ABIZ Seminar Series

Causal Inference for Business Research

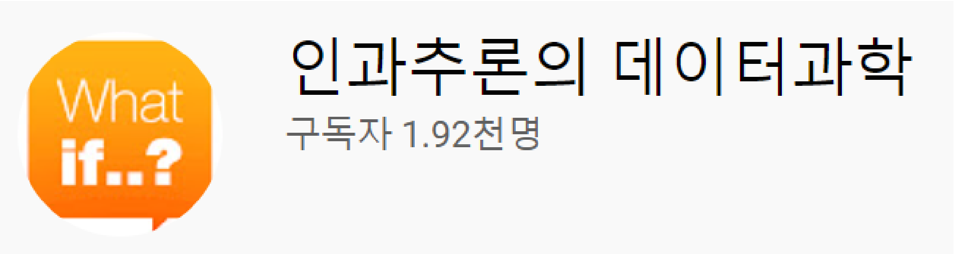
Jiyong Park

Bryan School of Business and Economics

University of North Carolina at Greensboro

Background: Causal Inference Workshops

- Korea Summer Workshop on Causal Inference 2022 (2022, Online)
- Korea Summer Session on Causal Inference 2021 (2021, Online)
- Social Science of COVID-19: From the Perspective of Causal Inference (2020, Online)
- MIS Summer Session 2019 - Experimental Empirical Methods (2019, KAIST)
- MIS Summer Session 2018 - Research Design for Data Analytics (2018, KAIST)
- MIS Summer Session 2017 - Introduction to Economics of IS and Research Methodology (2017, KAIST)



Credibility Revolution in Empirical Research

What is a Cause-and-Effect?

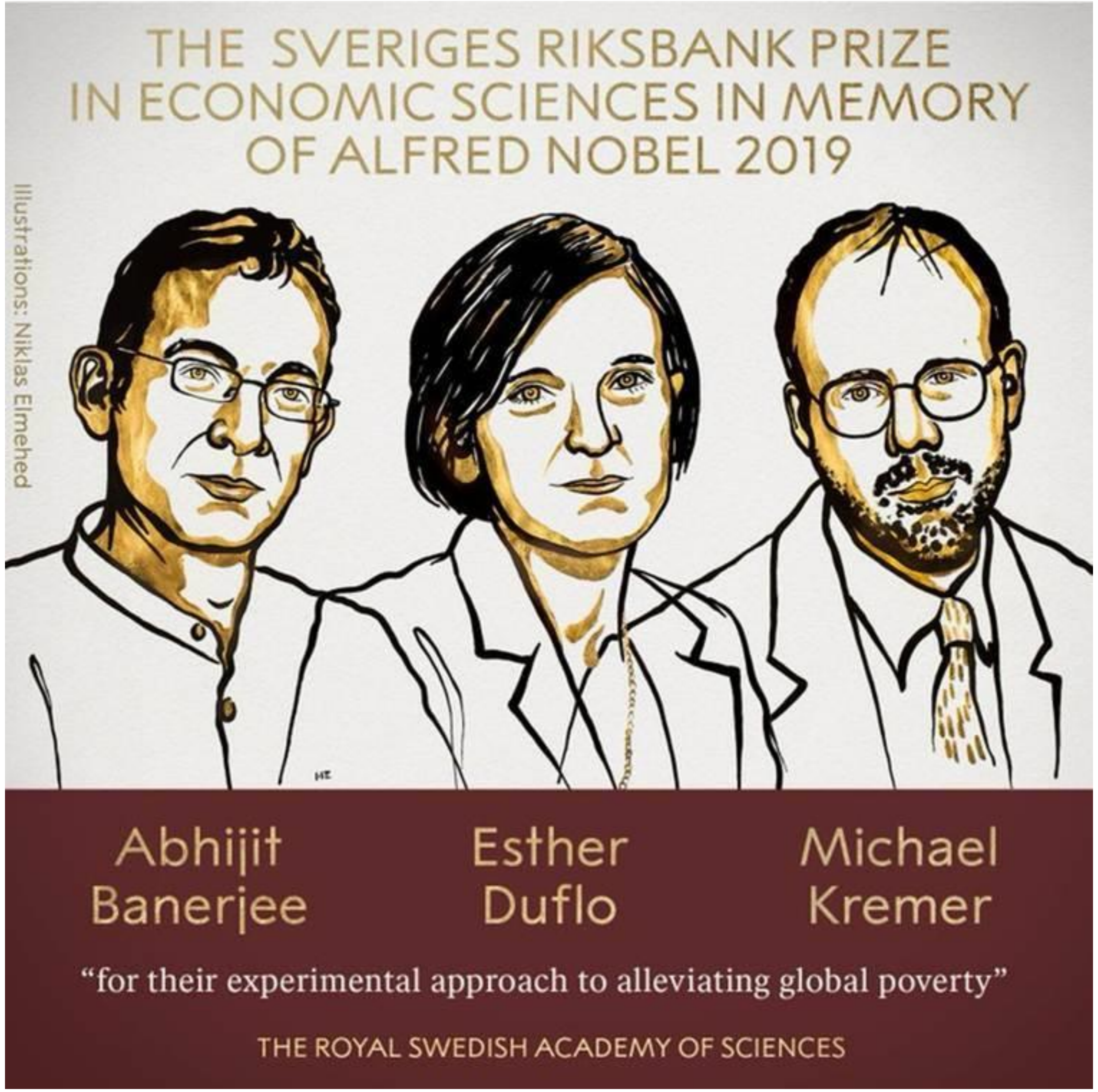
What is a Cause-and-Effect?

