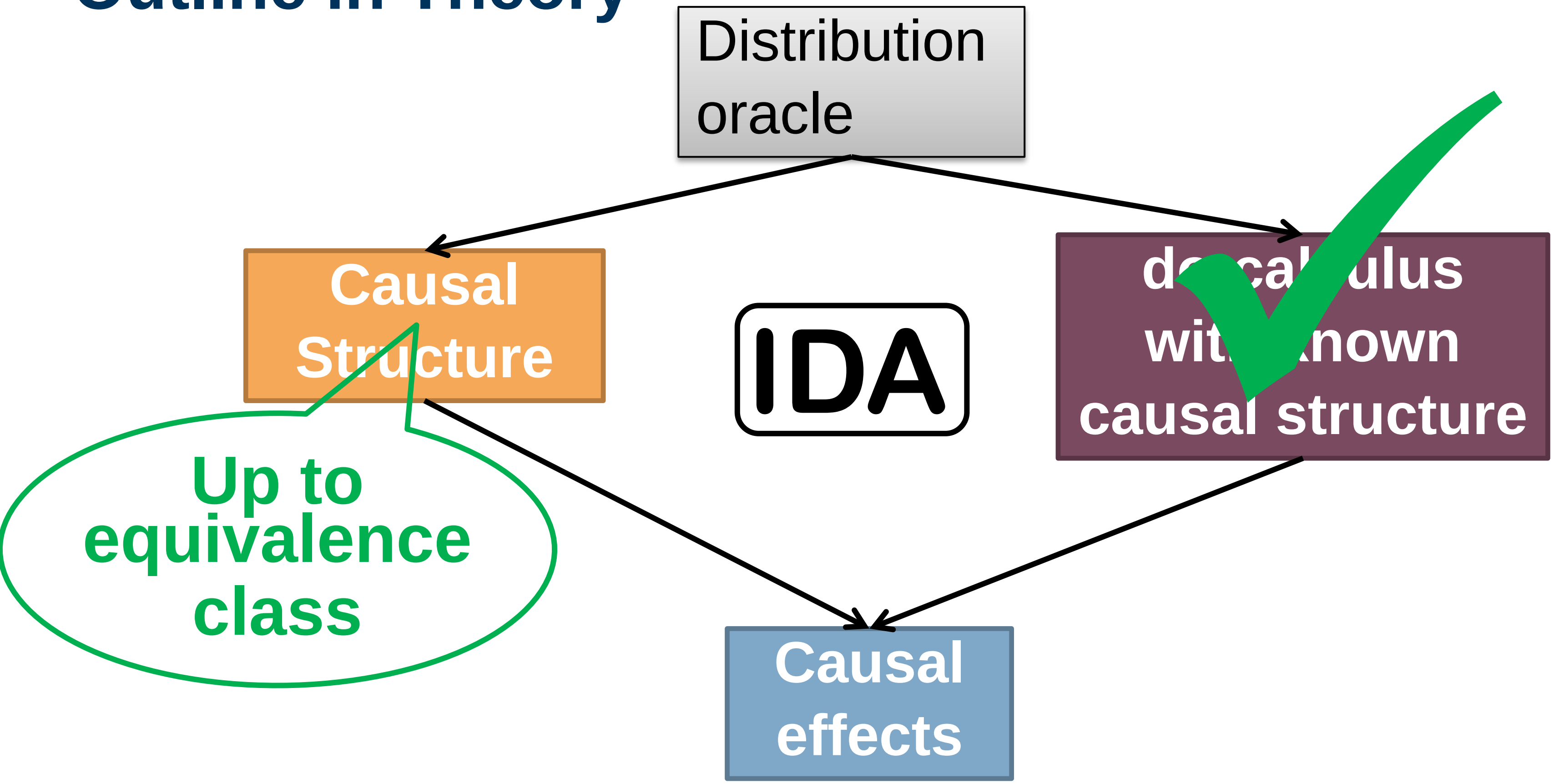
Some DAGs describe exactly the same list of independence relations



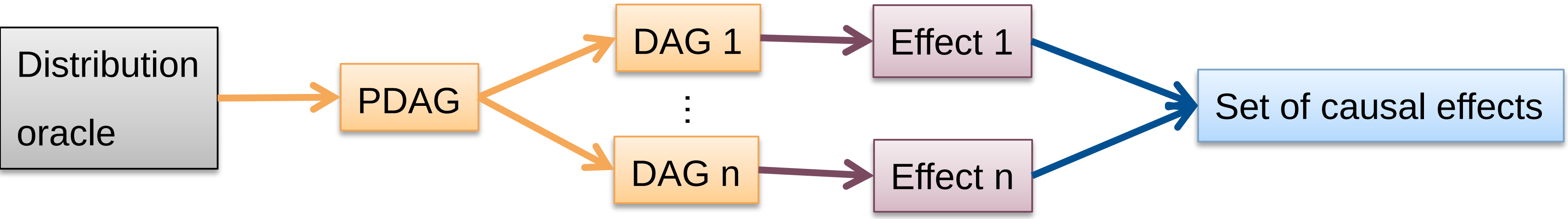
PC Algorithm finds equivalence class

Equivalence class: PARTIALLY Directed Acyclic Graph (PDAG)

