Lecture 7 – Assessing Assumptions in Causal Inference

Assessing Assumptions

Dr. Xi Chen

Rotterdam School of Management

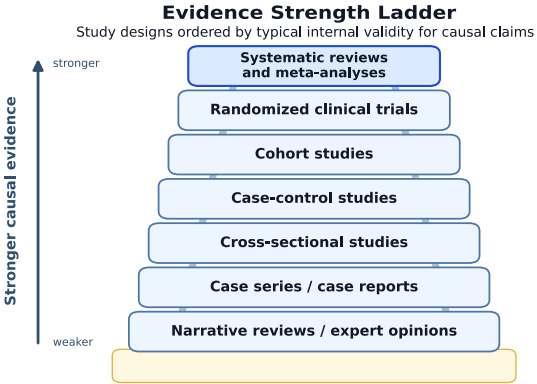
2026-06-20

Outline

1. Understanding assumptions in causal inference
   - Assumptions on potential outcomes are untestable
   - Assumptions are not unique and subjective
   - The law of decreasing credibility
2. Assessing assumptions
   - Sensitivity analysis
   - Consistency tests
3. Conclusions

The ladder of evidence

- This ranking feels intuitive: some designs seem "more credible" than others.
- It is useful, but incomplete.
- **Today's question**: what makes a rung on this ladder credible?



Reproduced from Raevschi, Penina, and Obreja, Basic Biostatistics and Research Methodology, Fig. 8.3, "Research Study Design Evidence Strength Pyramid."

Where the evidence ladder can mislead

Design label	What sounds reassuring	What still needs to be true
RCT	"Treatment is randomized."	No interference, no attrition bias, compliance or clear estimand.
DID	"We compare changes, not levels."	Parallel trends, no anticipation, stable composition.
IV	"The instrument moves treatment."	Exclusion, exogeneity, relevance, monotonicity.
Regression	"We control for confounders."	No omitted confounders, good measurement, reasonable functional form.

Status quo to question

Is a design higher in the ladder necessarily more credible?

Opening poll: same question, different evidence

Question: Does an entrepreneurship training program increase firm survival?

Study A: Lottery

Study description

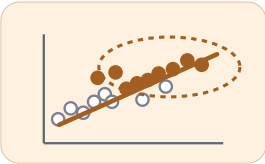
Applicants are randomly offered training, but many do not attend and some share materials with friends.

Study B: DID

Study description

One region introduces the program. Another region does not. Their survival rates moved similarly before the program.

Study C: Regression



Study description

We observe rich pre-program data, but entrepreneurs choose whether to participate.

Vote: Which study would you trust most?

What did you actually evaluate?

You probably did not only evaluate the method.

- Was the lottery implemented cleanly?
- Are spillovers likely?
- Are the two regions comparable?
- Could selection into training be explained by observed covariates?

You evaluated assumptions.

- SUTVA / no interference
- Compliance and estimand
- Parallel trends
- Conditional exogeneity

What did you actually evaluate?

You probably did not only evaluate the method.

- Was the lottery implemented cleanly?
- Are spillovers likely?
- Are the two regions comparable?
- Could selection into training be explained by observed covariates?

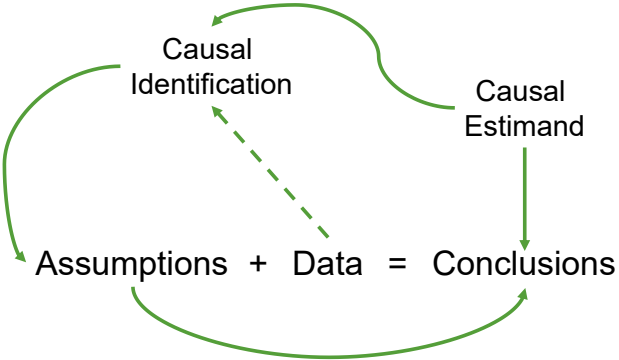
You evaluated assumptions.

- SUTVA / no interference
- Compliance and estimand
- Parallel trends
- Conditional exogeneity

Conclusion

Causal inference is not a contest between methods. It is an argument about assumptions.

The role of assumptions in causal inference



- Assumptions are always needed.
- Different assumptions ⇒ Different conclusions.
- Credibility of assumptions ⇒ Credibility of conclusions.

Assumptions determine conclusions

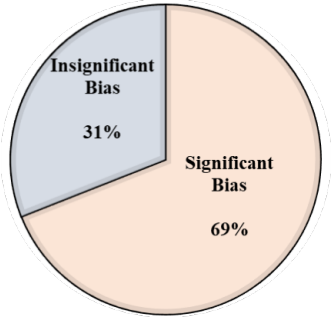
A recent study[1]:

- Asks teams of researchers to analyze the same data set - **football matches**.
- The data include information such as player clubs, birth date, height, weight, and positions...
- **Research question**: *are referees more likely to give red cards to players with darker rather than lighter skin tones*?

[1] Silberzahn, R. et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. Advances in Methods and Practices in Psychological Science, 1(3), 337-356.

Assumptions determine conclusions

Results of the study:



- Researchers make their own assumptions and get different results.
- "*What theoretical and/or statistical rationale was used for your choice of covariates?*"

Assumptions determine conclusions

Similar studies in other fields show consistent observations:

- Biology:
  - Gould, E. et al. (2023). Same data, different analysts: variation in effect sizes due to analytical decisions in ecology and evolutionary biology.
- Economics:
  - Breznau, N. et al. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. Proceedings of the National Academy of Sciences, 119(44), e2203150119.
- Neuroscience:
  - Botvinik-Nezer, R. et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. Nature, 582(7810), 84-88.

1 Understanding assumptions in causal inference
Assumptions on potential outcomes are untestable
Assumptions are not unique and subjective
The law of decreasing credibility

2 Assessing assumptions

3 Conclusions

1 Understanding assumptions in causal inference
Assumptions on potential outcomes are untestable
Assumptions are not unique and subjective
The law of decreasing credibility

2 Assessing assumptions

3 Conclusions

Examples of assumptions

Stable Unit Treatment Value Assumption (SUTVA)

For a completely randomized experiment (CRE), we must have:

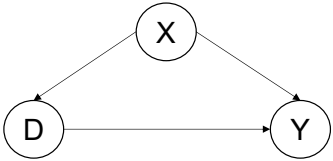
1. The potential outcomes for any unit do not vary with the treatments of others.
2. No different forms or versions of each treatment level.

