

A Practical Approach to Causal Inference

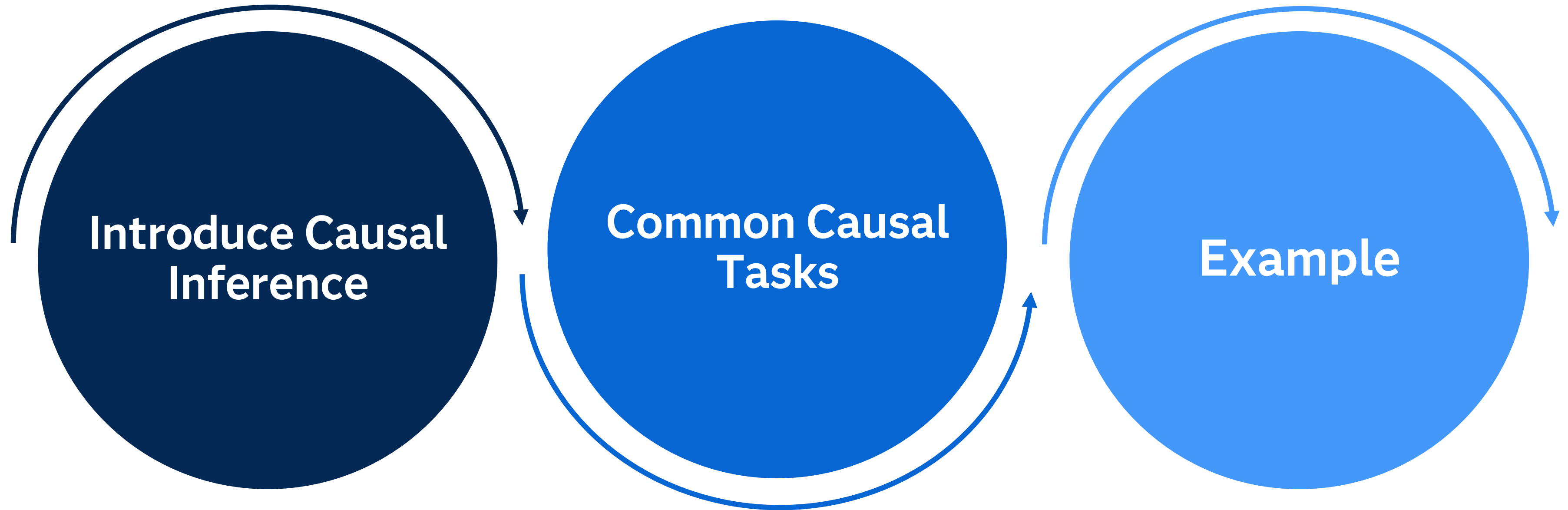
PHUSE 2026 Submission

Laura Watson Smith, Life Sciences Customer Advisor

Sherrine Eid, Global Head, Epidemiology, RWE, and Observational Research



Agenda



Define Causal Inference

Correlation $\neq$ Causation

Causality - a relationship where one event directly influences another, meaning the outcome would not occur without the cause.

Causal inference determines whether such cause-effect relationships exist in a population.

Practical Example: “Does quitting smoking cause weight change?”
US National Health Examination Survey (NHANES)

Statistical (Associational) Inference vs. Causal Inference

1. Answers Different Questions

Associational Inference

- What associations exist in the observed data?
(Associational Question)
- What associations exist between quitting smoking, weight change, and the other variables in the system?