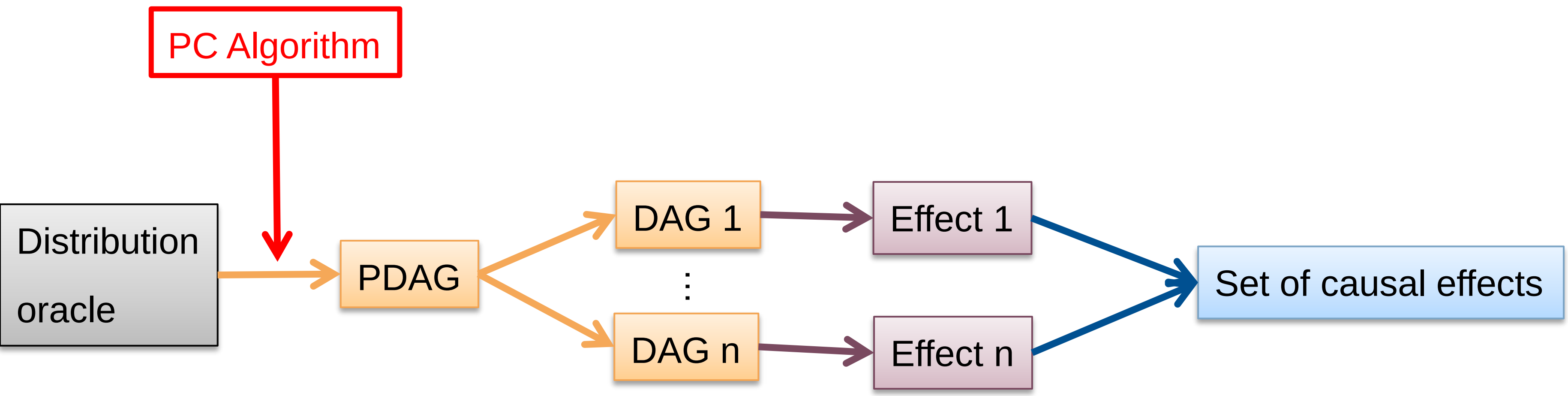
Outline in Theory



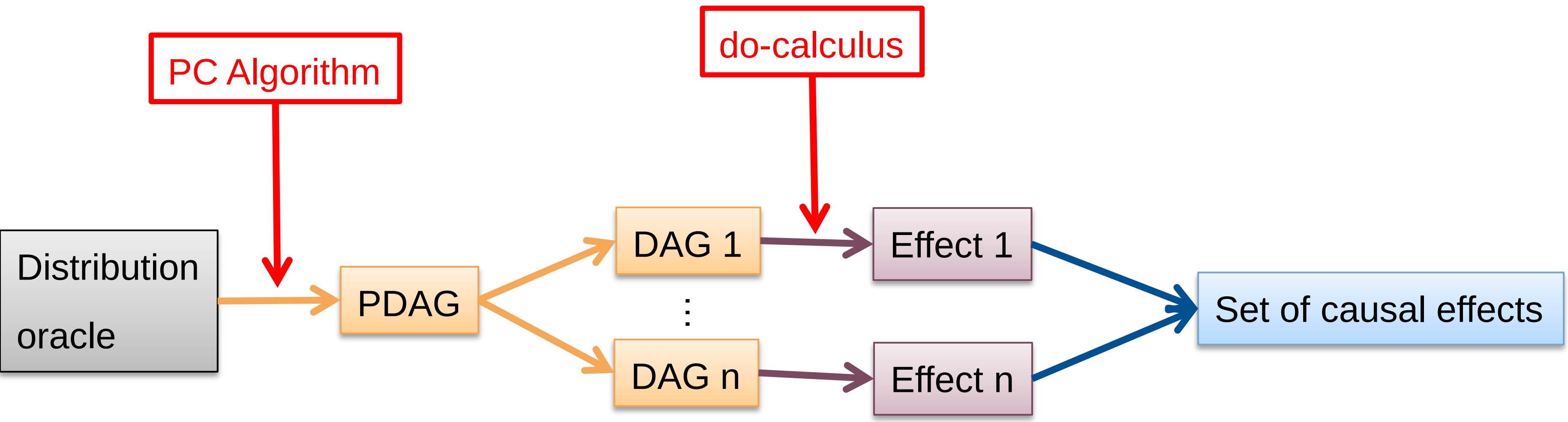
Putting everything together



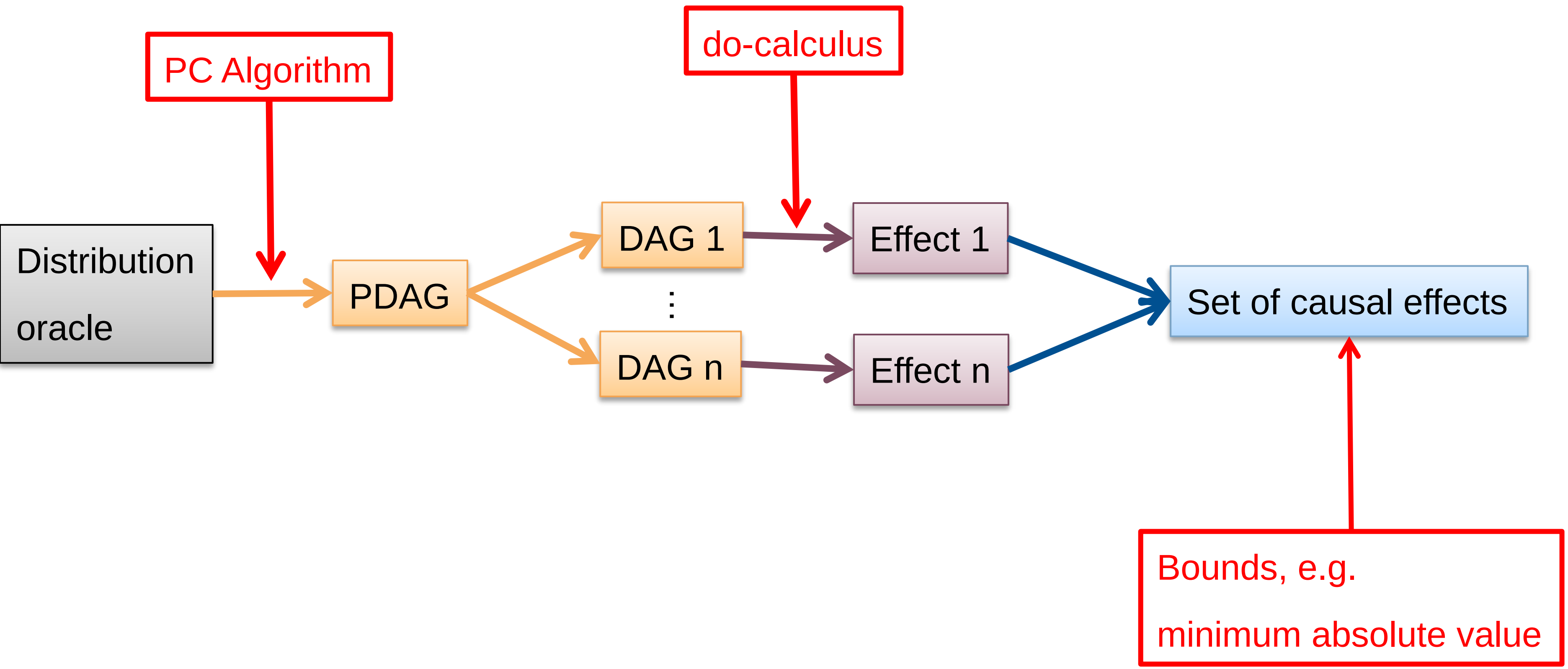
Putting everything together



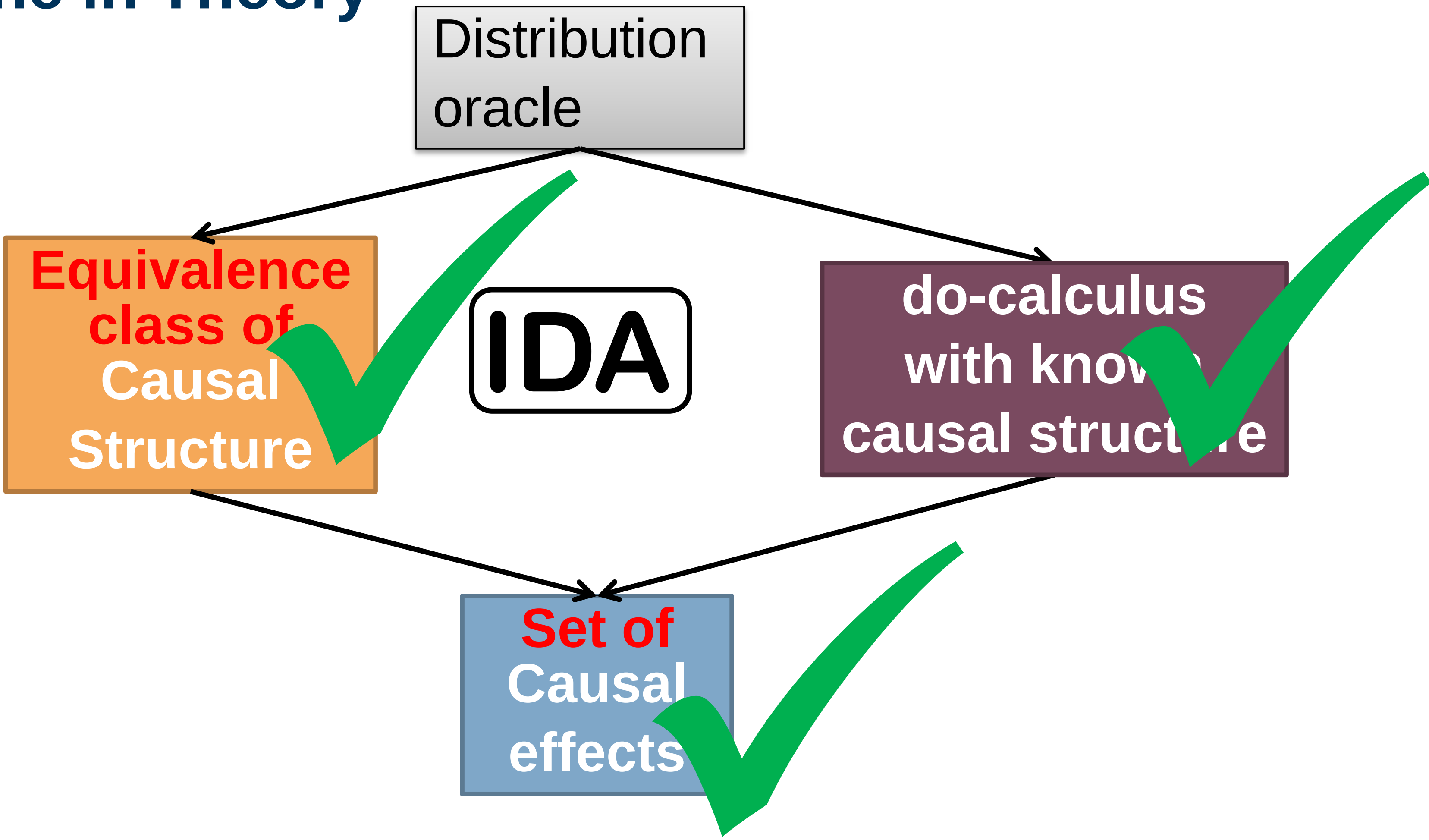
Putting everything together



Putting everything together

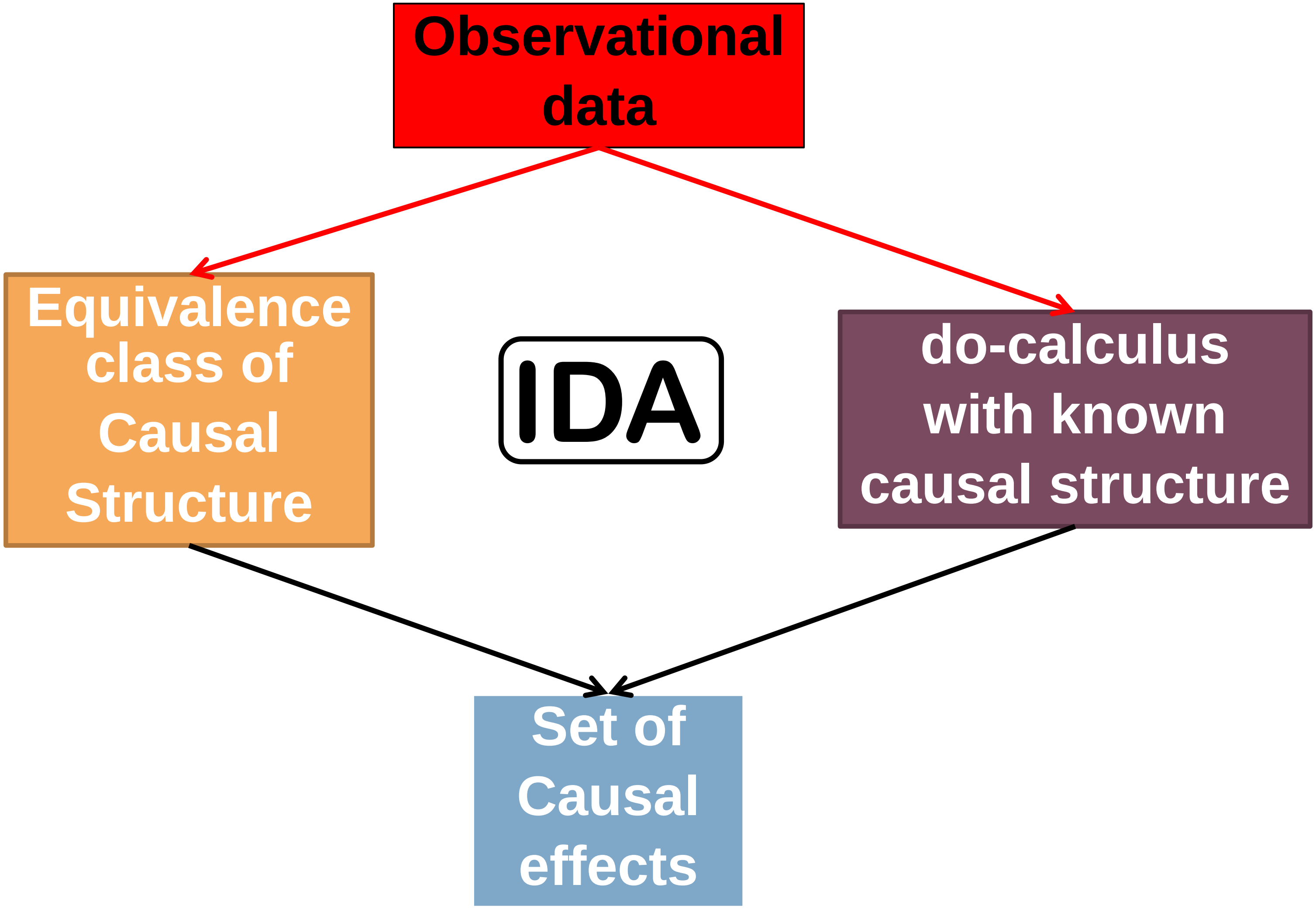


Outline in Theory

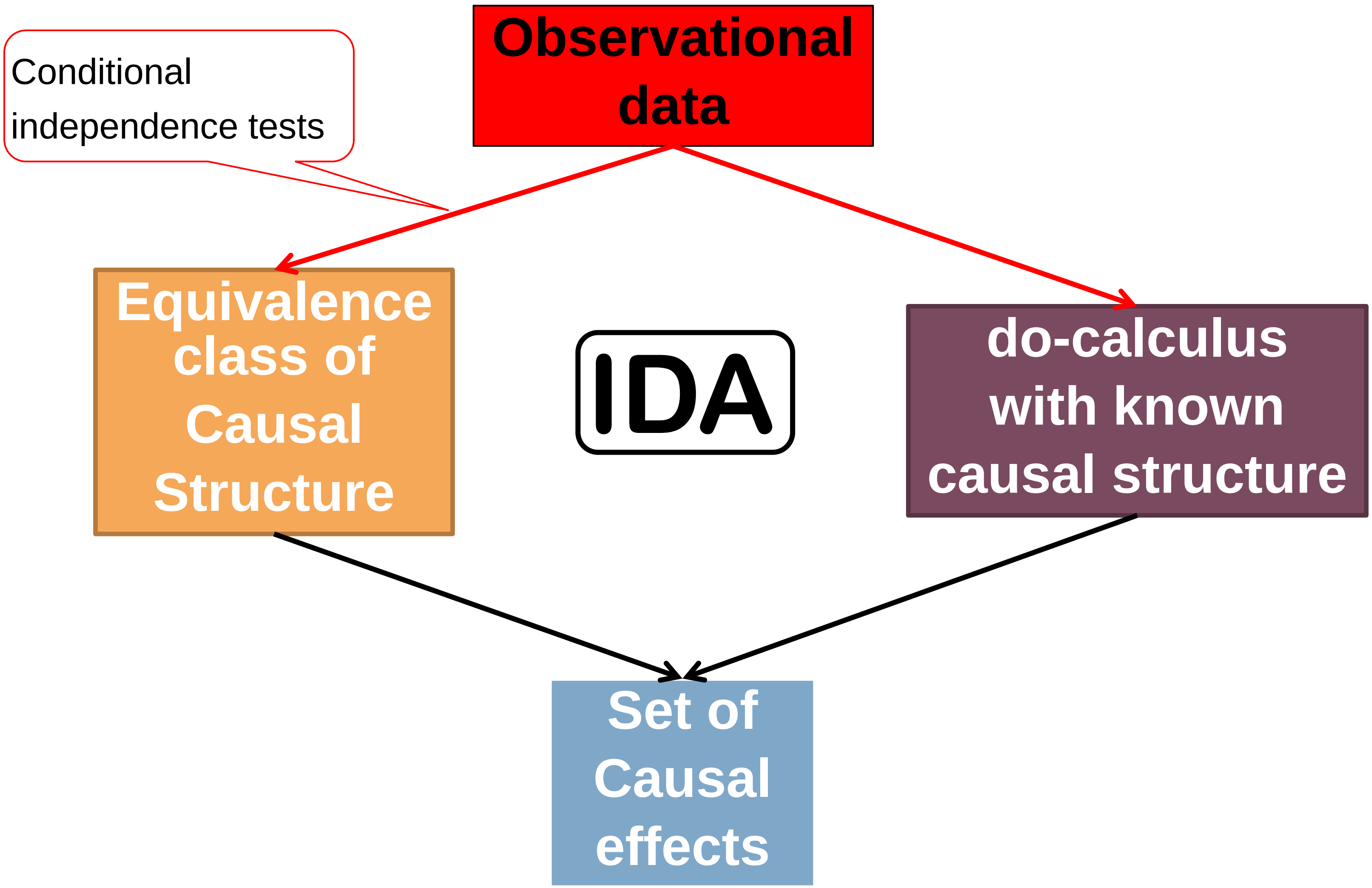



I’m busy!
Find your own information on the distribution…
horacek

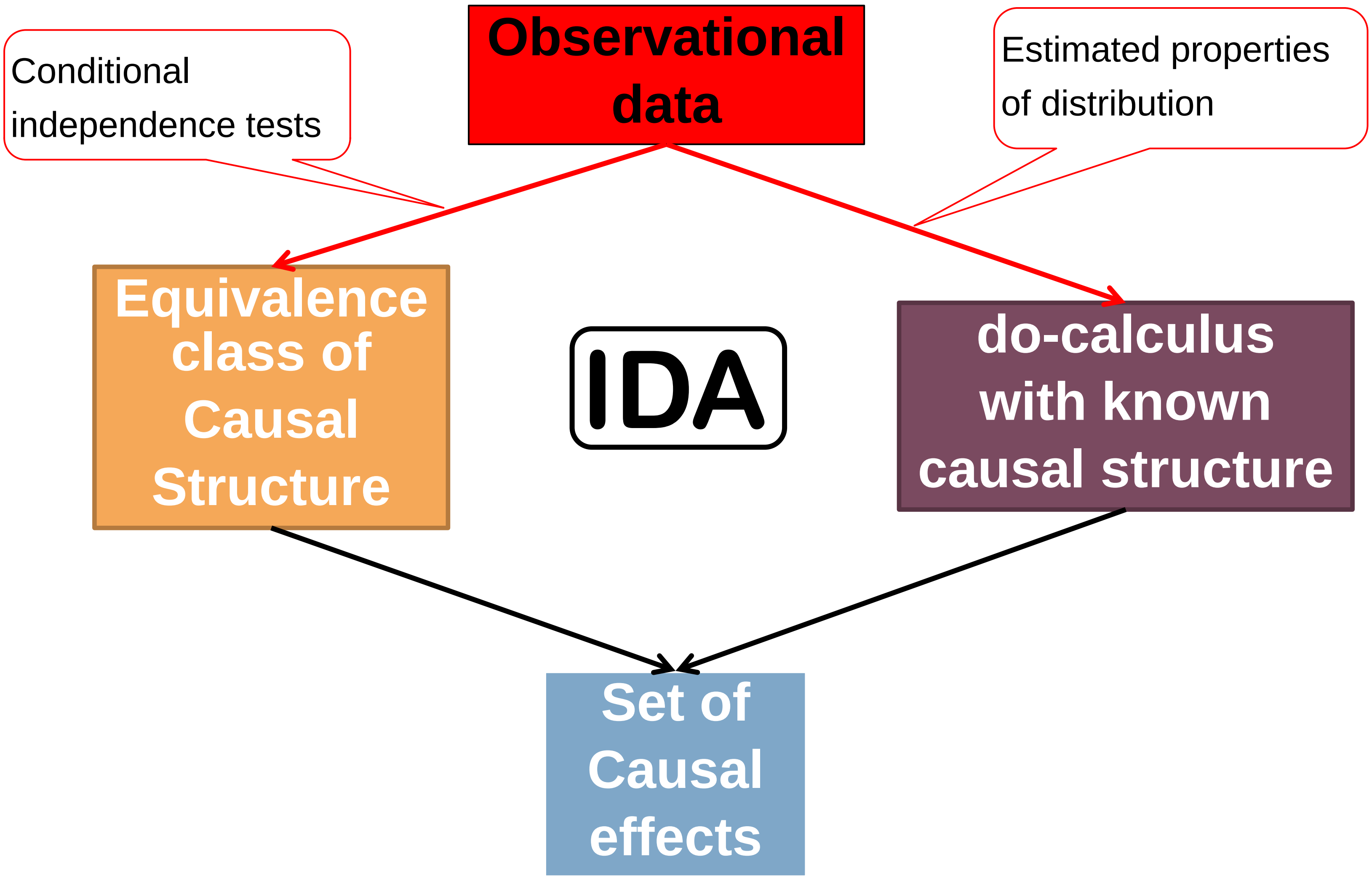
Outline in ~~Theory~~ Practice



Outline in ~~Theory~~ Practice

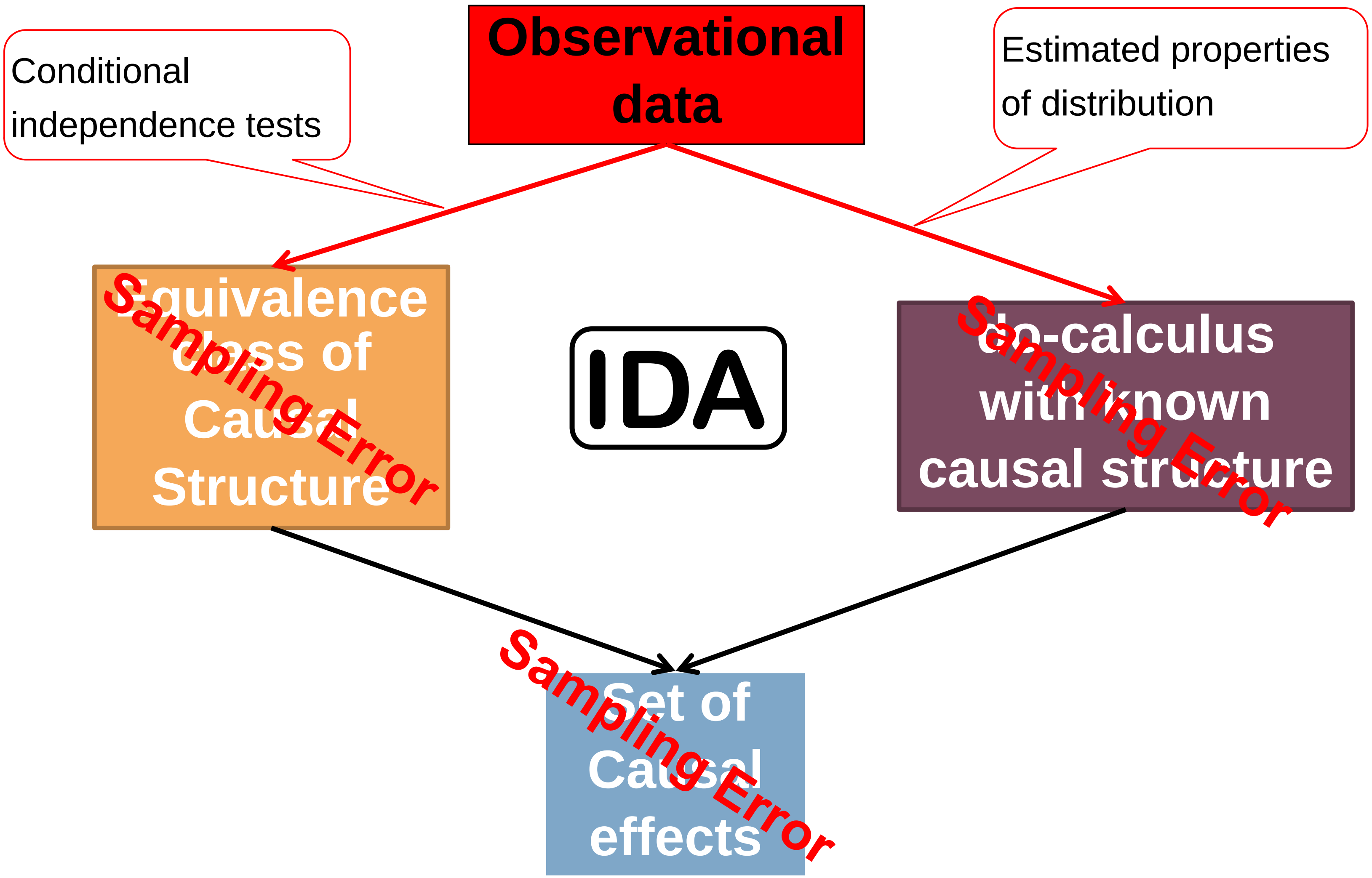


Outline in ~~Theory~~ Practice



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Outline in ~~Theory~~ Practice