Question: suppose we have a CRE, can we statistically test SUTVA with data?

Examples of assumptions

(Conditional) Unconfoundedness

The potential outcomes are independent of the treatment status (given a set of control variables X).

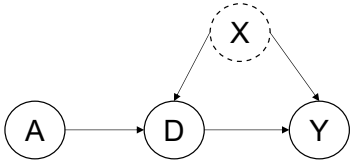


Question: With data on $\{Y, D, X\}$, can we statistically test the unconfoundedness assumption?

Examples of assumptions

Non-compliance (instrumental variables)

To identify the local average treatment effect, the treatment assignment (instrumental variable) must be 1) exogenous, 2) relevant, 3) excluded, and 4) monotonic (no defiers).

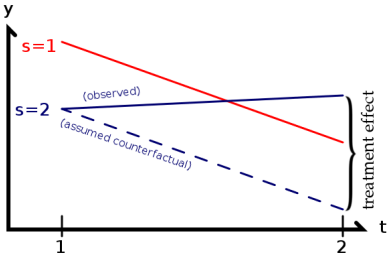


Question: With data on $\{Y, A, D\}$, can we statistically test the assumptions on instrumental variables?

Examples of assumptions

Parallel Trends Assumption in DID

In the absence of treatment, the difference between the treatment and control group is constant over time.

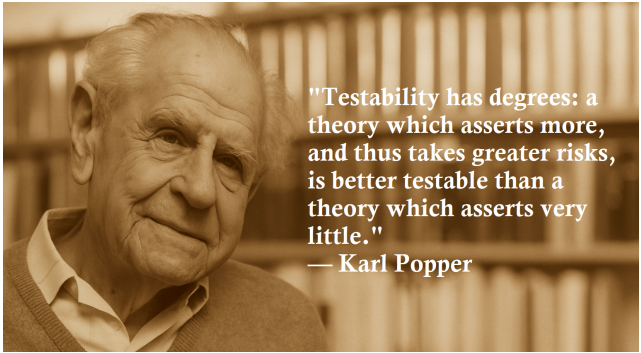


Question: With data on $\{Y_0, Y_1, D_0, D_1\}$, can we statistically test the parallel trends assumption?

Assumptions on potential outcomes are untestable!

Untestability of Assumptions

Given the fundamental problem of causal inference, any assumptions on potential outcomes are always statistically untestable.



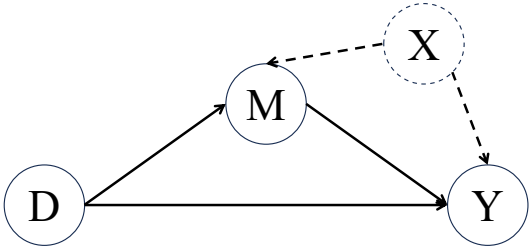
1 Understanding assumptions in causal inference
Assumptions on potential outcomes are untestable
Assumptions are not unique and subjective
The law of decreasing credibility

2 Assessing assumptions

3 Conclusions

Assumptions are not unique

The core problem of the traditional mediation analysis:

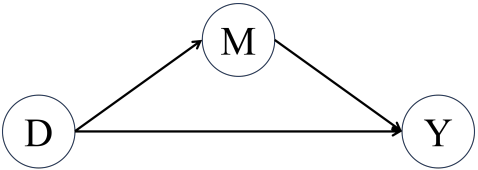


$$\begin{cases} Y &= \alpha_1 + \beta_1 D + e_1 \\ M &= \alpha_2 + \beta_2 D + e_2 \\ Y &= \alpha_3 + \beta_3 D + \beta_4 M + e_3 \end{cases}$$

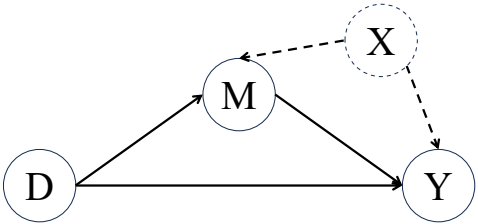
Question: what assumptions are needed for Baron and Kenny's (1986) approach?

Assumptions are not unique

(a) Baron DAG



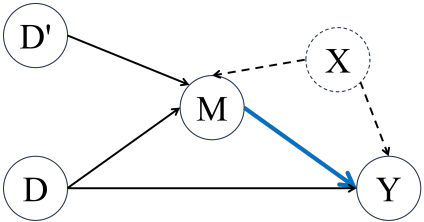
(b) Mediation DAG



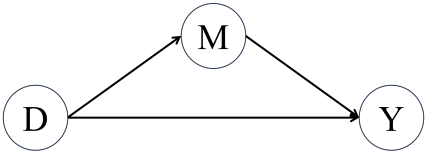
1. Treatment D is unconfounded.
2. Mediator M is unconfounded.

Assumptions are not unique

(a) Constructed IV



(b) Baron DAG

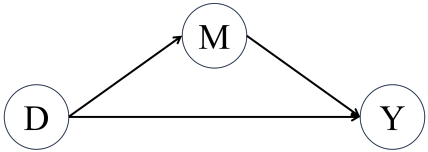


Constructed IV (D'):

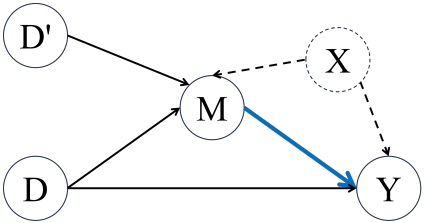
1. U is homoskedastic w.r.t. D or $\text{Cov}(De_2, e_3) = 0$.
2. V_2 is heteroskedastic w.r.t. D or $\text{Cov}(De_2, e_2) \neq 0$.