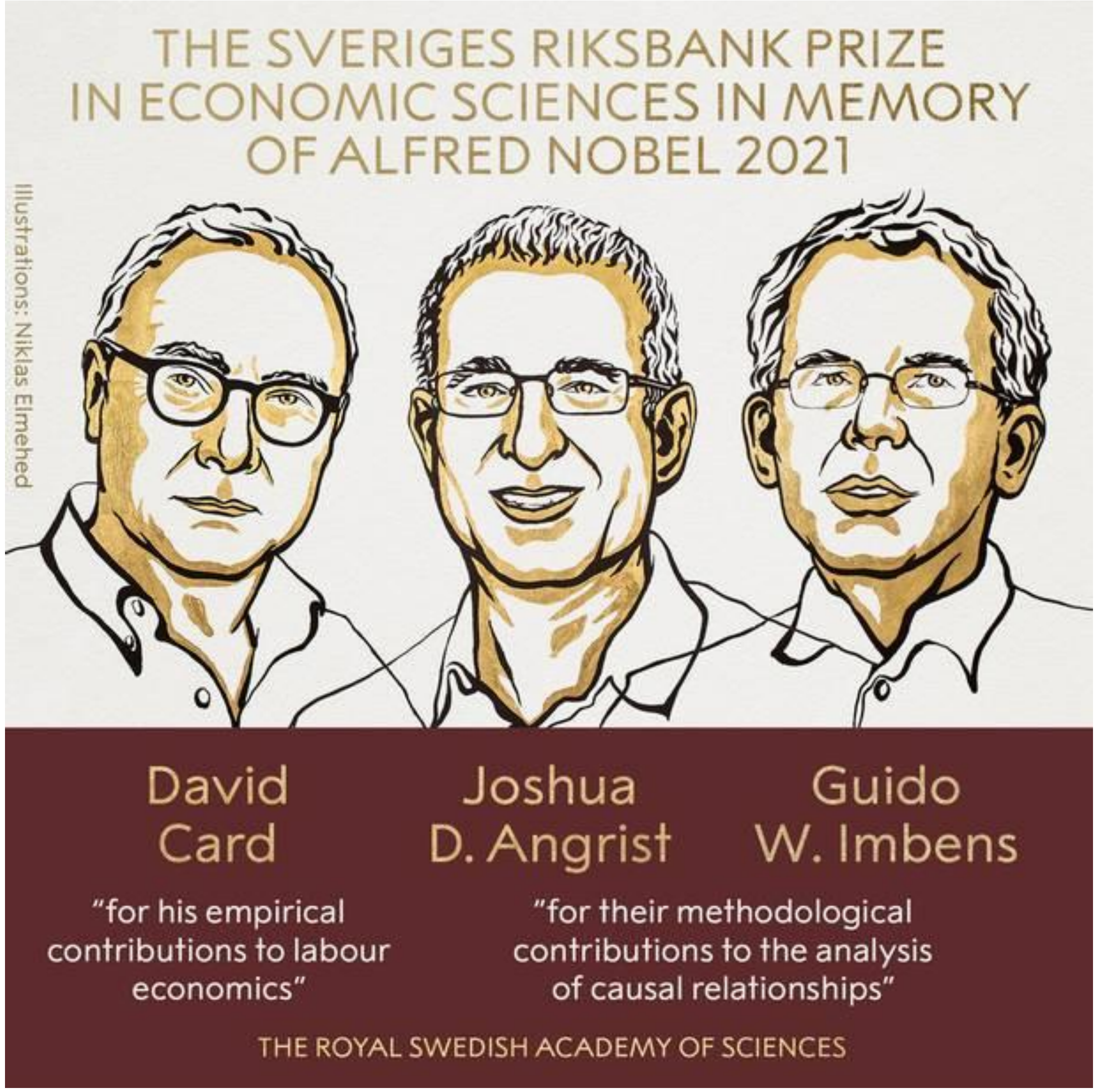
A relationship in which one event (the cause) makes another event happen (the effect)

What is Causal Inference?

The process of identifying a causal effect of a particular phenomenon from data

Nobel Prizes in Economics Recognizing Causal Inference





Credibility Revolution

Free exchange

The Nobel prize in economics celebrates an empirical revolution

David Card shares this year's award with Joshua Angrist and Guido Imbens

https://www.economist.com/finance-and-economics/2021/10/12/the-nobel-prize-in-economics-celebrates-an-empirical-revolution

SOCIAL SCIENCE SPACE

2021 Nobels in Economics a Victory in the Credibility Revolution

Published on 10/12/2021 By **David A. Jaeger**

The Nobel committee's decision to award its economics prize for 2021 to David Card, Josh Angrist and Guido Imbens marks the culmination of a revolution in the way economists approach the world that began more than 30 years ago. Until the 1980s, experiments were uncommon in economics. Most economists who worked on the applied side of the field relied on data from surveys (like the census) or administrative sources (like social security).

https://www.socialsciencespace.com/2021/10/2021-nobels-in-economics-a-victory-in-the-credibility-revolution

“Taking the Con out of Econometrics”