Consistency in high-dimensions: Gaussian case

Estimating graphical models with PC algorithm

M. Kalisch, P. Bühlmann, "*Estimating high-dimensional DAGs with the PC algorithm*", 2007, **JMLR** 8, 613 - 636

Do-calculus in high dimensions

M.H. Maathuis, M. Kalisch, P. Bühlmann, "*Estimating high-dimensional intervention effects from observational data*", 2009, **Annals of Statistics** 37, 3133 - 3164

Consistency in high-dimensions: Gaussian case

Estimating graphical models with PC algorithm

M. Kalisch, P. Bühlmann, "*Estimating high-dimensional DAGs with the PC algorithm*", 2007, **JMLR** 8, 613 - 636

Do-calculus in high dimensions

M.H. Maathuis, M. Kalisch, P. Bühlmann, "*Estimating high-dimensional intervention effects from observational data*", 2009, **Annals of Statistics** 37, 3133 - 3164

IDA

Intervention effects if

DAG is

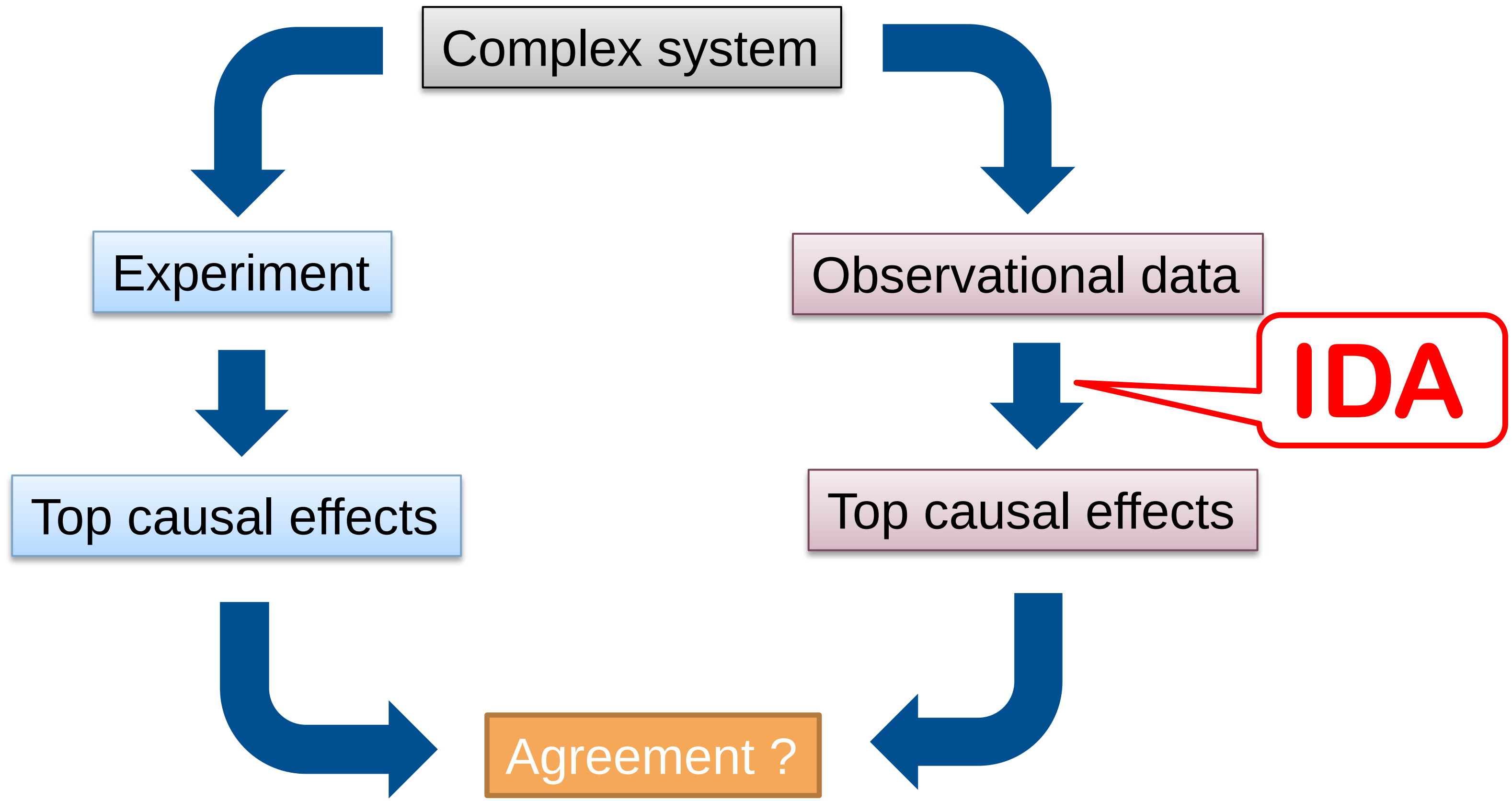
Absent

Main assumptions & requirements

- Involves a tuning parameter

- Gaussian data from unknown causal DAG
- Faithfulness to this DAG
- No hidden or selection variables

Experimental validation



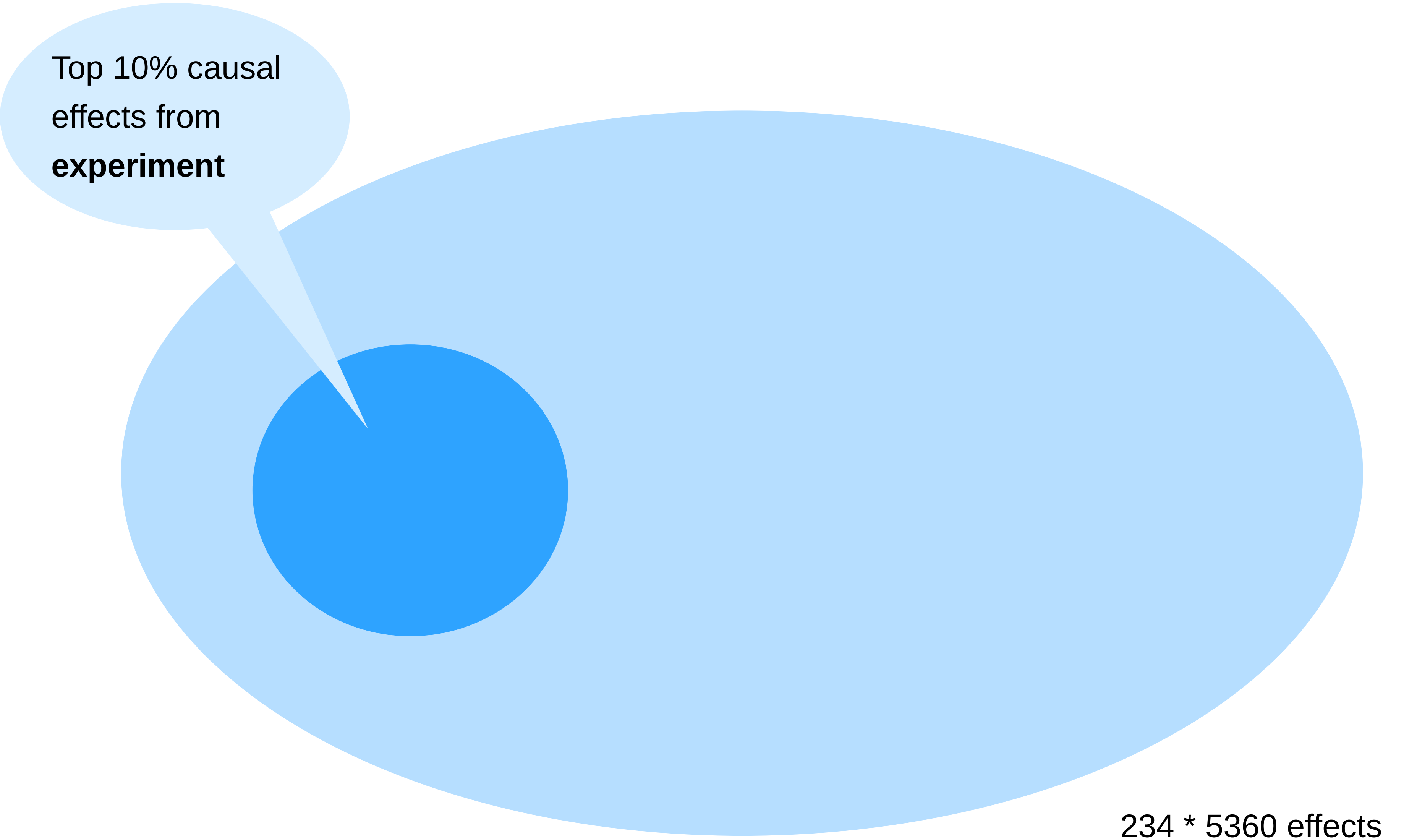
Back to the beer:

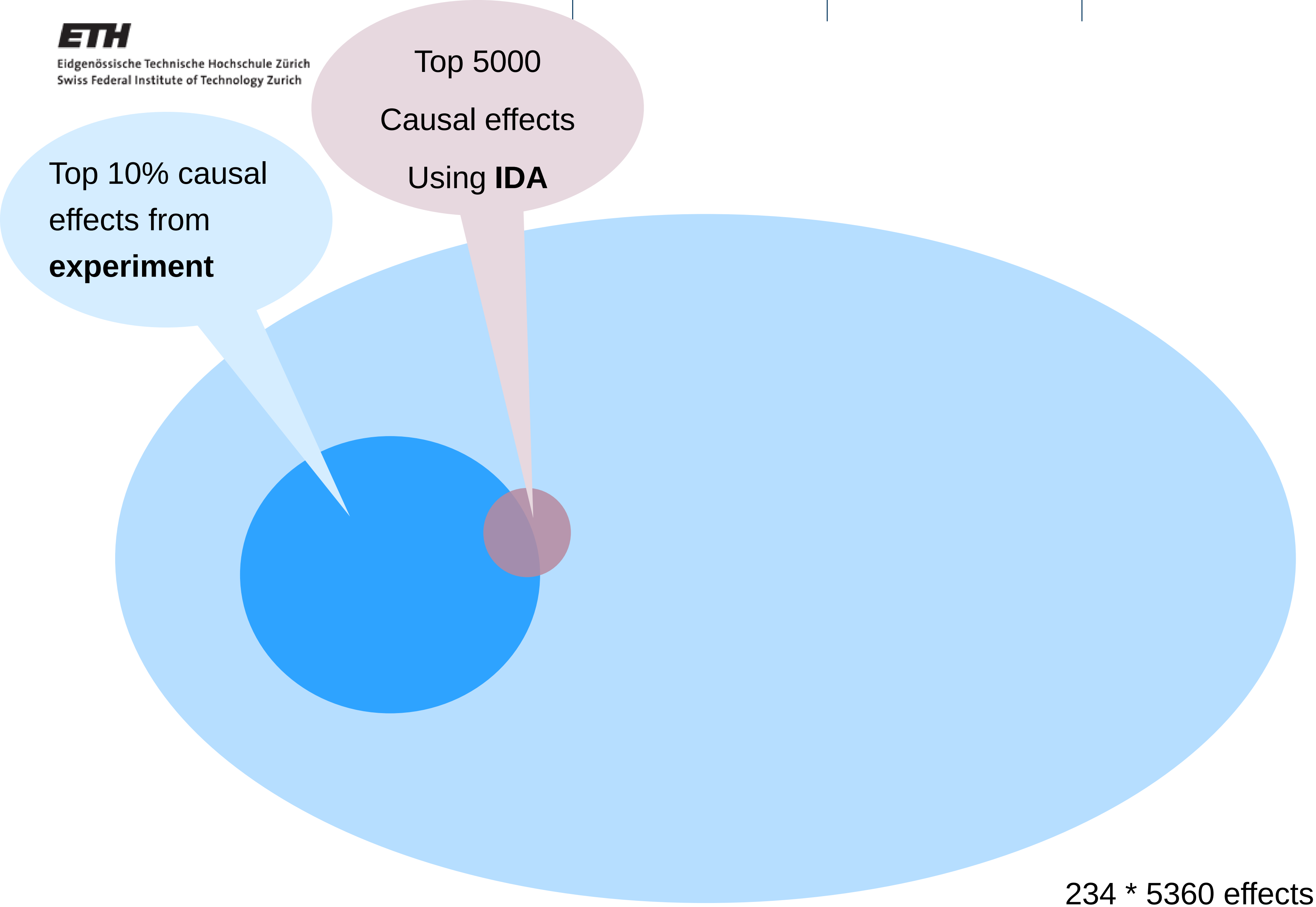
Experimental validation of IDA in *Saccharomyces cerevisiae*