Assumptions are not unique

(a) Baron DAG



(b) Constructed IV



Traditional Mediation	Constructed IV
$\rho(e_2, e_3) = 0$	$\rho(e_2, e_3) \neq 0$
-11.7 $(-20.2, -4.8)$	-5.2 $(-12.1, 0.4)$

The selection of assumptions is subjective

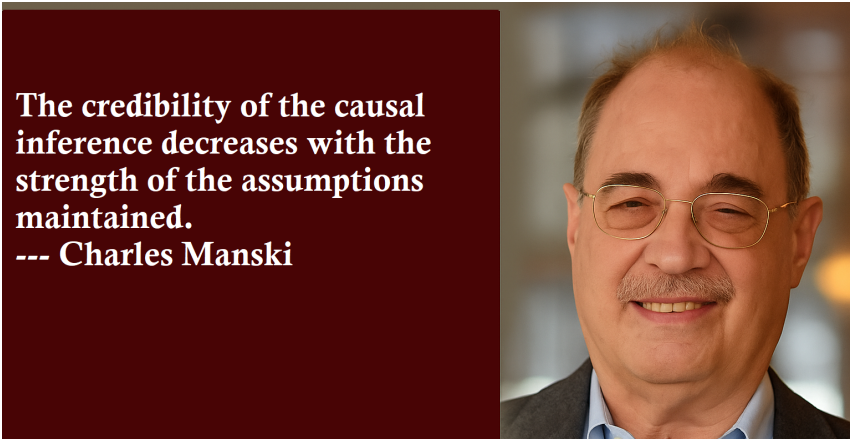
In Silberzahn et al. (2018), researchers reported the rationale of their analytic decisions:

1 Understanding assumptions in causal inference
Assumptions on potential outcomes are untestable
Assumptions are not unique and subjective
The law of decreasing credibility

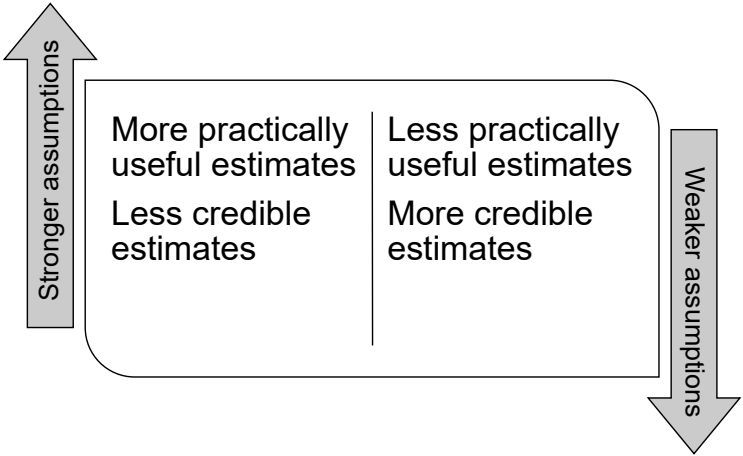
2 Assessing assumptions

3 Conclusions

The law of decreasing credibility

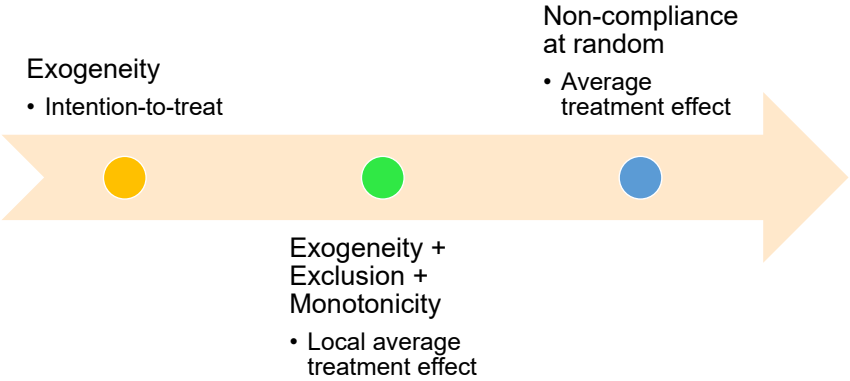


The law of decreasing credibility



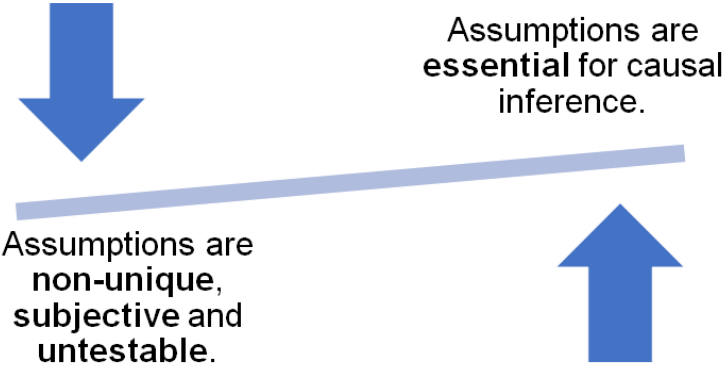
The law of decreasing credibility

Example (Non-compliance)

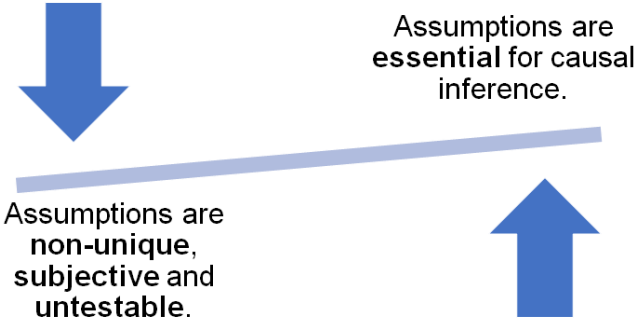


A recap on assumptions

We always need assumptions for causal inference practices!



A recap on assumptions



- How do we know the credibility of assumptions?
- What tools are available?

1 Understanding assumptions in causal inference

2 Assessing assumptions
Sensitivity analysis
Consistency tests

3 Conclusions

Tools in our toolkit

- The assumptions on potential outcomes are untestable.
- However, we can observe the implications of the assumptions.
- We assess assumptions by testing the implications.