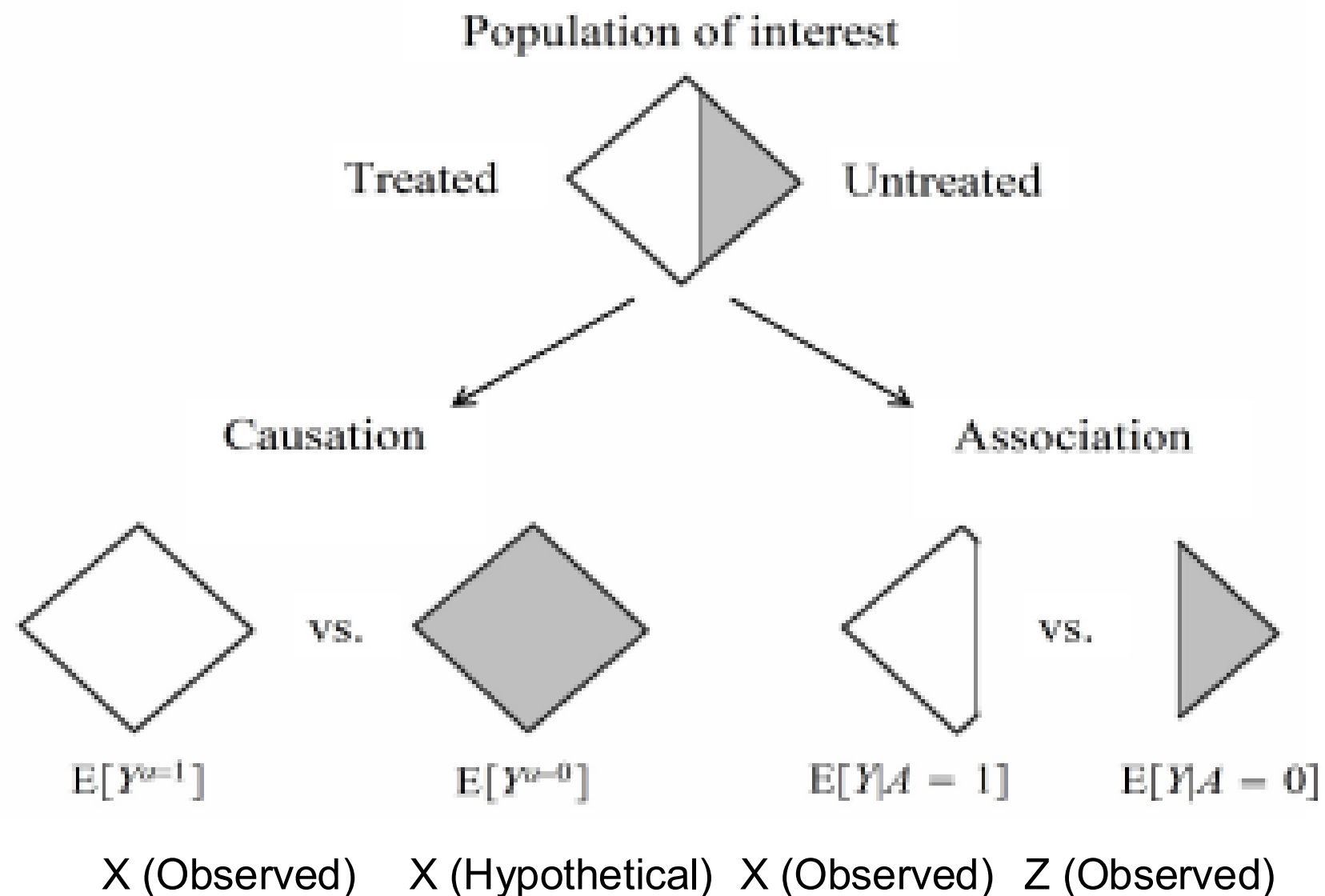
Causal Inference

- What would happen under a hypothetical intervention?
(Counterfactual Question)
- Would we see the same weight change regardless of if someone quit smoking or not?

Statistical Inference vs. Causal Inference

2. Considers Different Populations

“Is the outcome the same for treated and untreated individuals?”



Statistical Inference

- Compares Group X, who was treated, to Group Z, who was not, drawing conclusions from observed differences between real individuals
- Uses sample data to estimate what could be observed in the full population

Causal Inference

- Compares Person X, who was treated in reality, to that same Person X, had they not been treated (counterfactual) – for everybody in the data
- Uses sample data to infer outcomes in hypothetical scenarios that cannot be observed

Figure captured from Hernán MA, Robins JM (2020). Causal Inference: What If. Boca Raton: Chapman & Hall/CRC

Statistical Inference vs. Causal Inference

3. Asserts Different Assumptions

Statistical Inference

- Assumptions are related to distributions or model structures
 - Normality, Linearity, Independence, etc.
- Assumptions are testable in principle, given sufficient sample size

Causal Inference

- Requires additional assumptions beyond those for statistical analysis
- Identifying assumptions allow an observational study to be treated as a conditionally randomized experiment
- Such assumptions are untestable in principle

Why Perform Causal Inference?