Let’s Take the Con out of Econometrics

By Edward E. Leamer*

Econometricians would like to project the image of agricultural experimenters who divide a farm into a set of smaller plots of land and who select randomly the level of fertilizer to be used on each plot. If some plots are …

… the annual meeting of the American Ecological Association, another farmer in the audience objects that he used the same data but came up with the conclusion that moderate

One should not jump to the conclusion that there is necessarily a substantive difference between drawing inferences from experimental as opposed to nonexperimental data. The images I have drawn are de- …

… What is the real difference between these two settings? Randomization seems to be the answer. In the experimental setting, the fertilizer treatment is “randomly” assigned

“The econometric art as it is practiced at the computer terminal involves fitting many, perhaps thousands, of statistical models. One or several that the researcher finds pleasing are selected for reporting purposes.” (p. 36)

“Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else’s data analysis seriously.” (p. 37)

Leamer, E.E., 1983. Let’s Take the Con out of Econometrics. *American Economic Review*, *73*(1), pp.31-43.
Angrist, J.D. and Pischke, J.S., 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, *24*(2), pp.3-30.

"Taking the Con out of Econometrics"

Let's Take the Con out of Econometrics

By EDWARD E. LEAMER*

Econometricians would like to project the image of agricultural experimenters who divide a farm into a set of smaller plots of land and who select randomly the level of fertilizer to be used on each plot. If some plots are

One should not jump to the conclusion that there is necessarily a substantive difference between drawing inferences from experimental as opposed to nonexperimental data. The images I have drawn are de-

"The econometric art as it is practiced at the computer terminal involves fitting many, perhaps thousands, of statistical models. One or several that the researcher finds pleasing are selected for reporting purposes." (p. 36)

"Hardly anyone takes data analysis seriously. Or perhaps more accurately, hardly anyone takes anyone else's data analysis seriously." (p. 37)

the annual meeting of the American Ecological Association, another farmer in the audience objects that he used the same data but came up with the conclusion that moderate

What is the real difference between these two settings? Randomization seems to be the answer. In the experimental setting, the fertilizer treatment is "randomly" assigned

About 3 decades

The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics

Joshua D. Angrist and Jörn-Steffen Pischke

"But, we're happy to report that Leamer's complaint that "hardly anyone takes anyone else's data analysis seriously" no longer seems justified." (p. 3)

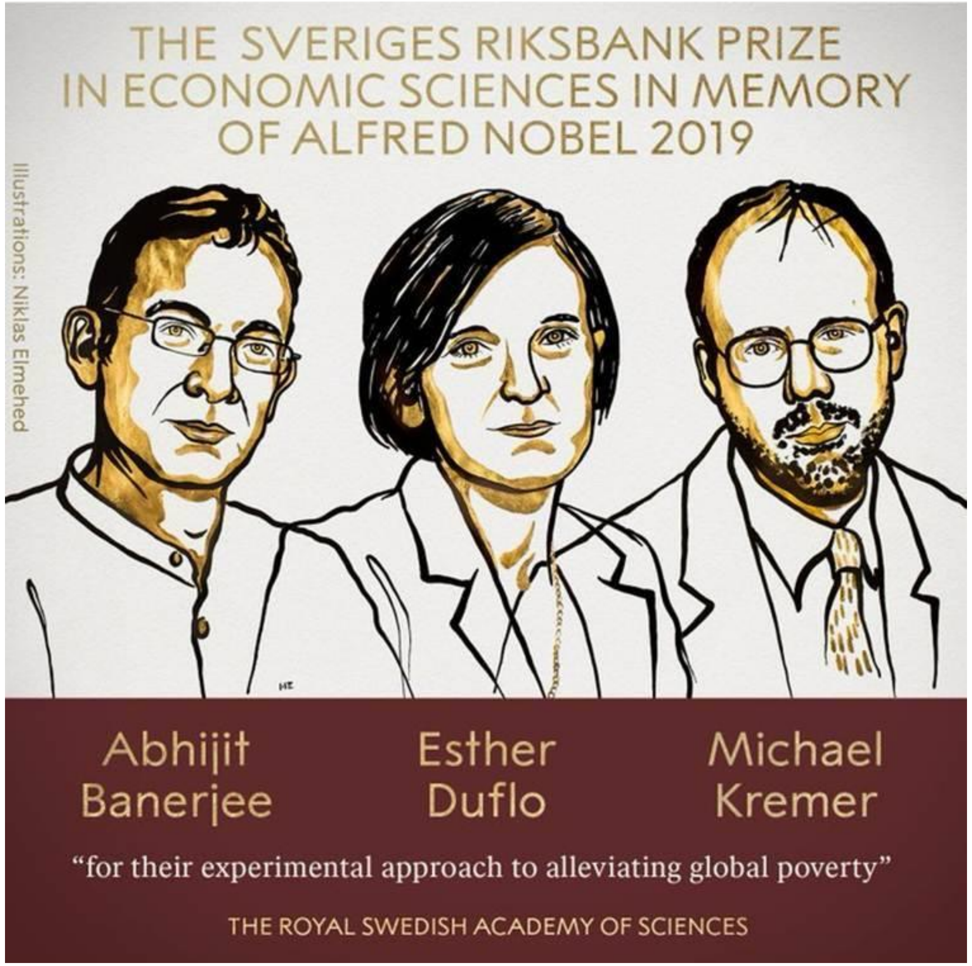
"**A clear-eyed focus on research design is at the heart of the credibility revolution in empirical economics.**" (p. 6)