Sensitivity analysis

Consistency test

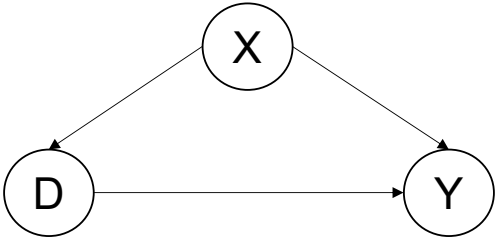
1 Understanding assumptions in causal inference

2 Assessing assumptions
Sensitivity analysis
Consistency tests

3 Conclusions

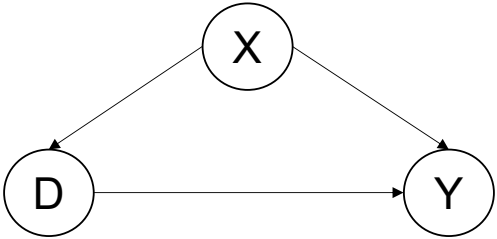
The basic idea of sensitivity analysis

- In practice, we often apply conditioning strategy and assume conditional unconfoundedness.



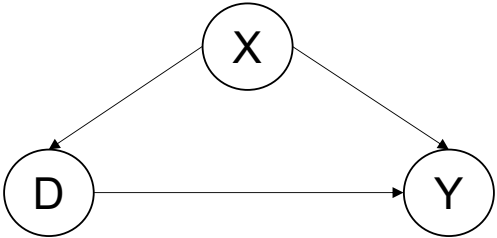
The basic idea of sensitivity analysis

- In practice, we often apply conditioning strategy and assume conditional unconfoundedness.
- This is a strong assumption, because it is always possible that confounders exist.



The basic idea of sensitivity analysis

- In practice, we often apply conditioning strategy and assume conditional unconfoundedness.
- This is a strong assumption, because it is always possible that confounders exist.
- **Question: if confounders exist, how do they change the ATE estimate?**



The logic of sensitivity analysis

Assume a confounder exists as U

Simulate the values of $U \sim f(U \mid D, Y, X)$

Use both U and X and re-estimate $P(Y \mid D)$

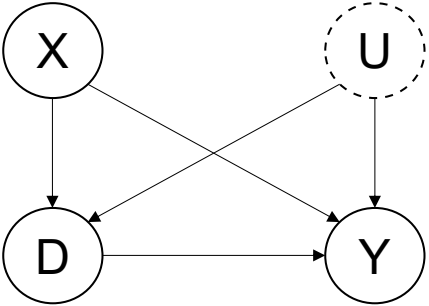
Examine the change in $P(Y \mid D)$ with vs. without U

The basic logic: the appeal to absurdity (*reductio ad absurdum*).

Sensitivity analysis: A toy model

A simple model with treatment D, outcome Y, observable X and a single unobservable $U \perp X$.

$$D := a_x X + a_u U$$
$$Y := b_x X + b_u U + \underbrace{\tau}_{\text{Objective}} D$$



Sensitivity analysis: A toy model

$$\begin{cases} D & := \alpha_x X + \alpha_u U \\ Y & := \beta_x X + \beta_u U + \tau D \end{cases}$$

If we observe both X and U, we can recover τ:

$$E\left(Y_1 - Y_0\right) = E_{X,U}\left[E\left(Y \mid D = 1, X, U\right) - E\left(Y \mid D = 0, X, U\right)\right] = \tau$$

Yet, we do not observe U. If condition on X only, we get:

$$E_X\left[E\left(Y \mid D = 1, X\right) - E\left(Y \mid D = 0, X\right)\right] = \tau + \underbrace{\frac{\beta_u}{\alpha_u}}_{\text{Bias}}$$

Sensitivity analysis: a toy model

Technical proof:

$$
\begin{aligned}
E_X\left[E\left(Y \mid D=d, X\right)\right] &= E_X\left[E\left(\beta_x X+\beta_u U+\tau D \mid D=d, X\right)\right] \\
&= E_X\left[\beta_x X+\beta_u E\left(U \mid D=d, X\right)+\tau d\right] \\
&= E_X\left[\beta_x X+\beta_u \underbrace{\left(\frac{d-\alpha_x X}{\alpha_u}\right)}_{\text{From the treatment equation}}+\tau d\right] \\
&= \left(\tau+\frac{\beta_u}{\alpha_u}\right) d+\left(\beta_x-\frac{\beta_u \alpha_x}{\alpha_u}\right) E\left(X\right)
\end{aligned}
$$

Therefore, we have the following,

$$
E_X\left[E\left(Y \mid D=1, X\right)-E\left(Y \mid D=0, X\right)\right]=\left(\tau+\frac{\beta_u}{\alpha_u}\right)(1-0)=\tau+\frac{\beta_u}{\alpha_u}
$$

Sensitivity analysis: a toy model

With a closed-form solution for the bias, how do we use it?

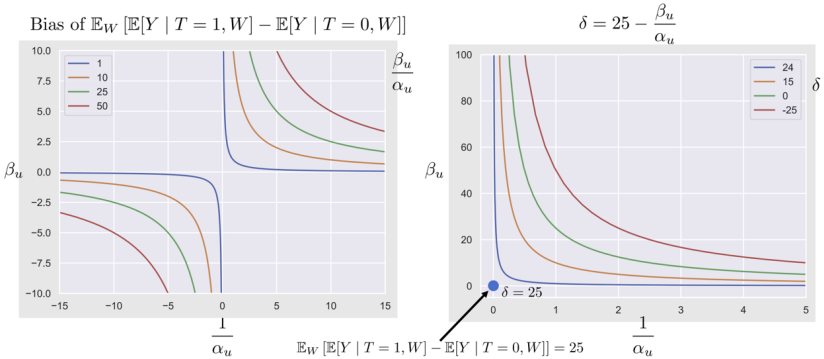
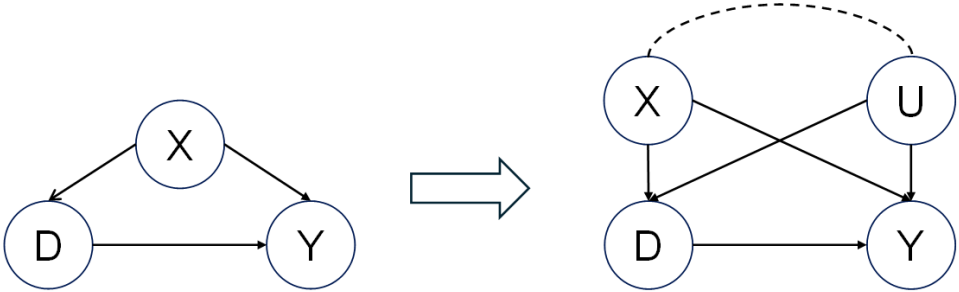


Figure 4: Contour Plots of Bias (left) and "True" Treatment Effect (right)

A unifying framework of sensitivity analysis

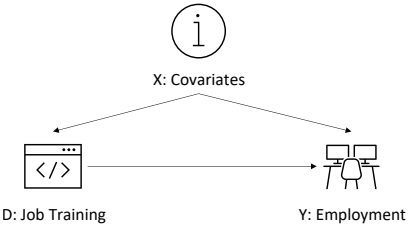


Different sensitivity analyses differ in how U is simulated.