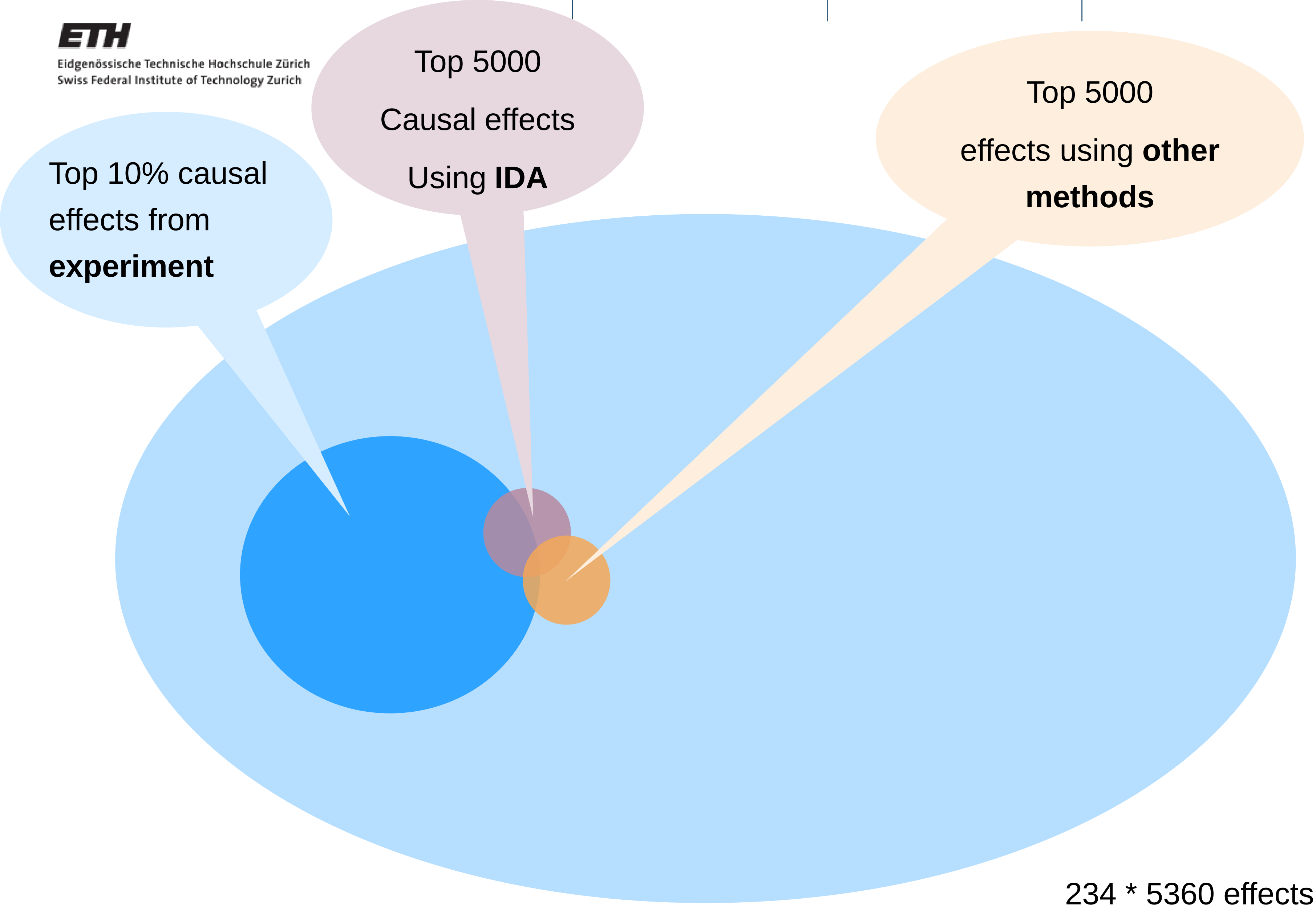
Setting

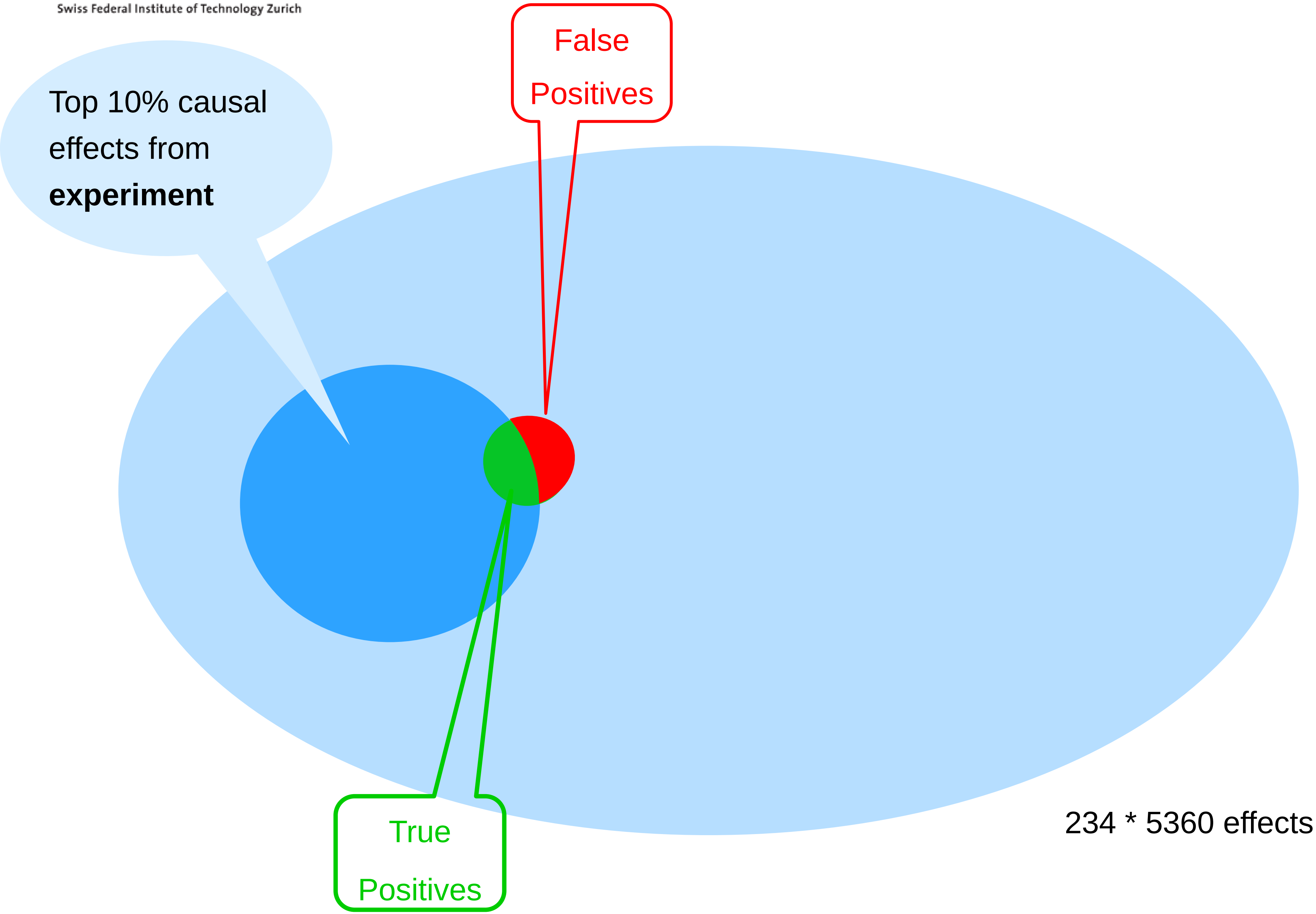
- 5361 observed genes
- Experiments: 234 single-gene deletion mutants
- Observational data: 63 wild-type cultures
- Very high dimensional: 5361 variables, 63 observations

234 * 5360 effects


Top 10% causal effects from **experiment**
234 * 5360 effects


Top 5000
Causal effects
Using **IDA**
Top 10% causal effects from **experiment**
234 * 5360 effects


Top 10% causal effects from **experiment**
Top 5000 Causal effects Using **IDA**
Top 5000 effects using **other methods**
234 * 5360 effects


Top 10% causal effects from **experiment**
False Positives
True Positives
234 * 5360 effects

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

M.H. Maathuis,

D. Colombo,

M. Kalisch,

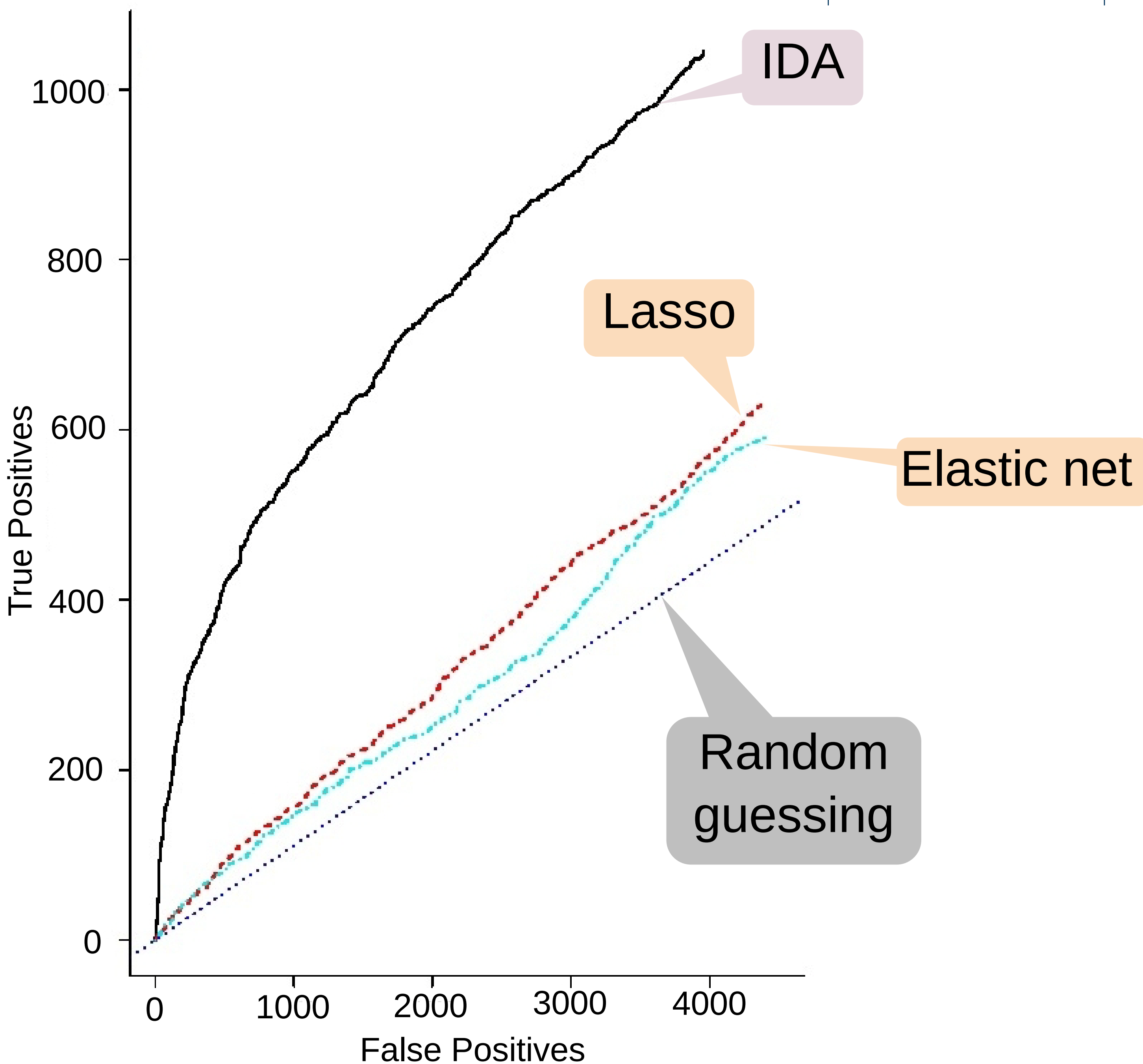
P. Bühlmann,

"*Predicting causal effects in large-scale systems from observational data*",

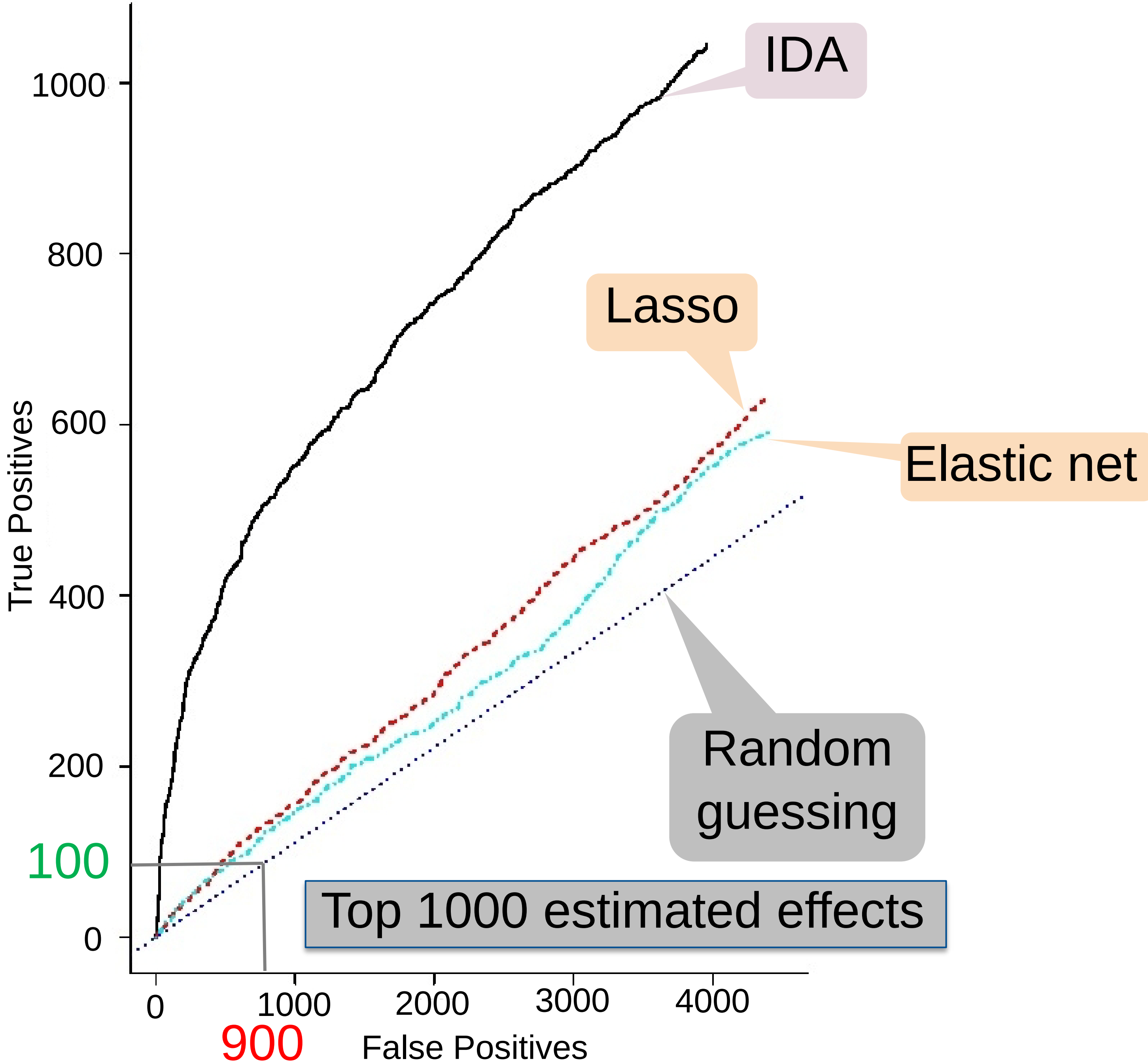
2010,

Nature Methods

7, 247 - 248



M.H. Maathuis,
D. Colombo,
M. Kalisch,
P. Bühlmann,
"*Predicting causal effects in large-scale systems from observational data*",
2010,
Nature Methods
7, 247 - 248