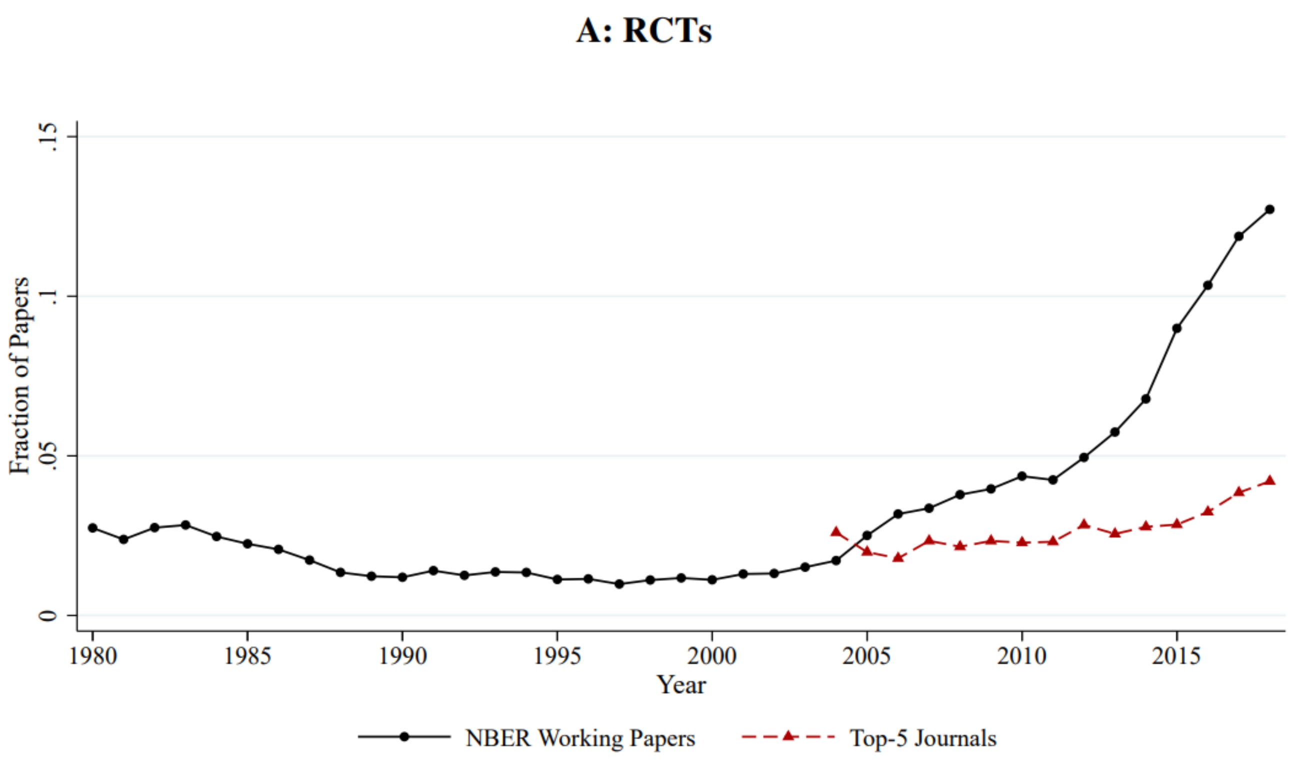
from a distressing lack of robustness to changes in key assumptions—assumptions he called "whimsical" because one seemed as good as another. The remedy he proposed was sensitivity analysis, in which researchers show how their results vary with changes in specification or functional form. Leamer's critique had a refreshing emperor's-new-clothes earthiness that we savored on first reading and still enjoy today. But we're happy to report that Leamer's complaint that "hardly anyone takes anyone else's data analysis seriously" no longer seems justified.

Leamer, E.E., 1983. Let's Take the Con out of Econometrics. *American Economic Review*, *73*(1), pp.31-43.
Angrist, J.D. and Pischke, J.S., 2010. The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, *24*(2), pp.3-30.

Methodological Advances: Experimental Approaches

- Randomized Controlled Trial (RCT)
  - Gold standard for causal inference



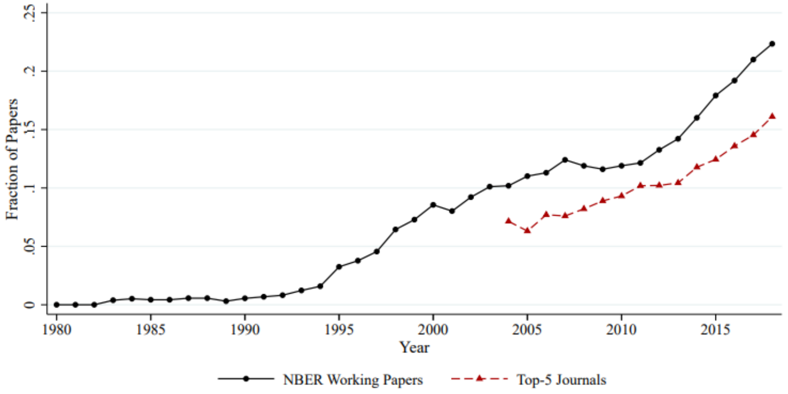


Currie, J., Kleven, H. and Zwiers, E., 2020, May. Technology and Big Data are Changing Economics: Mining Text to Track Methods. In *AEA Papers and Proceedings* (Vol. 110, pp. 42-48).