Sensitivity analysis: Rosenbaum's approach for matching

Rosenbaum (1987) developed a sensitivity analysis for matching.

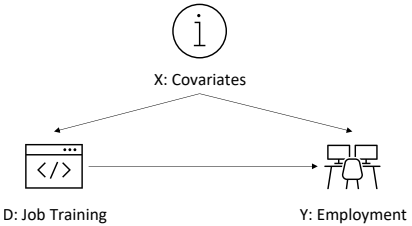
- How can **a job training program** (D) help unemployed people **find a job** (Y)?



Sensitivity analysis: Rosenbaum's approach for matching

Rosenbaum (1987) developed a sensitivity analysis for matching.

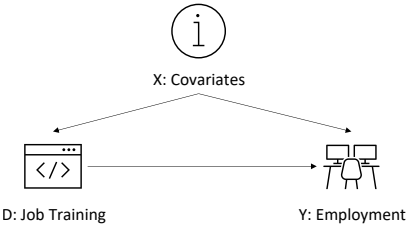
- How can **a job training program** (D) help unemployed people **find a job** (Y)?
- Participation is self-selected, but covariates X are observed.



Sensitivity analysis: Rosenbaum's approach for matching

Rosenbaum (1987) developed a sensitivity analysis for matching.

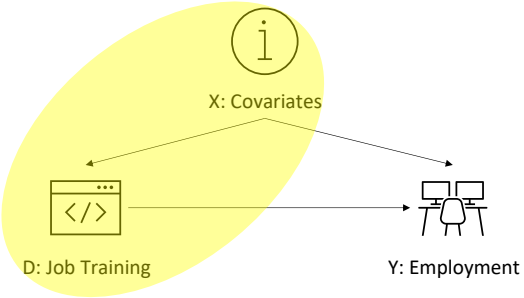
- How can **a job training program** (D) help unemployed people **find a job** (Y)?
- Participation is self-selected, but covariates X are observed.
- Define the propensity score $\pi_i = P(D_i = 1 \mid X_i)$.



Sensitivity analysis: Rosenbaum's approach for matching

Rosenbaum (1987) developed a sensitivity analysis for matching.

The confounder U is simulated based on $P(D_i|X_i)$.



Sensitivity analysis: Rosenbaum's approach

Observation

For **a matched pair** with the same covariates X_i (e.g., age, gender...), the propensity score should be equal, if there are no confounders.

$$\begin{cases} \text{Odds}_{\text{Bob}} &= \dfrac{\pi_{\text{Bob}}}{1-\pi_{\text{Bob}}} \\ \text{Odds}_{\text{Jim}} &= \dfrac{\pi_{\text{Jim}}}{1-\pi_{\text{Jim}}} \end{cases} \text{ and } \Gamma = \frac{\dfrac{\pi_{\text{Bob}}}{1-\pi_{\text{Bob}}}}{\dfrac{\pi_{\text{Jim}}}{1-\pi_{\text{Jim}}}} = 1$$

Sensitivity analysis: Rosenbaum's approach

Existence of confounders

If no confounders exist, $\Gamma = 1$. The existence of confounders $\rightarrow \Gamma \neq 1$.

$$\Gamma = \frac{\dfrac{\pi_{\text{Bob}}}{1 - \pi_{\text{Bob}}}}{\dfrac{\pi_{\text{Jim}}}{1 - \pi_{\text{Jim}}}} \neq 1$$

- Γ represents the severity of the confoundedness.
- Increase the value of Γ to examine changes in the treatment effects.

Sensitivity analysis: Rosenbaum's approach

		Job Training	
		$D = 0$	$D = 1$
Employment	$Y = 0$	10	c = 50
	$Y = 1$	b = 20	140

- Suppose we use McNemar's exact p-value to test the difference in the employment rate with vs. without the job training ($n = b + c$).

$$\text{p-value} = 2\sum_{k=b}^{n}\binom{n}{k}\left(\frac{\Gamma}{1+\Gamma}\right)^{k}\left(\frac{1}{1+\Gamma}\right)^{n-k}$$

- Under unconfoundedness, $\Gamma = 1$. A larger Γ implies more serious confounding.

Sensitivity analysis: Rosenbaum's approach

p-value	prob $\left(\frac{\Gamma}{1+\Gamma}\right)$	Γ
0.004	0.5000	1.00
0.0018	0.5238	1.10
0.0057	0.5455	1.20
0.0148	0.5652	1.30
0.0328	0.5833	1.40
0.0635	0.6000	1.50

- When $\Gamma = 1.5$, the p-value is larger than 0.05.
- To nullify the effect: a confounder that makes Bob 1.5 times more likely to join the program than Jim.
- Is $1.5\times$ a good indicator of validity? You are the judge!

Sensitivity analysis: Rosenbaum's approach

Technical proof: The probability in McNemar's exact p-value is equal to $\Gamma/(1+\Gamma)$.

The probability in the McNemar's test is θ_i, indicating $P(d_i = 1 \mid d_i + d_j = 1)$, or the probability that subject i is treated, conditional on one member of the matched pair $\langle i, j \rangle$ being treated.

$$\begin{aligned}\theta_i &= \frac{P(d_i = 1, d_j = 0)}{P(d_i = 1, d_j = 0) + P(d_i = 0, d_j = 1)} \\ &= \frac{\pi_i (1 - \pi_j)}{\pi_i (1 - \pi_j) + \pi_j (1 - \pi_i)}\end{aligned}$$

Then we must have

$$\frac{\theta_i}{1 - \theta_i} = \frac{\pi_i (1 - \pi_j)}{\pi_j (1 - \pi_i)} \overset{\text{Definition}}{=} \Gamma$$

Therefore, $\theta_i = \Gamma/(1+\Gamma)$.