M.H. Maathuis,

D. Colombo,

M. Kalisch,

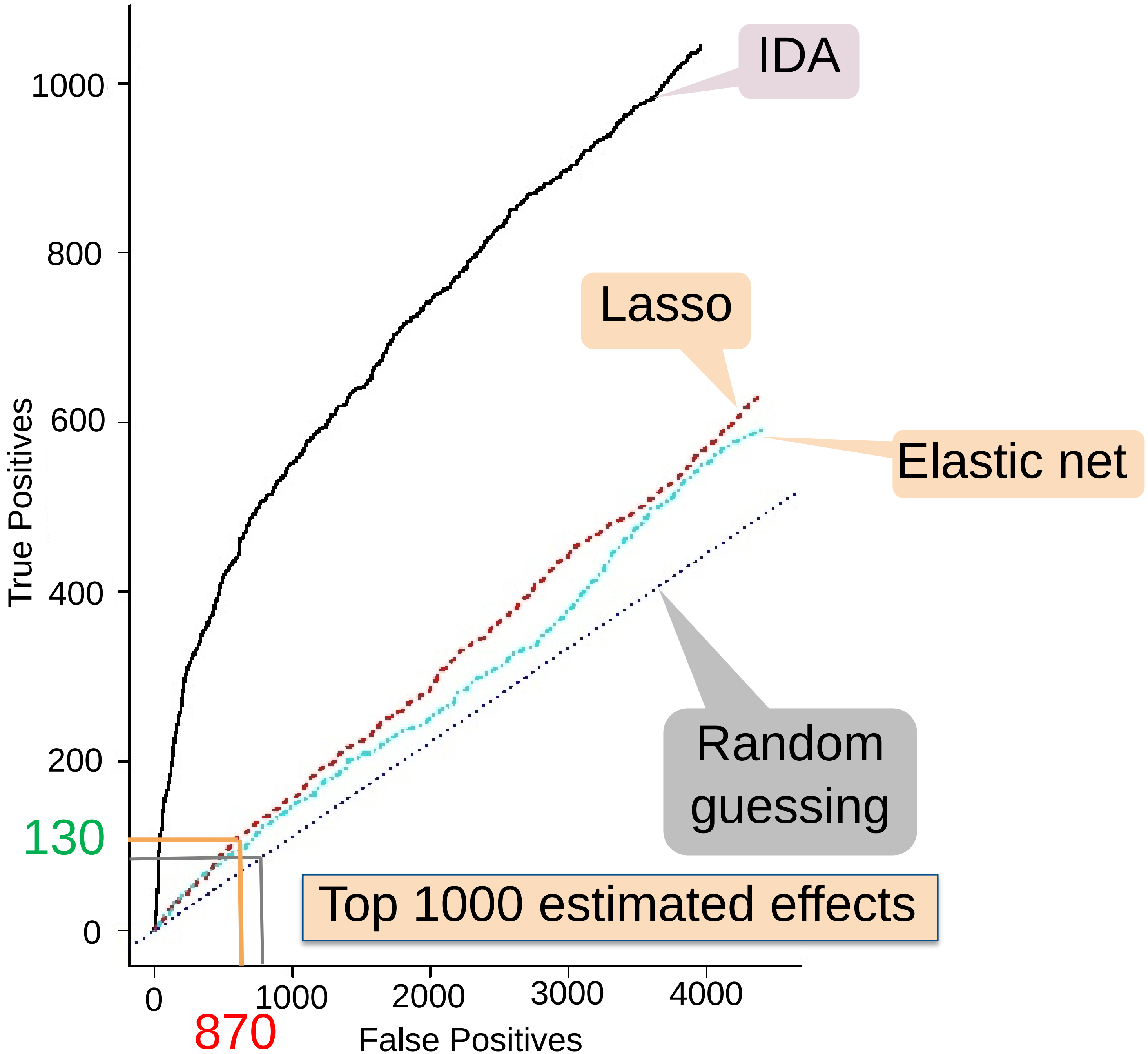
P. Bühlmann,

“*Predicting causal effects in large-scale systems from observational data*”,

2010,

Nature Methods

7, 247 - 248



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

M.H. Maathuis,

D. Colombo,

M. Kalisch,

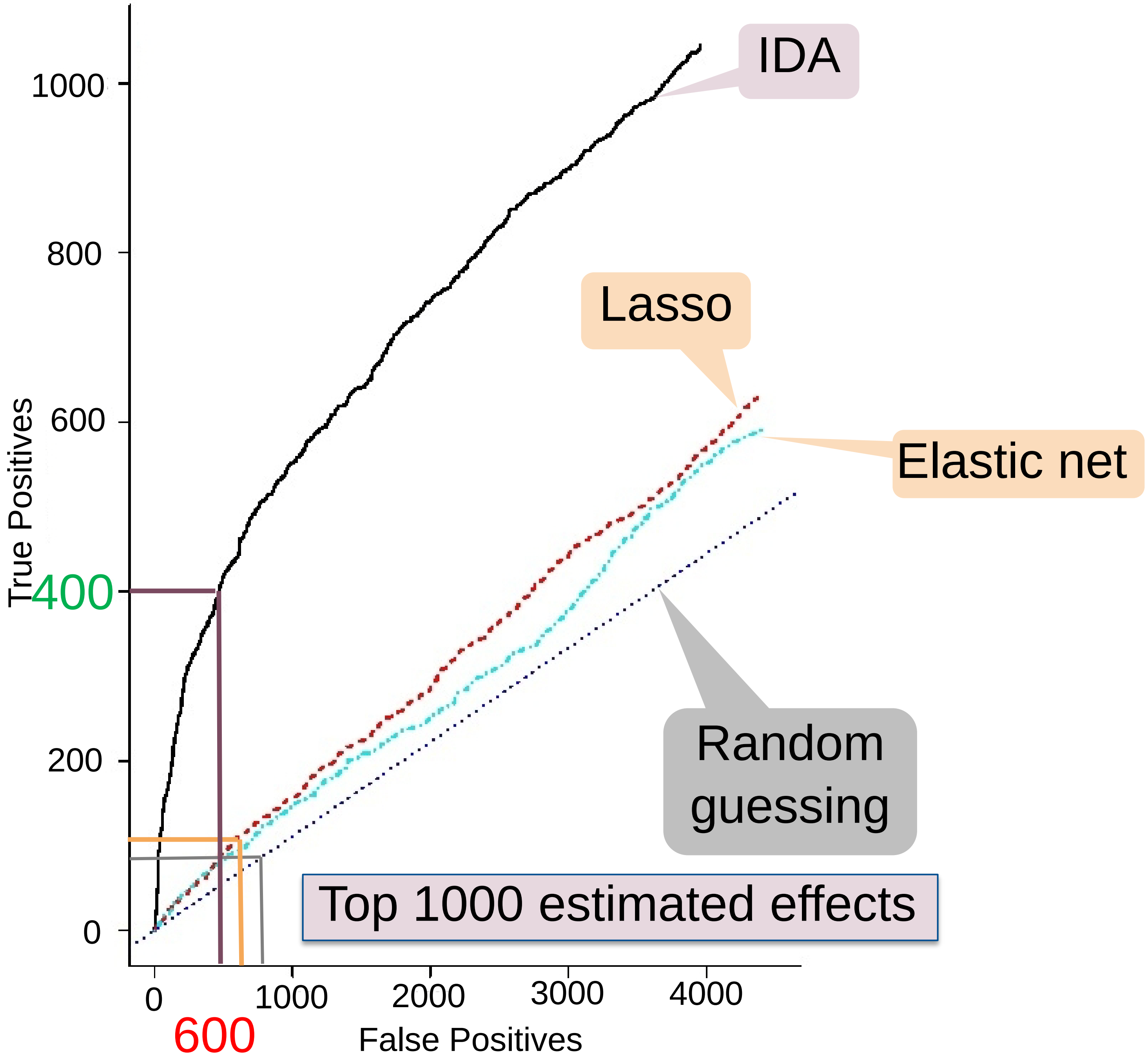
P. Bühlmann,

"*Predicting causal effects in large-scale systems from observational data*",

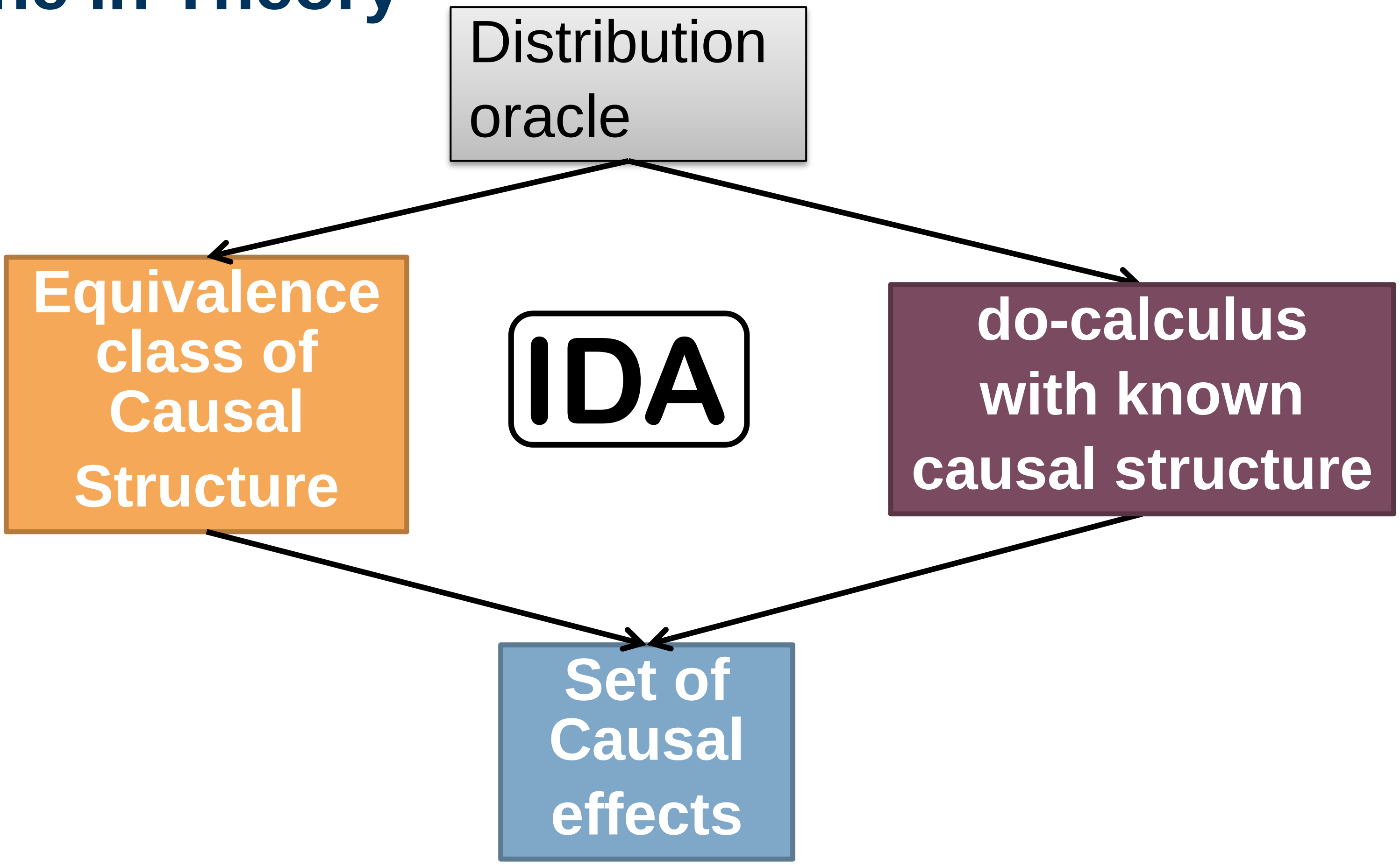
2010,

Nature Methods

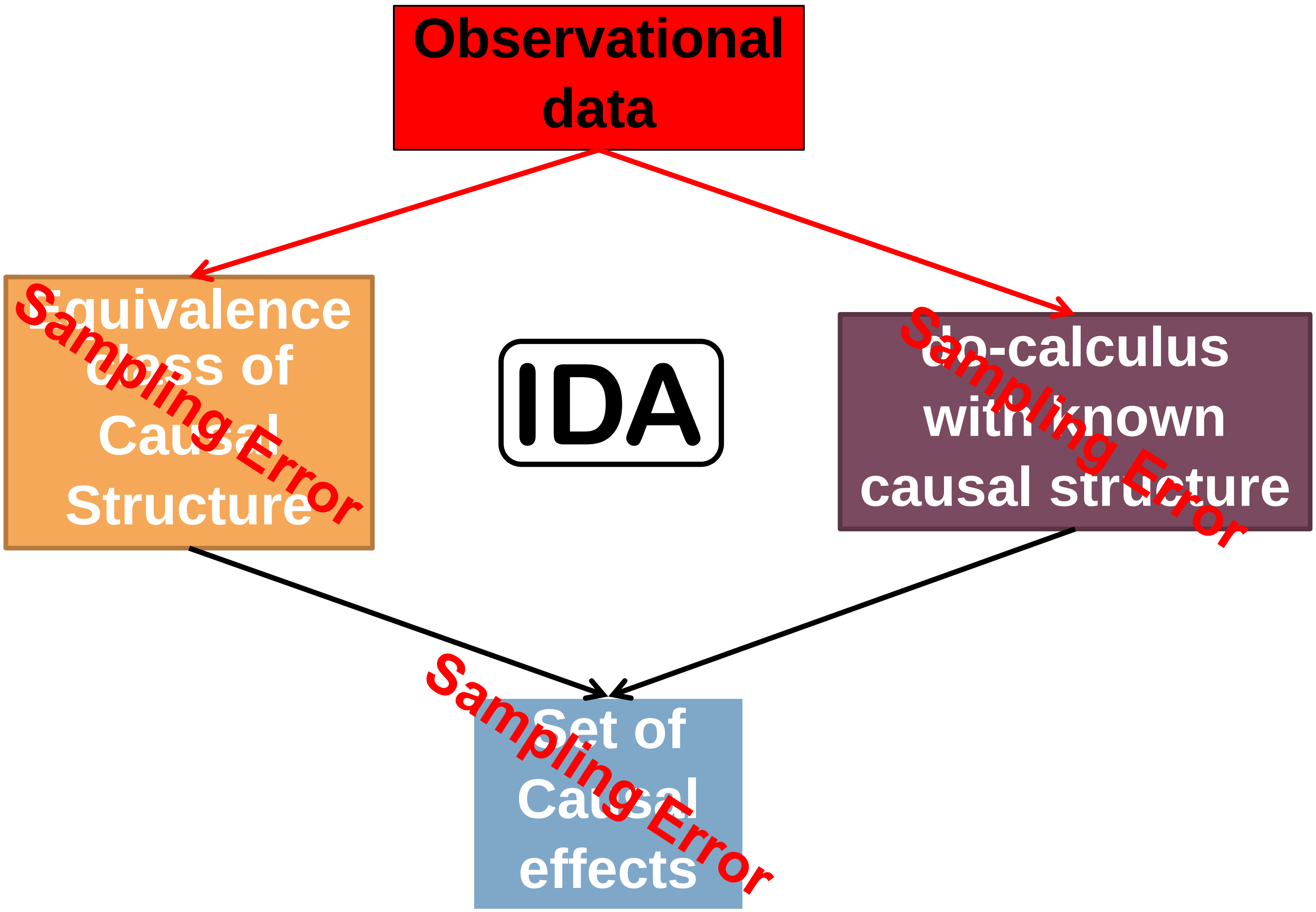
7, 247 - 248



Outline in Theory



Outline in ~~Theory~~ Practice



Summary of assumptions

- Data is faithful to an underlying causal DAG
- No hidden or selection variables
- Consistent in high-dimensions if
  - data multivariate normal
  - some regularity conditions on partial correlations
  - underlying DAG is sparse
- For IDA also: All conditional expectations are linear