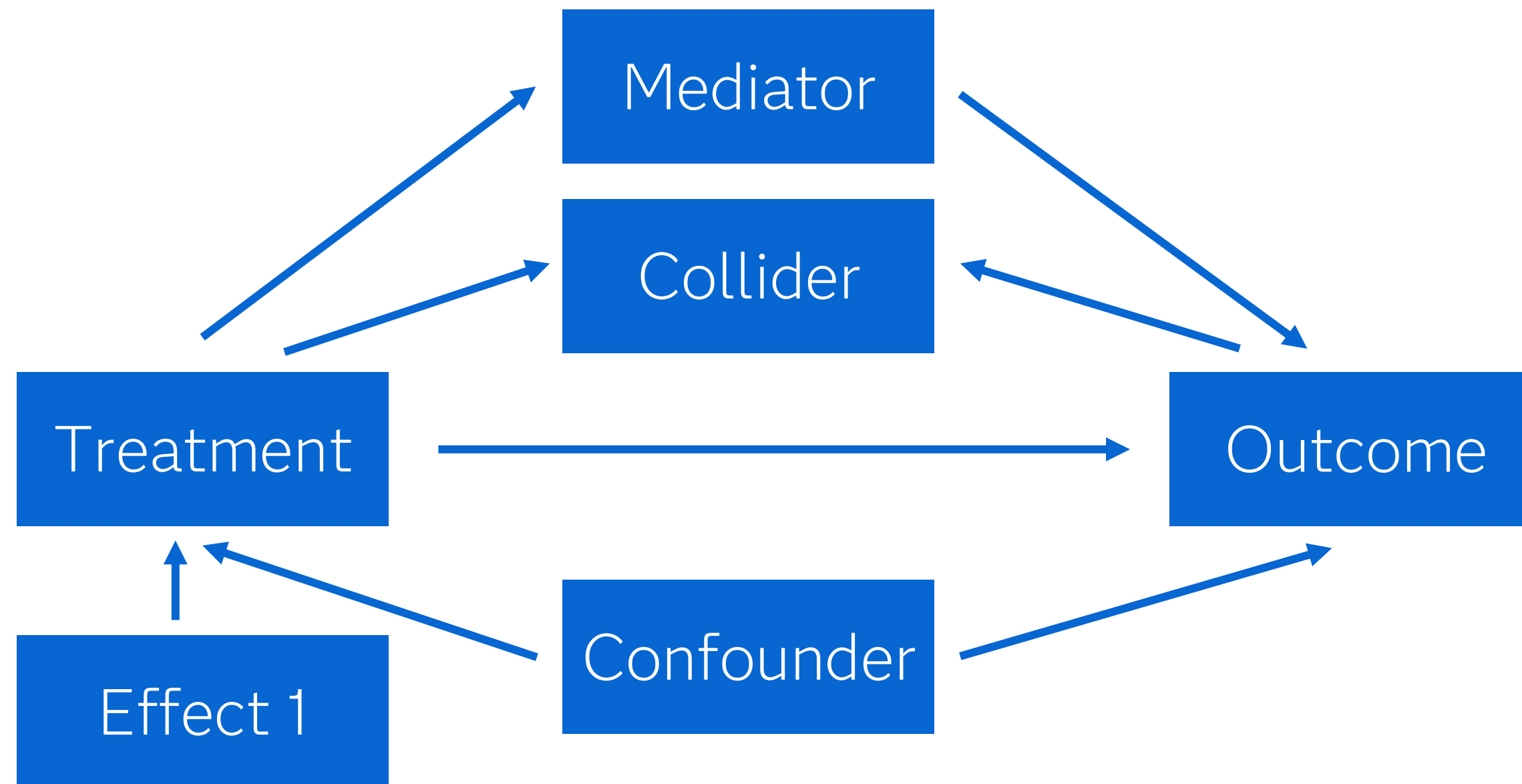
1. To answer specifically **causal questions**, for understanding what would happen in a system if certain conditions were changed
2. To make a **stronger claim** than one derived from statistical analysis
3. To generate insight in response to a **regulatory inquiry**

Causal Inference Challenges

1. **Unobservable Data:** we can never observe both potential outcomes for the same individual at the same time
2. **Association $\neq$ Causation:**
 - Different underlying populations
 - Association may imply other phenomena
3. **To Imply Causality:** Make a set of additional assumptions, including “identifying assumptions”

Common Causal Tasks

Construct a Causal Diagram – Directed Acyclic Graph (DAG) - Visualize the assumed relationships among variables included in the system as nodes linked through one-way directed arrows which form no loops or cycles.