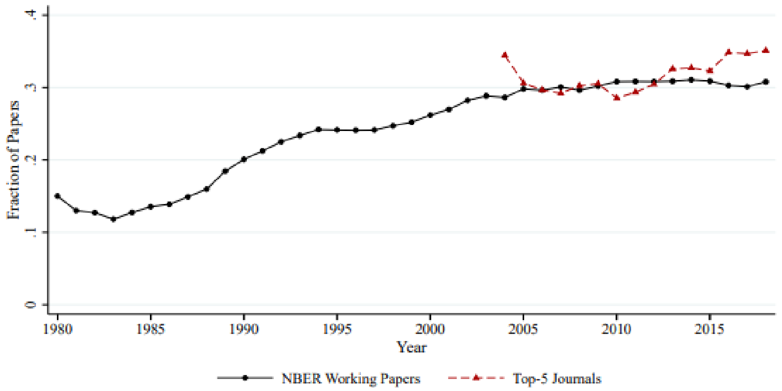
Methodological Advances: Experimental Approaches

- (Natural) Quasi-Experiments

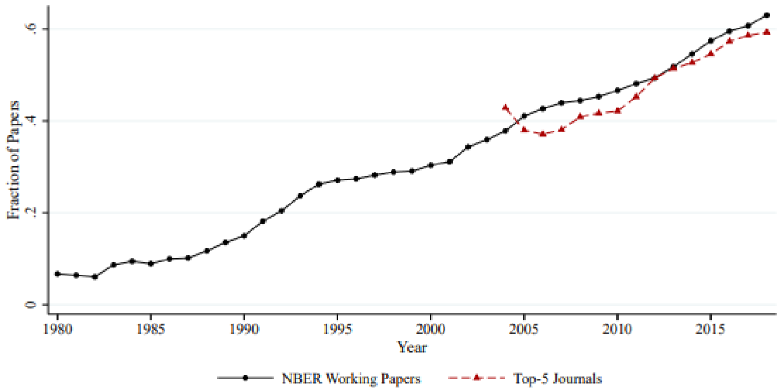
A: Difference-in-Differences



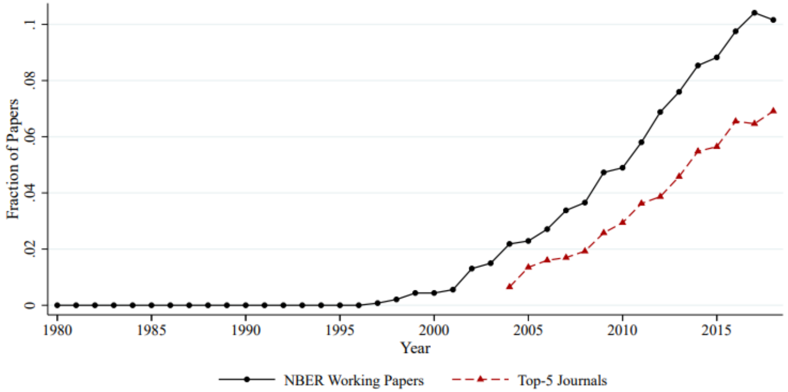
A: Instrumental Variables



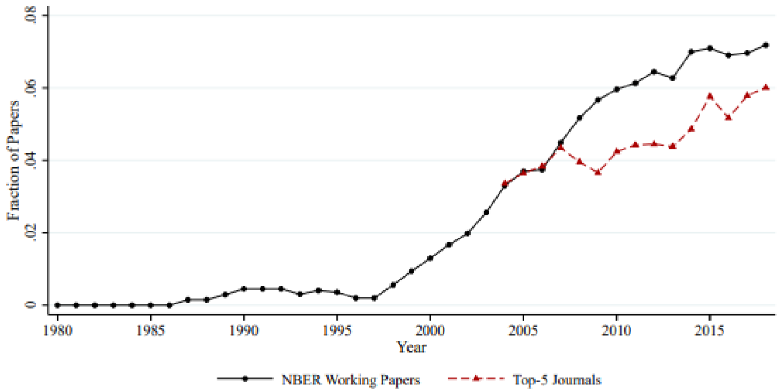
B: Fixed Effects



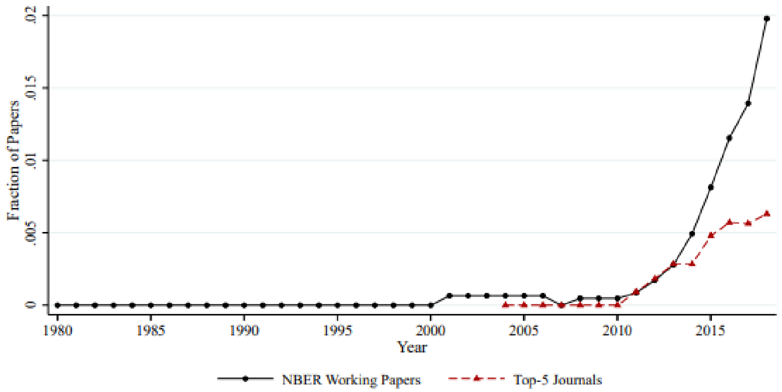
B: Regression Discontinuity



C: Matching

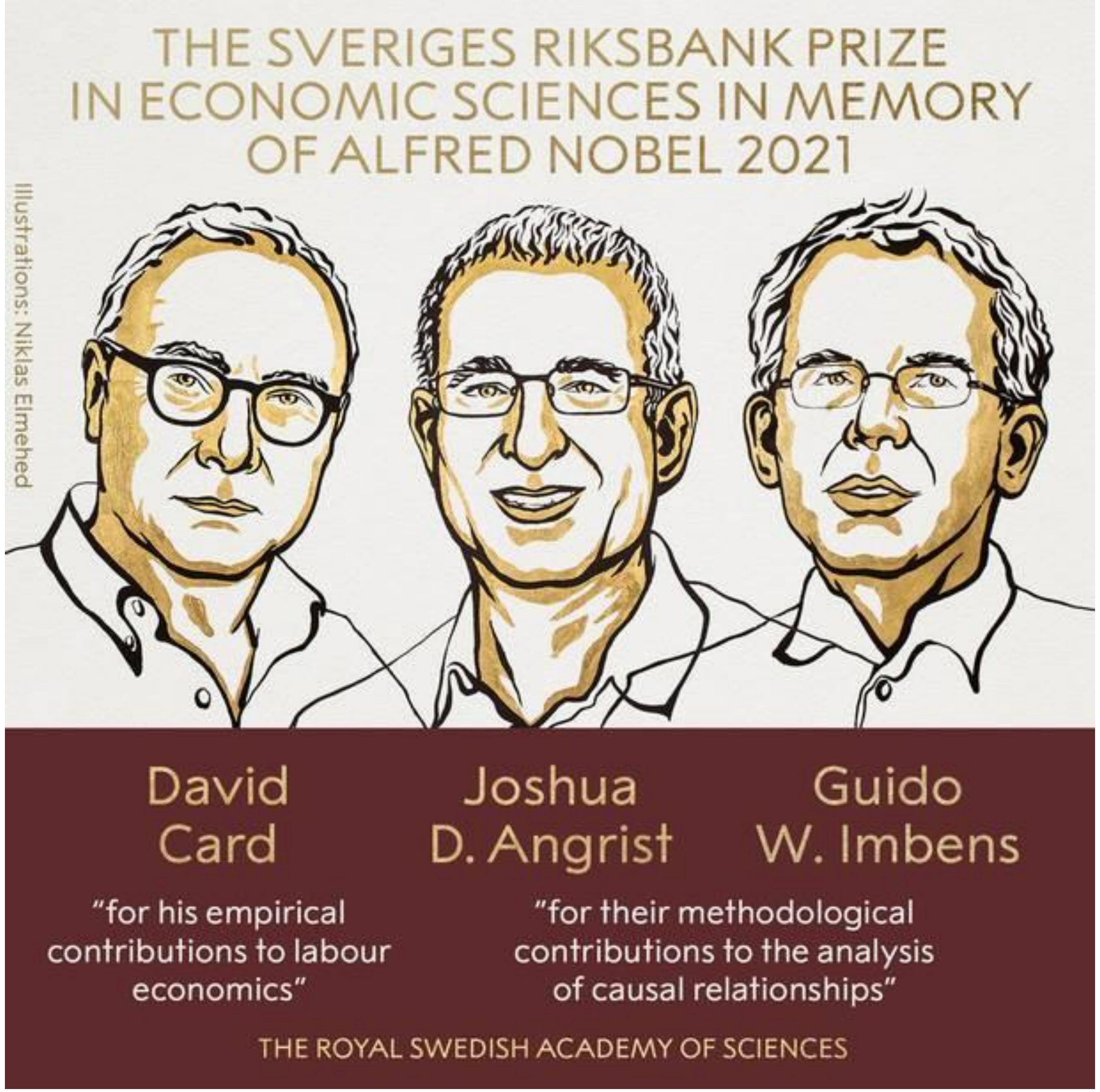


D: Synthetic Control



Currie, J., Kleven, H. and Zwiers, E., 2020, May. Technology and Big Data are Changing Economics: Mining Text to Track Methods. In *AEA Papers and Proceedings* (Vol. 110, pp. 42-48).

Methodological Advances: Experimental Approaches



Nobel Prize in economics goes to 'natural experiments' pioneers

US economists David Card, Joshua Angrist and Guido Imbens win the 2021 Nobel Prize in economics.

Natural experiments use real-life situations to work ou
Reuters]

https://www.aljazeera.com/news/2021/10/11/nobel-prize-in-economics-goes-to-natural-experiments-pioneers

KUNGL. VETENSKAPS-AKADEMIEN

THE ROYAL SWEDISH ACADEMY OF SCIENCES

11 OCTOBER 2021

THE ROYAL SWEDISH ACADEMY OF SCIENCES, founded in 1739, is an independent organisation whose overall objective is to promote the sciences and strengthen their influence in society. The Academy takes special responsibility for the natural sciences and mathematics, but endeavours to promote the exchange of ideas between various disciplines.