Conclusions

- Theoretical result: Under certain assumptions

  IDA is consistent, even in high dimensions
- Causal effect **not identified uniquely**

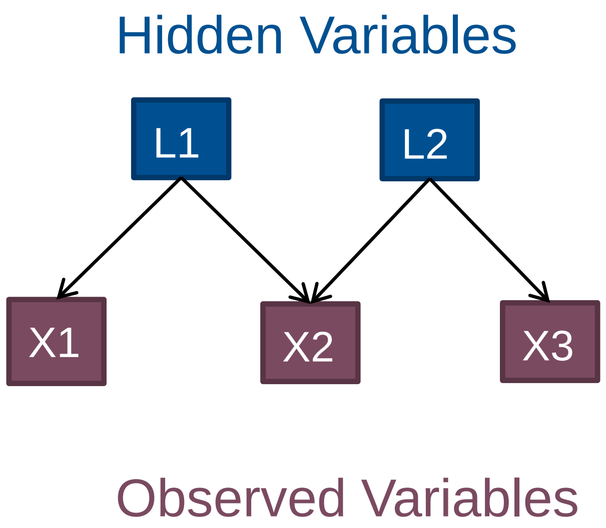
  (even without sampling error)
- **Validation on experimental data**
- IDA cannot replace randomized, controlled experiments
- IDA can help

  prioritizing and designing random experiments
- Software is available for free: **R-package “pcalg”**

Current research

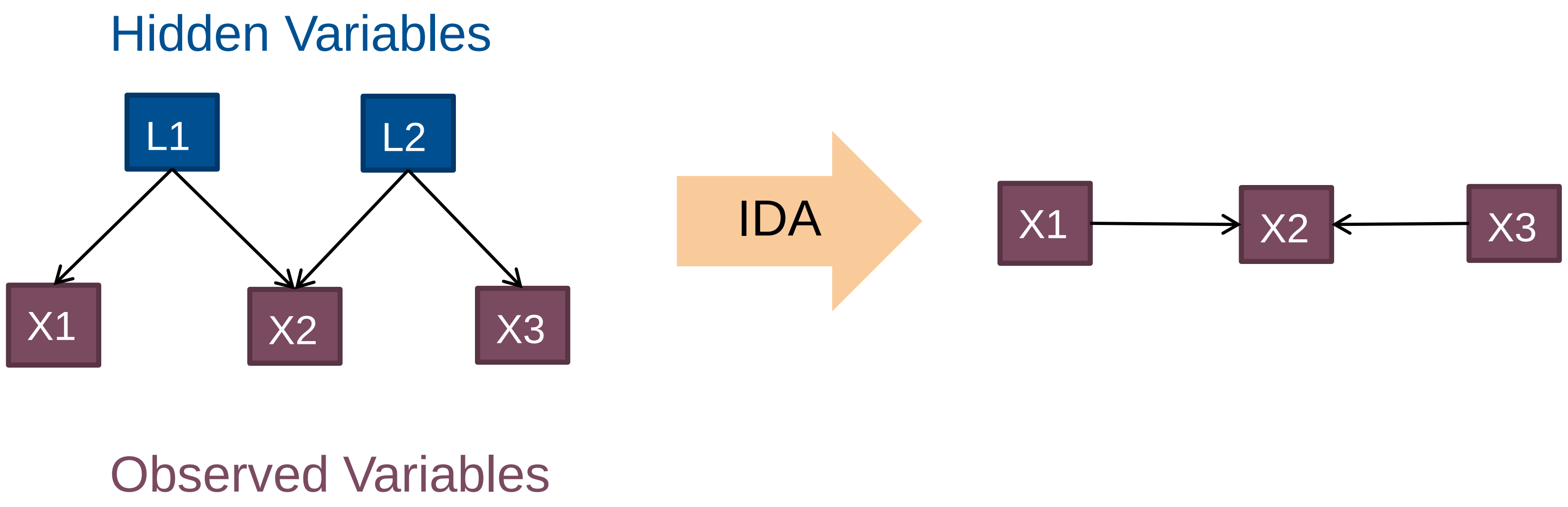
What if hidden and/or selection variables are present?

Hidden variables



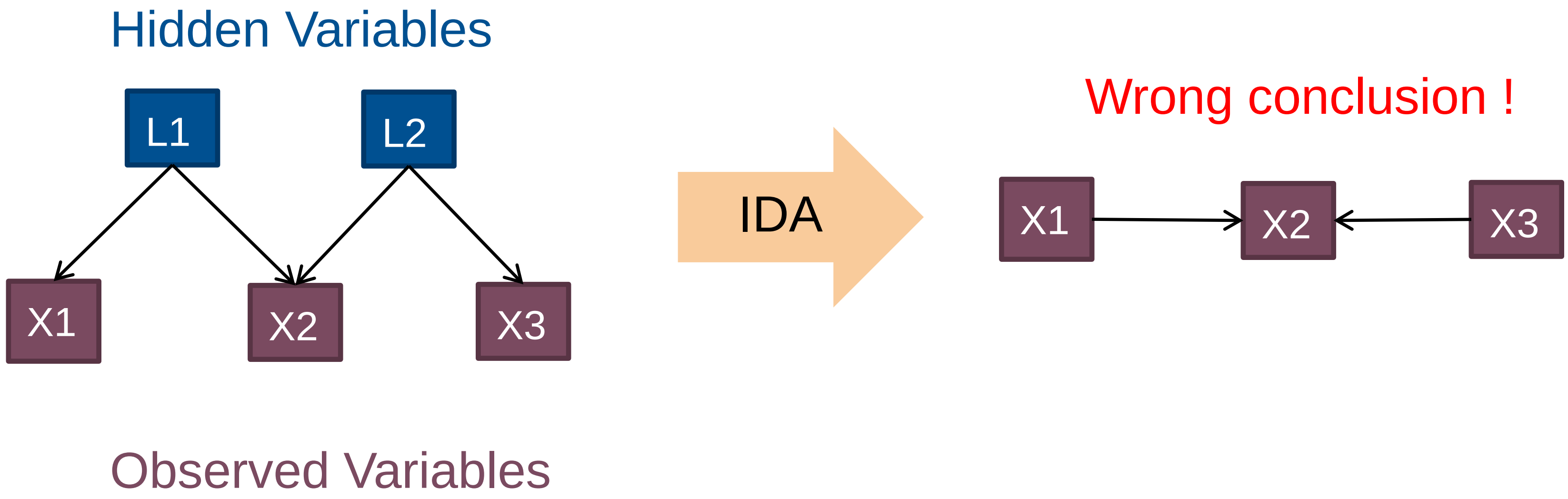
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Hidden variables



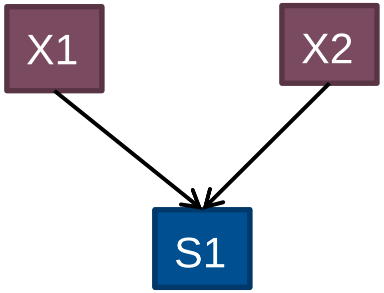
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Hidden variables



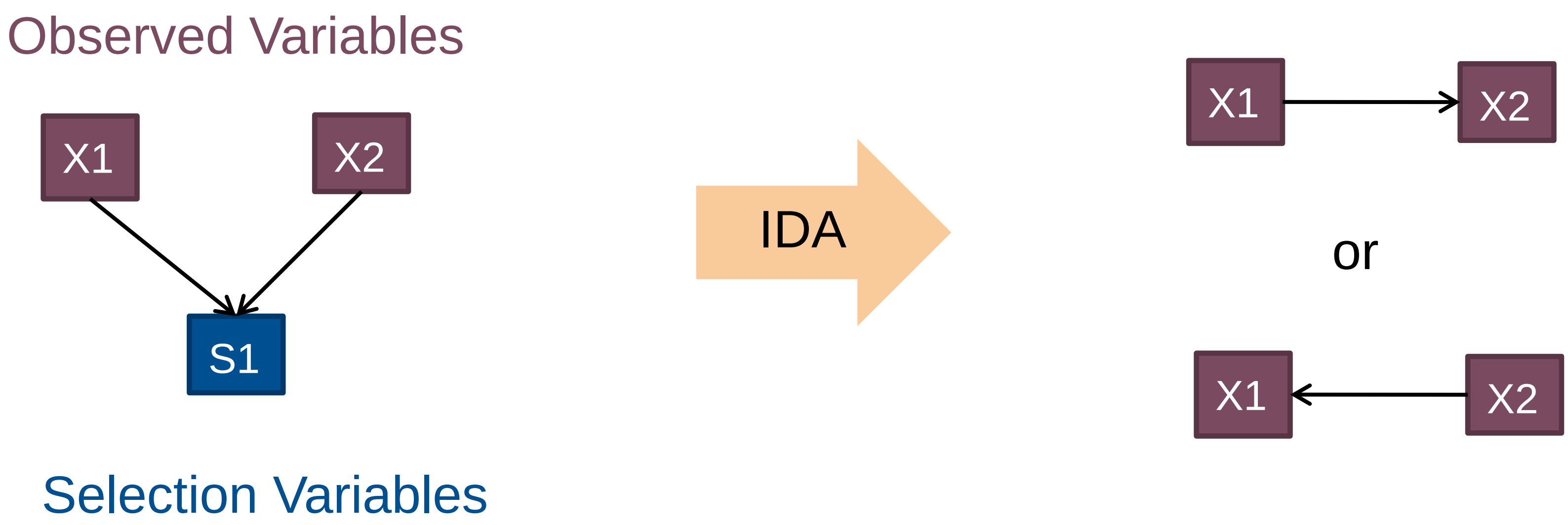
Selection variables

Observed Variables



Selection Variables

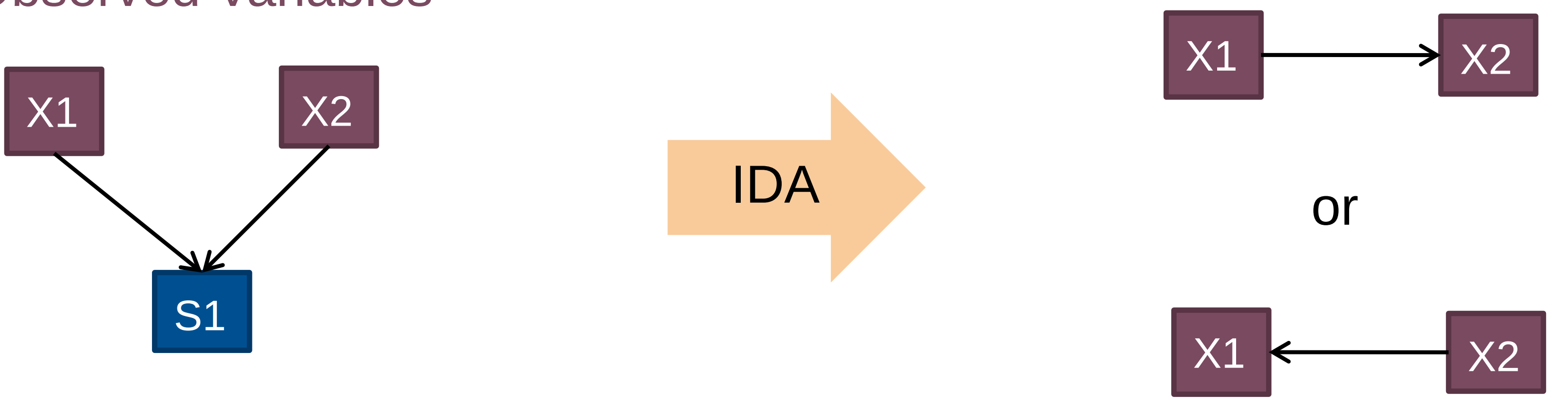
Selection variables



Selection variables

Wrong conclusion !

Observed Variables



Selection Variables

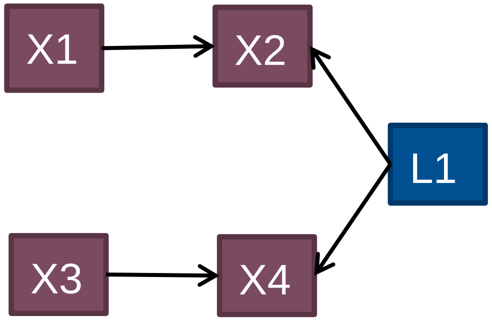
Question

How can we describe a system with arbitrarily many hidden or selection variables?