Common Causal Tasks

Identification – Determine whether the causal estimand (average treatment effect) can be expressed as a function of the observed data.

- Make a set of “identifying assumptions”
 - If they hold... causal estimand is “uniquely identified”
 - If not... analysis halts
- Identification assumptions allow us to assert that the exposed and unexposed groups are identical in all respects, except for their treatment assignments

Effect Estimation – Estimate the effect of treatment or another covariate on the outcome for the entire population, or subset of the entire population.

Effect Estimation Example

Leveraging NHANES Smoking & Weight Change Data

Published Example – NHANES Data

Variable	Description
Activity	Level of daily activity (0, 1, or 2)
Age	Age in 1971
BaseWeight	Weight in kilograms in 1971
Education	Level of education (0, 1, 2, 3, or 4)
Exercise	Amount of regular recreational exercise (0, 1, or 2)
PerDay	Number of cigarettes smoked per day in 1971
Quit	1 if the individual quit smoking between interviews; 0 otherwise
Race	0 for white; 1 otherwise
Sex	0 for male; 1 for female
Weight	Weight in kilograms at the follow-up interview
YearsSmoke	Number of years the individual has smoked
Change	Difference in weight between follow-up and baseline (in kilograms)

- Based on [NHANES Epidemiologic Follow-up Study](#)
- Outcome: Change (10-year weight change)
- Treatment: Quit (1 = quit smoking, 0 = continued)

Sex	Age	Race	Education	Exercise	Change	Activity	YearsSmoke	PerDay	Quit	BMI
0	42	1	1	2	-10.0940	0	29	30	0	26.0503
0	36	0	2	0	2.6050	0	24	20	0	23.0823
1	56	1	2	2	9.4145	0	26	20	0	20.0090

Construct a Causal Diagram

Effect Estimation Dashboard: PROC CAUSALDISCOVERY

SAS PROC CAUSALDISCOVERY Procedure

The CAUSALDISCOVERY Procedure

Algorithm Information	
Algorithm	Topological Order Permutation
Initialization Method	Variance
Maximum Iterations	200
Parallel	Yes
Seed	1
GPU Active	No

Number of Observations Read	1566
Number of Observations Used	1566

Topological Orders		
Variable	Initial Order	Final Order
Sex	3	3
Age	11	11
Race	1	1
Education	6	6
Exercise	5	5
Change	7	7
Activity	4	4
YearsSmoke	10	10
PerDay	8	9
Quit	2	2
BMI	9	8

Estimated Full Adjacency Matrix												
Variable	Sex	Age	Race	Education	Exercise	Change	Activity	YearsSmoke	PerDay	Quit	BMI	
Sex	0	1	0	1	1	1	1	1	1	0	1	