BOX 50005 [LILLA FRESCATIVÄGEN 4 A], SE-104 05 STOCKHOLM, SWEDEN
TEL +46 8 673 95 00, KVA@KVA.SE • WWW.KVA.SE

Answering Causal Questions Using Observational Data

Most applied science is concerned with uncovering causal relationships. In many fields, randomized controlled trials (RCTs) are considered the gold standard for achieving this. The systematic use of RCTs to study causal relationships — assessing the efficacy of a medical treatment for example — has resulted in tremendous welfare gains in society. However, due to financial, ethical, or practical constraints, many important questions — particularly in the social sciences — cannot be studied using a controlled randomized experiment. For example, what is the impact of school closures on student learning and the spread of the COVID-19 virus? What is the

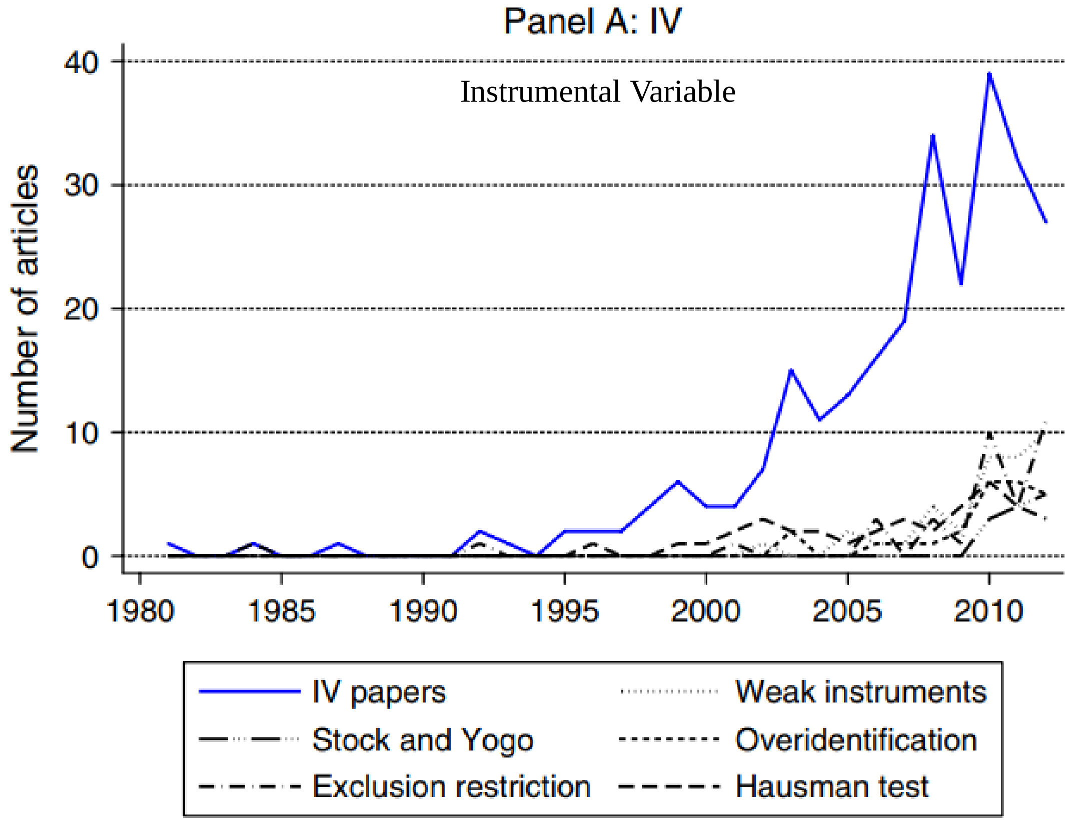
to uncover causal effects in important domains. This early work thus played a crucial role in shifting the focus in empirical research using observational data towards relying on quasi-experimental variation to establish causal effects. The framework developed by Angrist and Imbens, in turn, significantly altered how researchers approach empirical questions using data generated from either natural experiments or randomized experiments with incomplete compliance to the assigned treatment. At the core, the LATE interpretation clarifies what can and