DAGs are not ideal

- Would have to know all hidden and selection variables
- Even if we knew them, there might be a problem: Space of DAGs not closed under marginalization and conditioning

Space of DAGs is not closed under marginalization



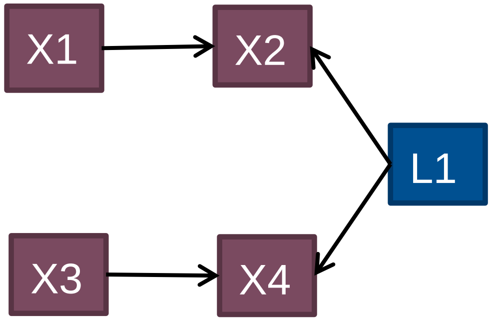
Implies

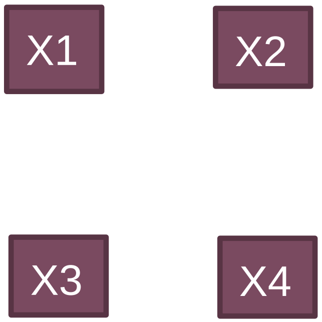
X1 indep. of X3 and X4

X2 indep. of X3

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Space of DAGs is not closed under marginalization





Implies

X1 indep. of X3 and X4

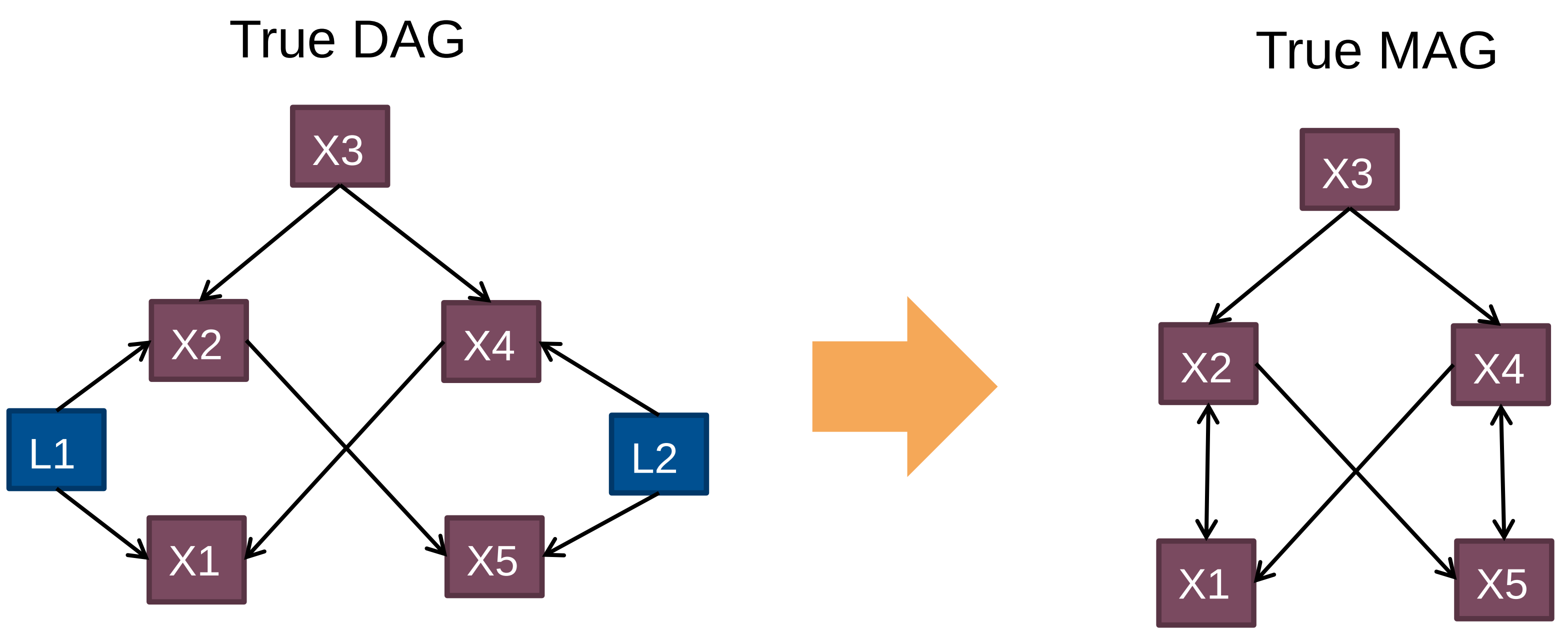
X2 indep. of X3

No DAG on the observed variables that implies the same conditional independencies

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Alternative that works: Maximal Ancestral Graphs (MAGs)

Richardson, T.S., Spirtes, P., Ancestral Graph Models, 2002, **Ann. Stat.** 30, 962-1030

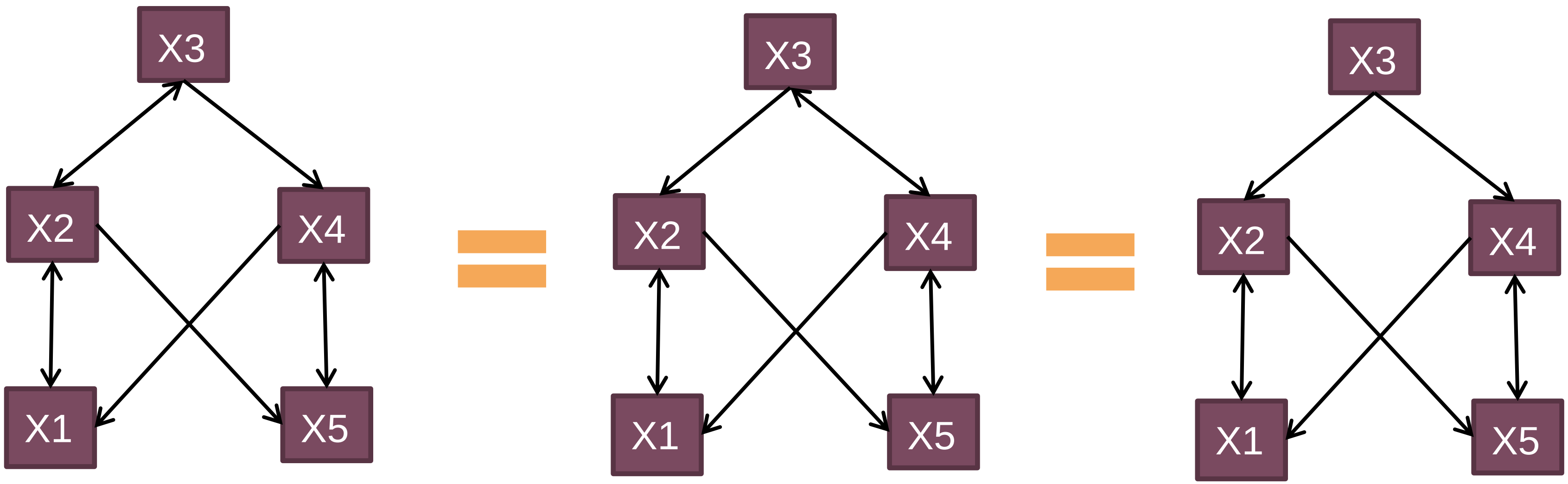


Arrowhead: X2 is no ancestor of X3 or a selection variable in true DAG

Arrowtail: X2 is an ancestor of X5 or a selection variable in true DAG

In practice: Equivalence class again…

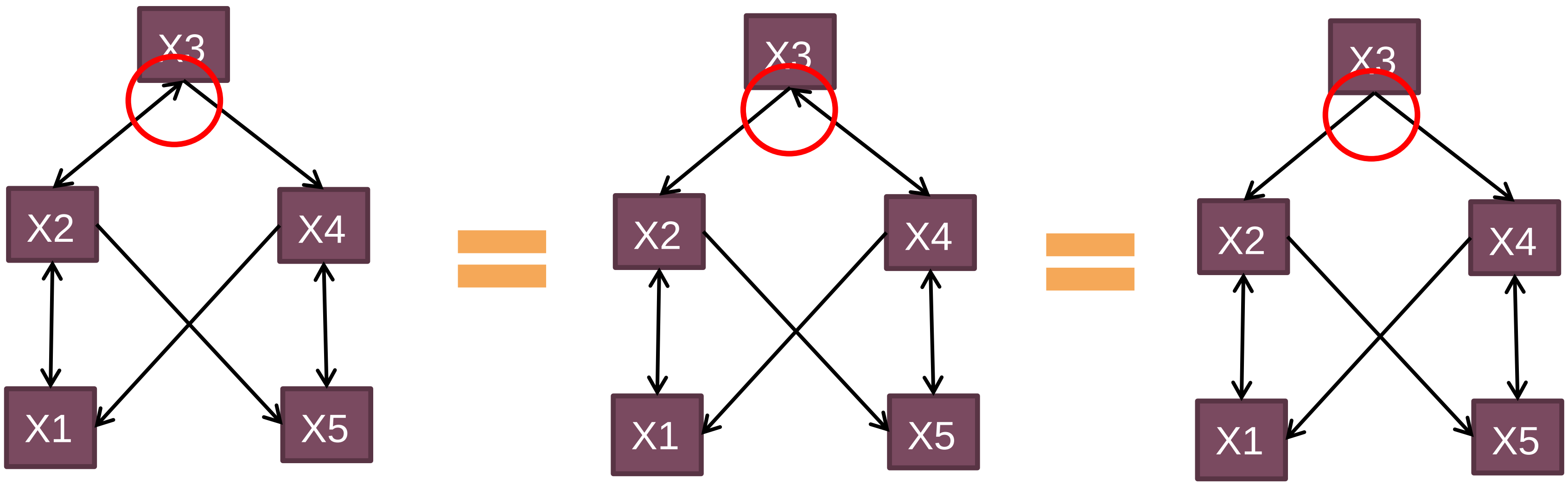
Several MAGs describe exactly the same list of independence relations



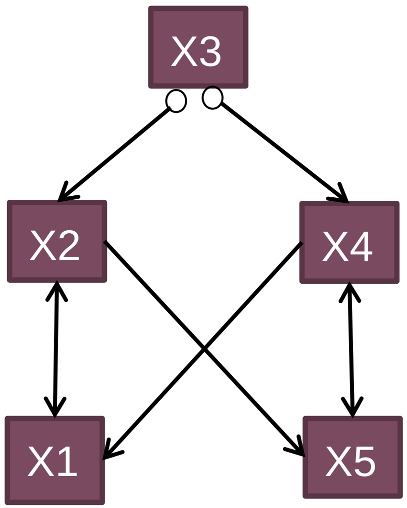
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

In practice: Equivalence class again…

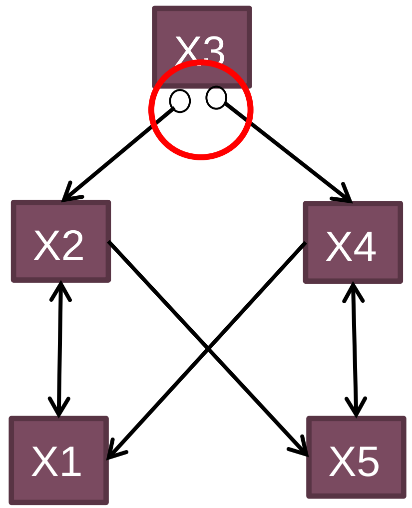
Several MAGs describe exactly the same list of independence relations



Equivalence class represented by a Partial Ancestral Graph (PAG)



Equivalence class represented by a Partial Ancestral Graph (PAG)

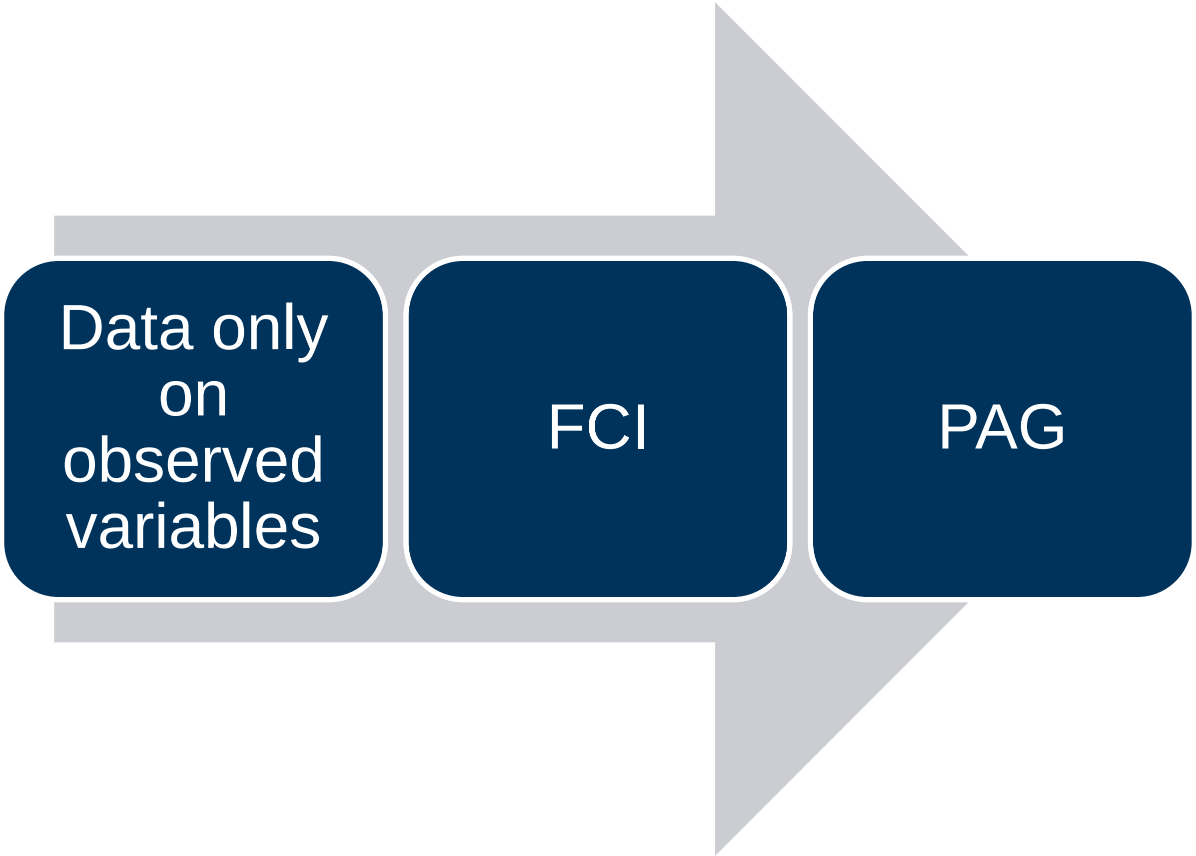


Is X1 ancestor of X4? – No!

Is X3 ancestor of X2? – Don’t know!

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Algorithm to find PAG from data: FCI



FCI is correct, but:

- Given distribution oracle: FCI is correct

Spirtes, P., Glymour, C. and Scheines, R., 2000, Causation, Prediction and Search, MIT Press

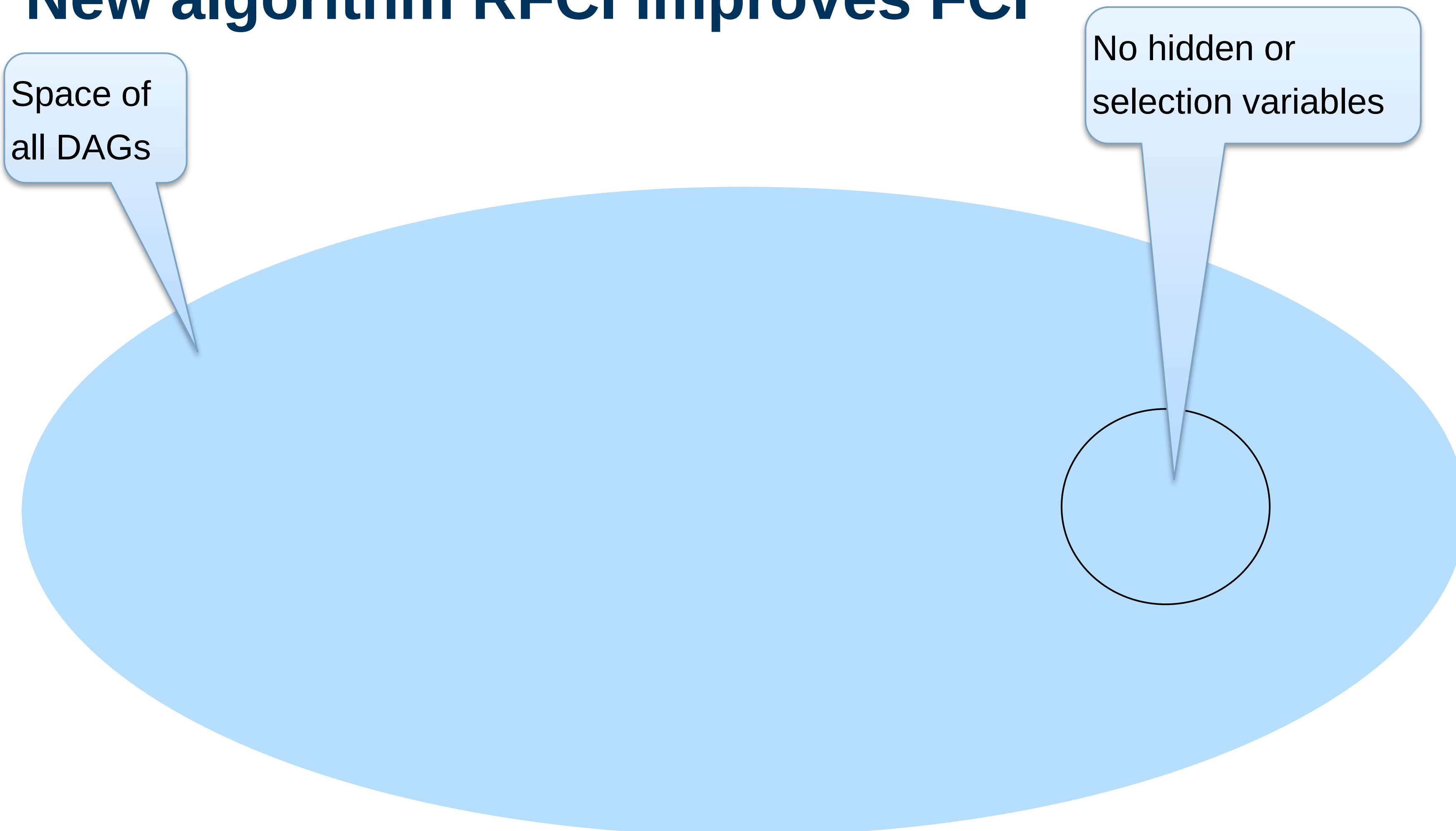
- Given data:
  - consistent under strict assumptions
  - only feasible for a handful of variables: Very slow