Set Job Parameters

PROC CAUSALDISCOVERY

LEARN Statement

Input Table for Learning DAG Structure *
CASUSER.SMOKINGWEIGHT

Algorithm
☐ MCV
☒ TOP

Alpha
0.05

Init Method
☐ DATA
☐ RANDOM
☒ VARIANCE

OK

Cancel

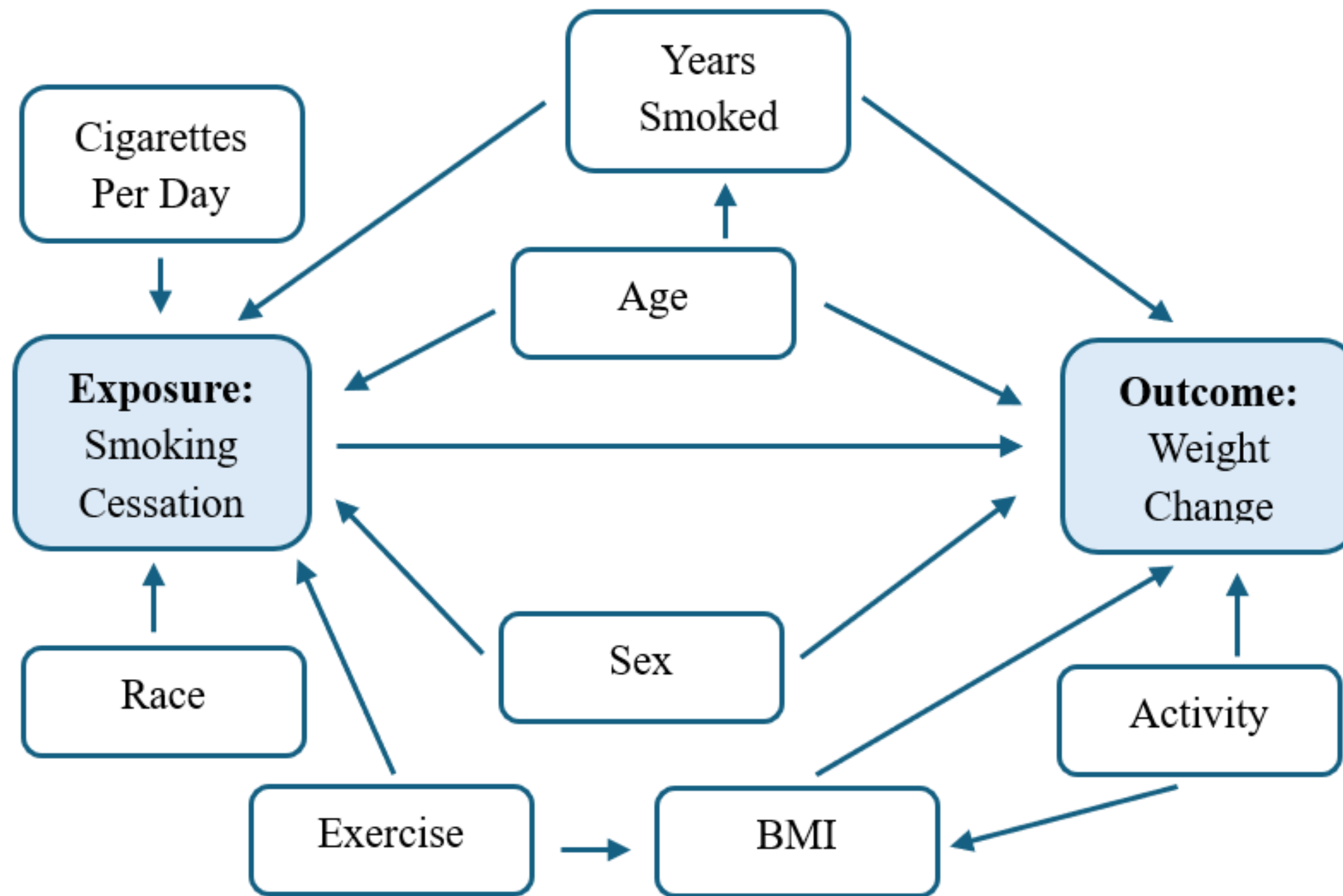
Construct a Causal Diagram

PROC CAUSALDISCOVERY Output: Estimated Adjacency Matrix

Estimated Adjacency Matrices												
Alpha	Variable	Sex	Age	Race	Education	Exercise	Change	Activity	YearsSmoke	PerDay	Quit	BMI
1E-6	Sex	0	1	0	1	1	0	1	0	1	0	1
	Age	0	0	0	0	0	0	0	0	0	0	0
	Race	1	0	0	0	1	0	0	0	1	1	1
	Education	0	1	0	0	0	1	0	1	1	0	1
	Exercise	0	0	0	1	0	0	0	1	0	0	1
	Change	0	0	0	0	0	0	0	0	0	0	0
	Activity	0	0	0	1	1	0	0	0	0	0	1
	YearsSmoke	0	1	0	0	0	0	0	0	0	0	0
	PerDay	0	0	0	0	0	0	0	1	0	0	0
	Quit	1	1	0	1	1	1	1	0	1	0	1
	BMI	0	1	0	0	0	0	0	1	1	0	0

Construct a Causal Diagram

Manually Constructed Causal Diagram of NHANES Example



```
/*B. CausalGraph*/  
/*Manual*/  
proc causalgraph method=backdoor;  
model 'TESTID Demo'  
SmokingCessation ==> WeightChange,  
YearsSmoke => SmokingCessation WeightChange,  
Sex => SmokingCessation WeightChange,  
  
...  
run;
```

Causal Identification

Effect Estimation Dashboard: Causal Graph

The CAUSALGRAPH Procedure

Variables in Model		
	N	Variables
Measured	11	Activity Age BMI Change Education Exercise PerDay Quit Race Sex YearsSmoke
Unmeasured	0	