Sensitivity analysis: selection bias

- Rosenbaum's approach focuses primarily on matching.
- In practice, we often use a linear regression: $Y = \beta X + \tau D + \varepsilon$.
- Oster (2019) developed a general method based on Altonji et al. 2005.

Variables	**Model 1**	**Model 2**	**...**	**Full Model**
Treatment D	1.00	1.20	...	1.50
X_1	Excl.	Incl.	...	Incl.
...	...	...	...	...
X_K	...	...	...	Incl.

Table 1: Stability of Treatment Effects across Models

Sensitivity analysis: selection bias

Is the stability of the treatment effect sufficient?

Variables	Null Model	Full Model
Treatment D	1.49	1.50
X_1	Excl.	Incl.
...	...	...
X_K	Excl.	Incl.

Two alternative explanations for stability

1. The bias from confounders is minimal.
2. Control variables are not "controlling anything."

Sensitivity analysis: general selection bias

To examine the "efficacy" of the control variables, we need to check the model fit.

Variables	Null Model	Full Model
D	1.49	1.50
X_1	Excl.	Incl.
...	...	...
X_K	Excl.	Incl.
R^2	10.00%	10.01%

Bad Controls

Variables	Null Model	Full Model
D	1.49	1.50
X_1	Excl.	Incl.
...	...	...
X_K	Excl.	Incl.
R^2	10.00%	90.01%

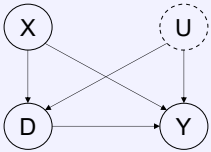
Good Controls

Sensitivity analysis: selection bias

- Observed variables could inform the selection on unobservables.
- We need to consider the **changes in model fit** (e.g., R^2), along with the **changes in coefficients**.

The setup of Oster's sensitivity analysis

$$Y = \alpha + \beta D + \varphi X + \omega U + \varepsilon$$



Both X and U can include multiple variables.

Sensitivity analysis: selection bias

Assumption in Oster's sensitivity analysis

The degree of selection on unobservables can be approximated by the degree of selection on observables using the model fit changes.

$$\underbrace{(\beta_1 - \beta^*)}_{\text{Selection on Unobs.}} = \boldsymbol{\delta} \underbrace{(\beta_0 - \beta_1)}_{\text{Selection on Obs.}} \frac{R^2_{max} - R^2_1}{R^2_1 - R^2_0}$$

- δ is the sensitivity parameter, indicating the relative importance of unobs. over the observables.
- β_0 and R^2_0 are the treatment effect and R-squared of the uncontrolled model $Y \sim D$.
- β_1 and R^2_1 are the treatment effect and R-squared of the controlled model $Y \sim D + X$.

Sensitivity analysis: selection bias

To understand the result, we do a little algebra and assume $\delta = 1$.

$$\beta^* = \beta_1 - (\beta_0 - \beta_1)\frac{R^2_{max} - R^2_1}{R^2_1 - R^2_0}$$

$$\implies \frac{\beta_1 - \beta^*}{\beta_0 - \beta_1} = \frac{R^2_{max} - R^2_1}{R^2_1 - R^2_0}$$

- Under the equal selection assumption, the ratio of the movement in coefficients is equal to the ratio of the movement in R-squared.
- The changes in model fit are informative for the changes in coefficients!

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;
3. Examine the value of δ that would nullify the effect or $\beta^* = 0$.

Sensitivity analysis: selection bias

To apply the result, we can follow these steps:

1. Run a regression of $Y \sim D$ and record β_0 and R_0^2;
2. Run a regression of $Y \sim D + X$ and record β_1 and R_1^2;
3. Examine the value of δ that would nullify the effect or $\beta^* = 0$.
4. Discuss whether the value of δ is reasonable.

Sensitivity analysis: challenges

- **Subjective parameter benchmarks**: Hard to interpret sensitivity parameters without clear, objective guidelines

Sensitivity analysis: challenges

- **Subjective parameter benchmarks**: Hard to interpret sensitivity parameters without clear, objective guidelines
- **Modeling unobserved confounders**: Simulating hidden confounders relies on strong assumptions.

Sensitivity analysis: challenges

- **Subjective parameter benchmarks**: Hard to interpret sensitivity parameters without clear, objective guidelines
- **Modeling unobserved confounders**: Simulating hidden confounders relies on strong assumptions.
- **Mixing power with identification**: Sensitivity analysis can be misleading in very small or very large samples.

1 Understanding assumptions in causal inference

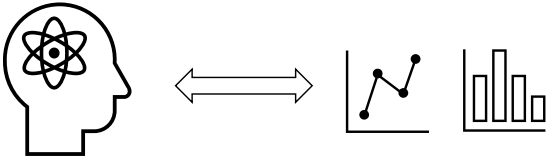
2 Assessing assumptions
Sensitivity analysis
Consistency tests

3 Conclusions

Consistency tests

Definition (Consistency tests)

If unconfoundedness is valid, we should observe patterns that are consistent with it.



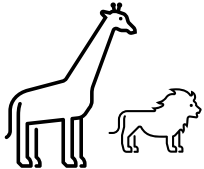
Theoretical Predictions Data Patterns

Consistency tests

Identification assumptions and data as a measurement tool for causal effects.

A Measurement Tool

Identification:
Assumptions + Data



Prior Knowledge:
Giraffe is taller!

Measure of Giraffe
> Measure of Lion

Three types of consistency tests

Using pseudo-outcomes, pseudo-treatments, or alternative groups where researchers have prior knowledge of the treatment effect.

Pseudo Outcomes

Pseudo Treatment

Alternative Groups

Consistency tests: pseudo-outcomes

Pseudo-outcomes (Dube et al. 2013)

Objective: Spillover effects of US gun laws on gun-related crimes in Mexico.

Population: Mexican municipalities close to the U.S. border.