Colombo, D., Maathuis, M.H., Kalisch, M., Richardson, T.S.,

Learning high-dimensional DAGs with latent and selection variables, to be submitted

New algorithm RFCI improves FCI

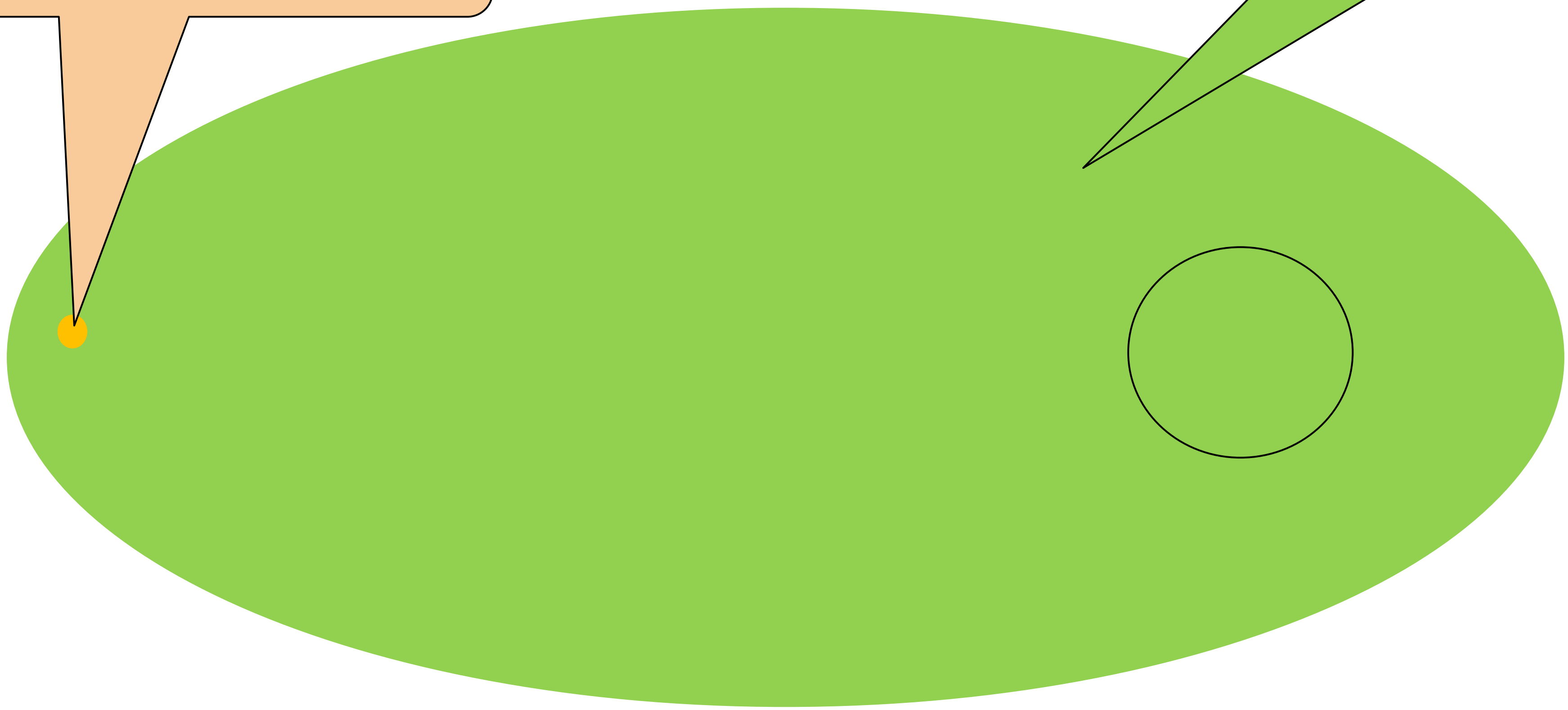
New algorithm RFCI improves FCI



New algorithm RFCI improves FCI

RFCI gives correct, but less information than FCI

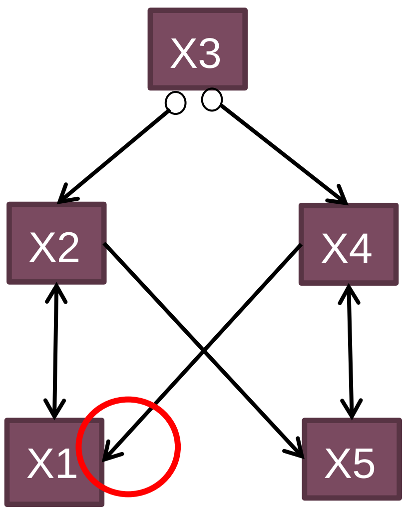
RFCI and FCI yield same result



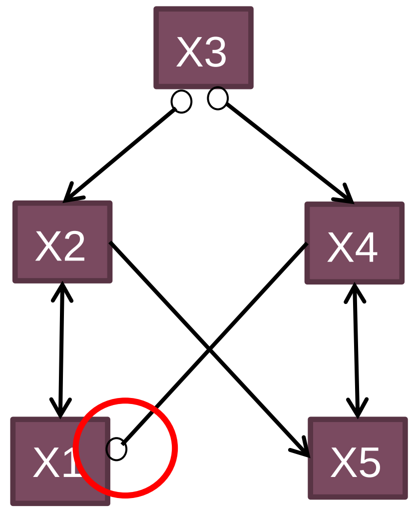
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

For the intuition:

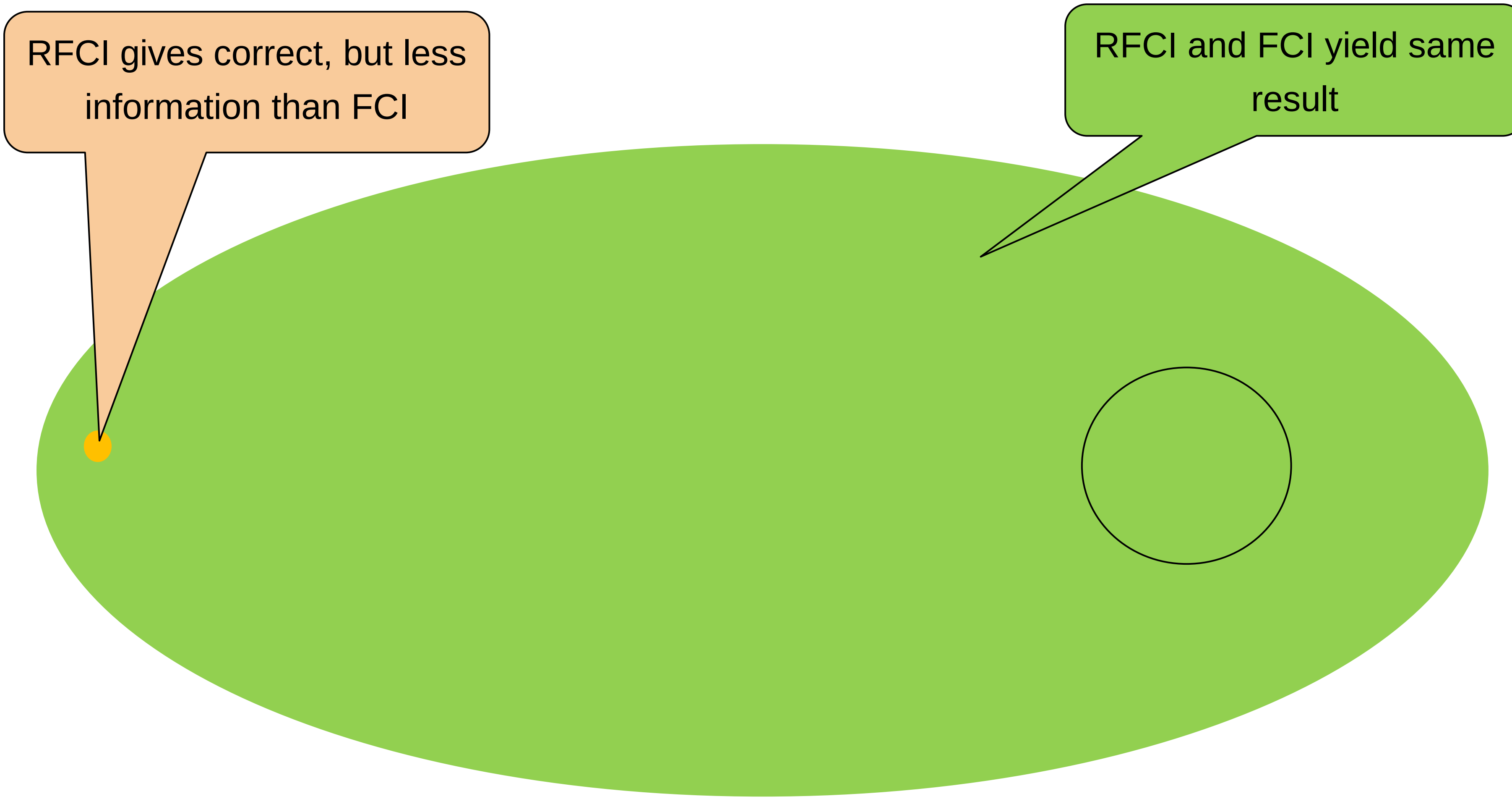
FCI



RFCI



New algorithm RFCI improves FCI



Simulation: 5 : 100’000

RFCI

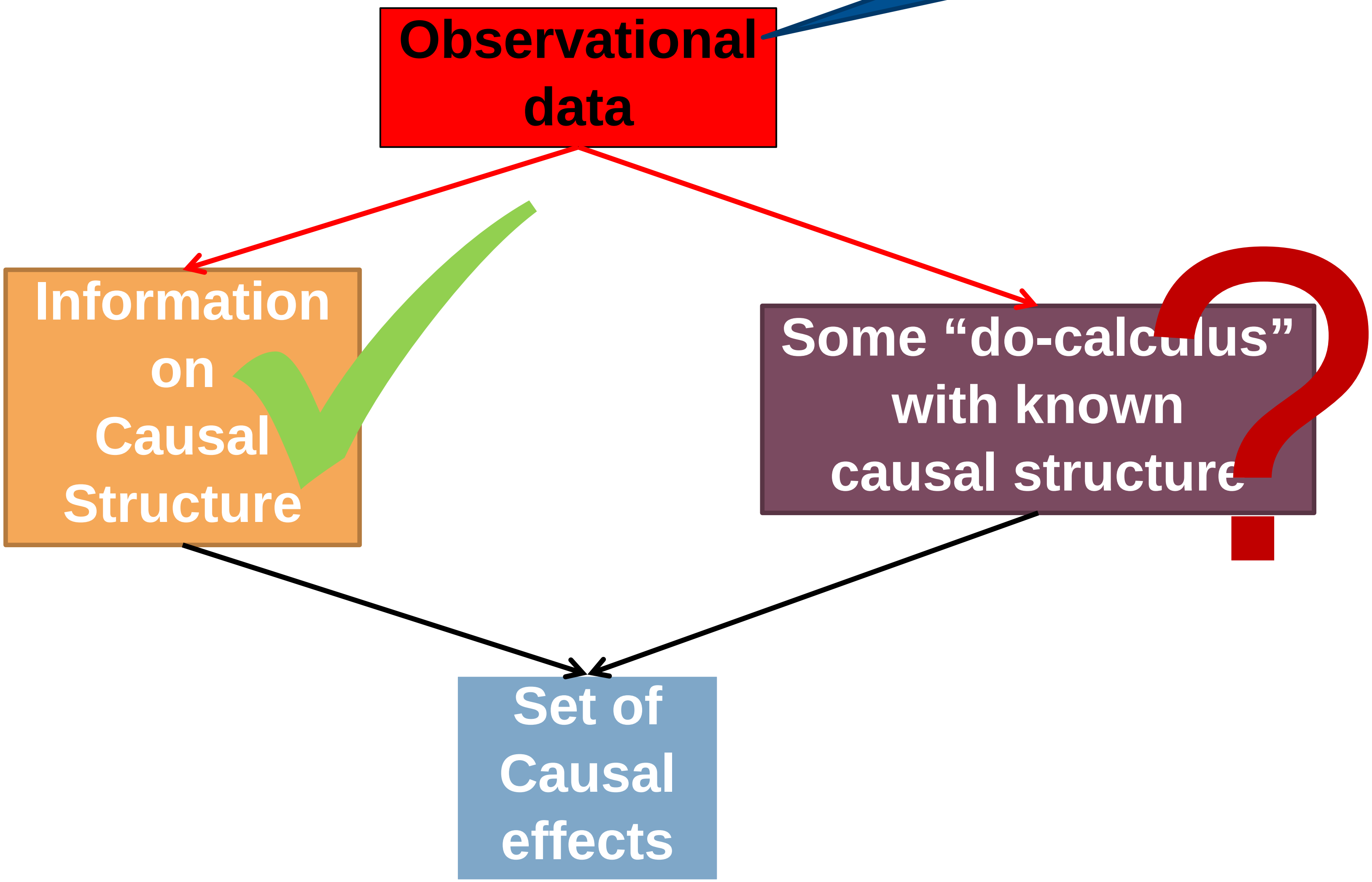
- Finds correct ancestral information of the true underlying DAG even if arbitrarily many (unknown) hidden and/or selection variables are present
- In general, it does not find all relations
- Very often as informative as FCI
- Consistent even in high dimensional setting (much weaker assumptions than in FCI)
- Computationally fast (FCI up to ~10 variables; RFCI up to thousands of variables)

Colombo, D., Maathuis, M.H., Kalisch, M., Richardson, T.S.,

Learning high-dimensional DAGs with latent and selection variables, to be submitted

Conclusion: Current research

Arbitrarily many, unknown hidden or selection variables



Thank you!