Graphical Model Summary						
Model	Nodes	Edges	Treatments	Outcomes	Measured	Unmeasured
&Model	11	35	1	1	11	0

The CAUSALGRAPH Procedure

Covariate Adjustment Sets for &Model											
Causal Effect of Quit on Change											
	Covariates										
	Size	Minimal	Activity	Age	BMI	Education	Exercise	PerDay	Race	Sex	YearsSmoke
1	1	Yes							*		

Causal Identification

PROC CAUSALGRAPH Output: Covariate Adjustment Sets

Covariate Adjustment Sets for TESTID Demo										
Causal Effect of SmokingCessation on WeightChange										
	Size	Minimal	Covariates							
			Activity	Age	BMI	Exercise	PerDay	Race	Sex	YearsSmoke
1	4	Yes		*	*				*	*
2	4	Yes		*		*			*	*
3	5	No	*	*	*				*	*
4	5	No	*	*		*			*	*
5	5	No		*	*	*			*	*
6	5	No		*	*		*		*	*
7	5	No		*	*			*	*	*
8	5	No		*		*	*		*	*
9	5	No		*		*		*	*	*
10	6	No	*	*	*	*			*	*
11	6	No	*	*	*		*		*	*
12	6	No	*	*	*			*	*	*
13	6	No	*	*		*	*		*	*

Effect Estimation

Effect Estimation Dashboard: CausalTrt

PROC CAUSALTRT Procedure

The CAUSALTRT Procedure

Model Information	
Data Set	CASUSER.SMOKINGWEIGHT
Distribution	Normal
Link Function	Identity
Estimation Method	REGADJ
Treatment Variable	Quit
Outcome Variable	Change

Number of Observations Read	1566
Number of Observations Used	1566

Class Level Information			
Class	Levels	Values	
Sex	2	0 1	
Race	2	0 1	
Education	5	1 2 3 4 5	
Exercise	3	0 1 2	
Activity	3	0 1 2	
Quit	2	0 1	

Treatment Profile		
Ordered Value	Quit	Total Frequency
1	1	403
2	0	1163

PROC CAUSALTRT is modeling the probability that Quit='1'.

The CAUSAL TRT Procedure

Set Job Parameters

PSMODEL Statement

Exposure Variable *  

 Quit

☒ Click to model Event='1'

Variables for Modelling Exposu...  

☐  Sex
☐  Age
☐  Exercise
☐  YearsSmoke

Cutoff Value to Flag Large Weights

50

OK Cancel

Set Job Parameters

MODEL Statement

Outcome Variable *  

 Change

Variables for Modelling Outcome  

☐  Sex
☐  Age
☐  Race
☐  Education
☐  Exercise
☐  Activity
☐  YearsSmoke
☐  PerDay
☐  ...

Probability Distribution

☐ BIN
☐ GAMMA
☐ IGAUSSIAN
☒ NORMAL
☐ POISSON

OK Cancel

Effect Estimation

PROC CAUSALTRT Output: Analysis of Causal Effects

The CAUSALTRT Procedure

Analysis of Causal Effect							
Parameter	Treatment Level	Estimate	Robust Std Err	Wald 95% Confidence Limits		Z	Pr > Z
POM	1	5.1467	0.4177	4.3282	5.9653	12.32	<.0001
POM	0	1.7782	0.2152	1.3564	2.2001	8.26	<.0001
ATE		3.3685	0.4621	2.4628	4.2742	7.29	<.0001

Recap

We have introduced and explored the principles of causal inference:

1. **Discovery:** Build a causal diagram to visualize and understand the relationships of the variables in your system
 - Algorithmic and Manual methods possible
2. **Identification:** Verify that the average treatment effect can be computed given our data
 - Determine a group of confounders to satisfy the identification assumptions
3. **Estimation:** Compute the average treatment effect for the general population
 - We found a significant effect of smoking cessation on weight change!

Resources



Laura Watson Smith
Customer Advisor

Email:

laura.watson@sas.com

LinkedIn:



Sherrine Eid
Advisory Industry Consultant

Email:

sherrine.eid@sas.com

LinkedIn:



Acknowledgement

The authors would like to thank Clay Thompson and Joe Zhao, subject matter experts who were consulted in that capacity and provided guidance to ensure the technical correctness and clarity of this paper. The authors also thank Jim Box for editorial support and mentorship throughout the development of the paper.

Thank you / Questions 😊