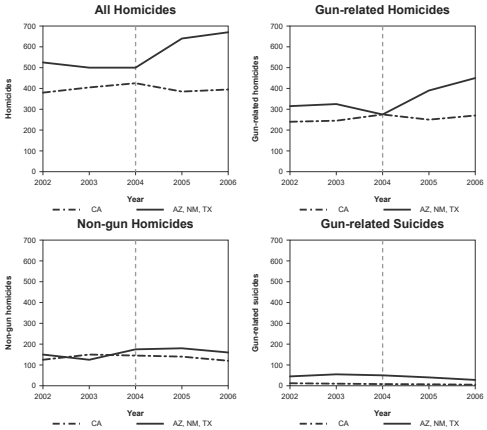
Treatment: Access to assault weapons from neighboring U.S. states.

Outcome: Gun-related homicides.

Pseudo-outcomes: Accidents, non-gun homicides and suicides.

Consistency tests: pseudo-outcomes

Pseudo-outcomes (Dube et al. 2013)



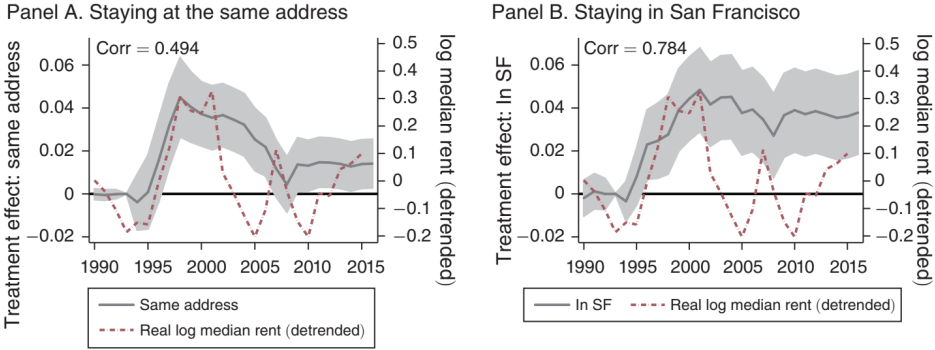
Consistency tests: pseudo-treatment

Pseudo-treatment

- In DID, falsification tests assume that treatment happens in the pre-treatment periods.
- We expect no effect of these pseudo-treatments.
- In RDD, falsification tests assume different cutoffs of the forcing variable.
- We expect no effect from pseudo-treatments based on these cutoffs.

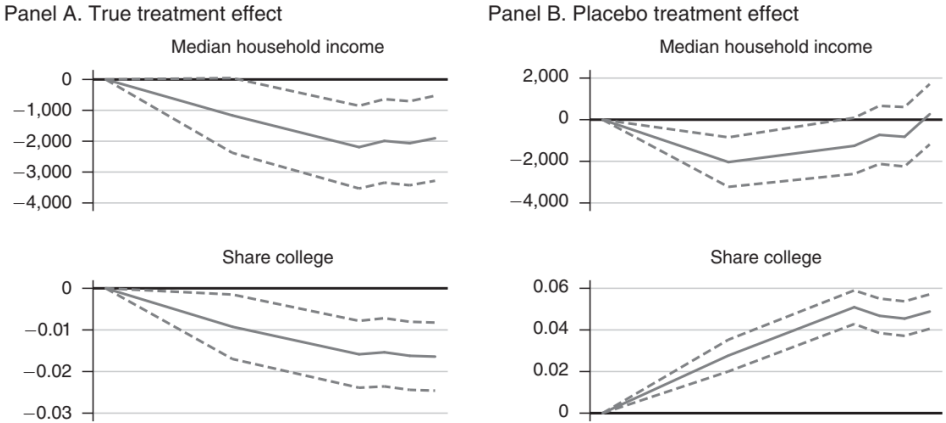
Consistency tests: alternative group

Diamond et al. (2019) compare rent-controlled **buildings before 1980** (treated) to similar **buildings after 1980** (control) to estimate the impact of rent control (D) on tenant mobility and housing supply (Y).

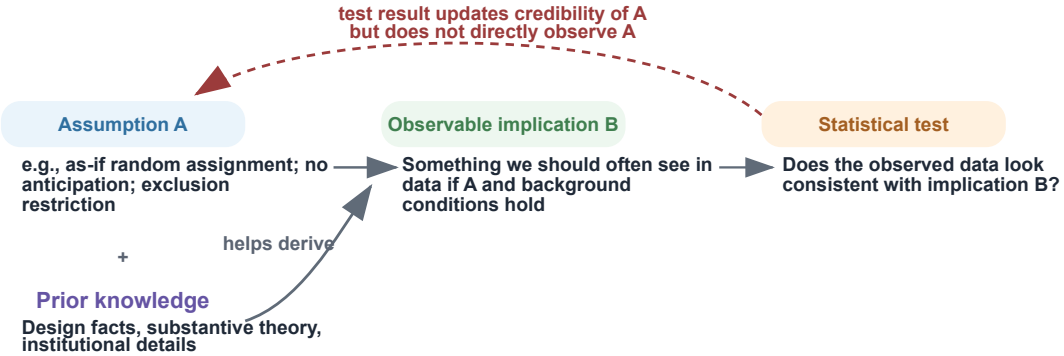


Consistency tests: alternative group

People who move away are used as a "pseudo" treatment group - no treatment effect expected.



Consistency tests test implications, not assumptions



Statistical testing as plausibility reasoning

- $A =$ The person is the burglar.
- $B =$ The broken window, the mask and the loot.

If A is true, then B is true.
B is true.

A becomes more plausible.

Statistical testing as plausibility reasoning

Deductive logic

$$\frac{\begin{array}{c}\text{If } A \text{ is true, then } B \text{ is true.}\\ A \text{ is true}\end{array}}{B \text{ is true}}$$

This is valid logic.

Plausibility reasoning

$$\frac{\begin{array}{c}\text{If } A \text{ is true, then } B \text{ is true.}\\ B \text{ is observed}\end{array}}{A \text{ becomes more plausible}}$$

This is not proof. It is evidence.

Key point

A consistency test usually asks whether an implication B is consistent with the data. It does not "prove" whether assumption A is true.

Examples: exogeneity assumptions and their implications

Test	Assumption A	Observable implication B
Balance test	**A** As-if random assignment	**B** Treated and control similar on covariates.
Pre-treatment placebo	**A** No treatment effect before treatment	**B** Zero estimated effects before treatment.
Placebo group or action	**A** Treatment cannot affect this group / outcome	**B** Zero estimated effect for the group / outcome.

Consistency tests examine the implications of A, but not directly test A.

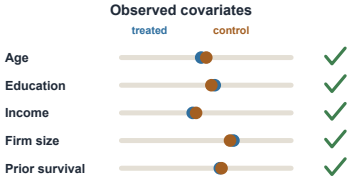
Why does one passed test feel more convincing than another?

The test is more informative when:

1. **Seeing B is hard to explain without A.**
2. **Not seeing B is hard to reconcile with A.**

Example 1: Balance test

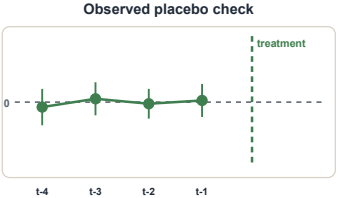
A as-if random assignment → **B** similar before treatment



Still hidden
motivation • ability • networks • access

Example 2: Pre-treatment placebo

A no effect + exogeneity → **B** placebo effects near zero



If pre-treatment effects are large and systematic: serious warning for A

Takeaway: what consistency tests can and cannot do

Consistency tests update credibility; they do not certify assumptions.

If a test passes

- The assumption becomes more plausible but stay non-proven.
- Passing is not sufficient for validity.

If a test fails

- The assumption becomes less plausible, but stay probable.
- Failing is not necessary for invalidity.

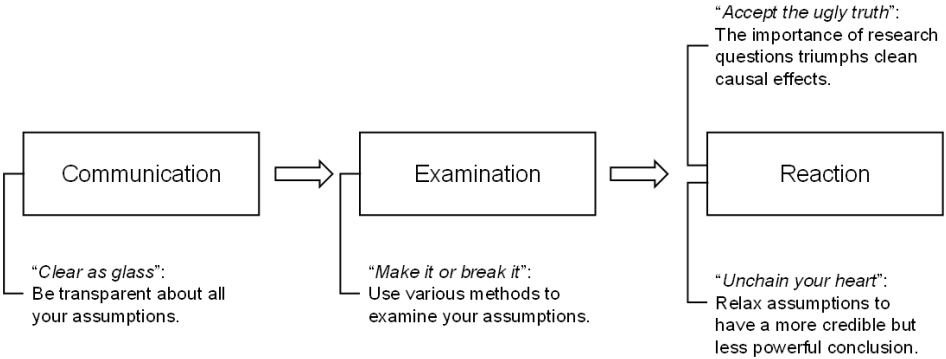
Best practice: combine consistency tests with prior knowledge and design details.

1 Understanding assumptions in causal inference

2 Assessing assumptions

3 Conclusions

Summary: a framework for assumptions



Summary: for your own projects

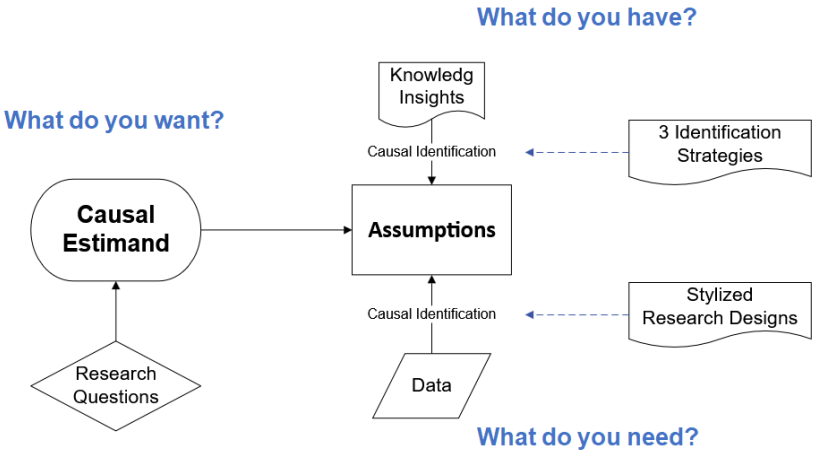
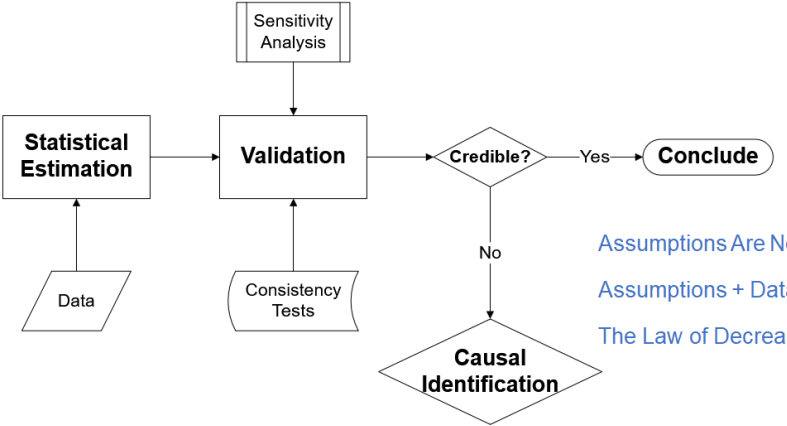


Figure 5: The Identification Stage

Summary: for your own projects



Assumptions Are Not Unique

Assumptions + Data = Conclusions

The Law of Decreasing Credibility

Figure 6: The Estimation Stage

References I

- Dube, A., Dube, O., & García-Ponce, O. (2013). Cross-border spillover: US gun laws and violence in Mexico. American Political Science Review, 107(3), 397-417.
- Diamond, R., McQuade, T., & Qian, F. (2019). The effects of rent control expansion on tenants, landlords, and inequality: Evidence from San Francisco. American Economic Review, 109(9), 3365-3394.
- Rosenbaum, P. R. (1987). Model-based direct adjustment. Journal of the American Statistical Association, 82(398), 387-394.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business & Economic Statistics, 37(2), 187-204.

References II

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. Journal of Political Economy, 113(1), 151-184.