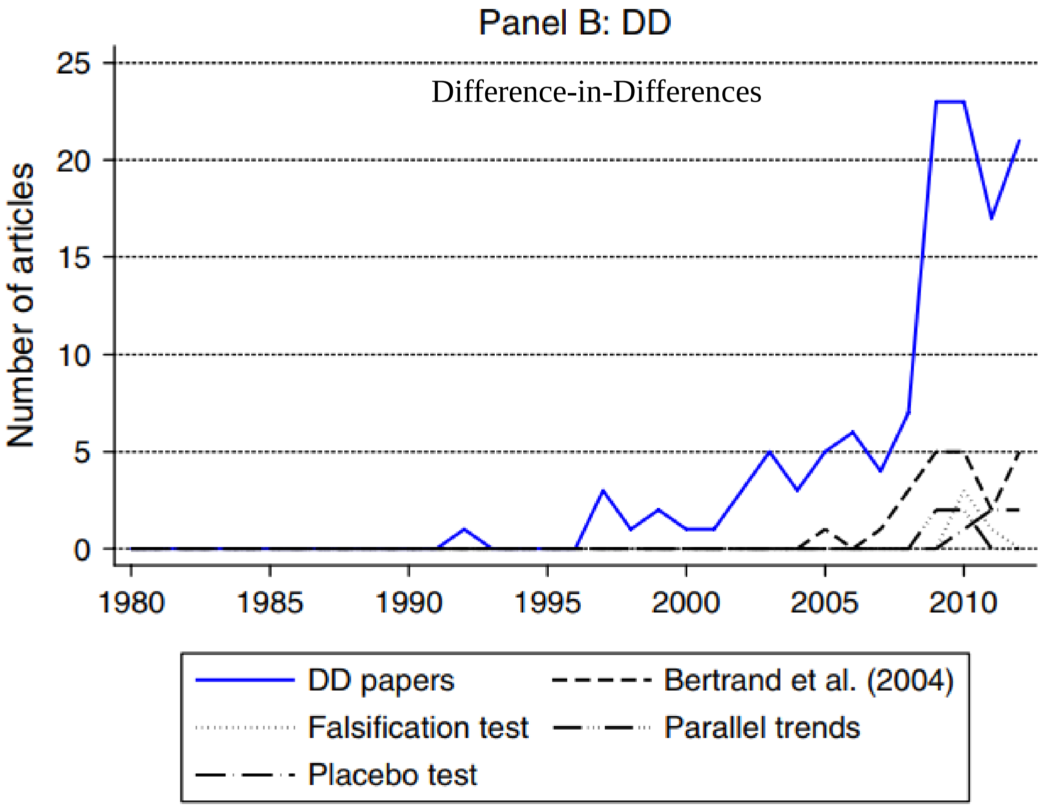
1

"Taken together, therefore, the Laureates' contributions have played a central role in establishing the so-called **design-based approach** in economics. This approach – aimed at **emulating a randomized experiment to answer a causal question using observational data** – has transformed applied work and improved researchers' ability to answer causal questions of great importance for economic and social policy using observational data." (P. 2)

https://www.nobelprize.org/uploads/2021/10/advanced-economicssciencesprize2021.pdf

Credibility Revolution is Ongoing in Business Research





* Among top finance journals (Journal of Finance, Journal of Financial Economics, Review of Financial Studies)

Bowen III, D.E., Frésard, L. and Taillard, J.P., 2016. What's Your Identification Strategy? Innovation in Corporate Finance Research. *Management Science*, 63(8), pp.2529-2548.

Credibility Revolution is Ongoing in Business Research

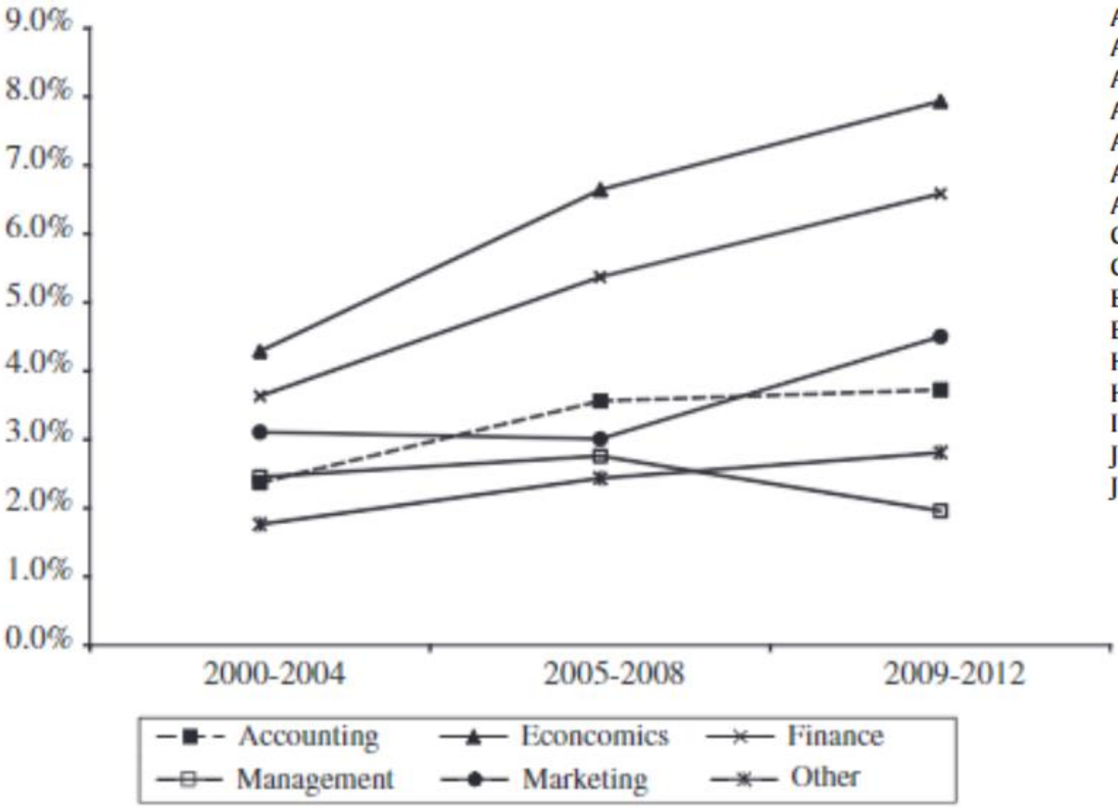


Fig. 1. Relative frequency of causal studies by field and publication period.

Journal title	Field
Academy of Management Journal	Management
Academy of Management Perspectives	Management
Academy of Management Review	Management
Accounting Organizations and Society	Accounting
Accounting Review	Accounting
Administrative Science Quarterly	Management
American Economic Review	Economics
California Management Review	Management
Contemporary Accounting Research	Accounting
Econometrica	Economics
Entrepreneurship Theory and Practice	Other
Harvard Business Review	Management
Human Resource Management	Other
Information Systems Research	Other
Journal of Accounting Economics	Accounting
Journal of Accounting Research	Accounting

Journal title	Field
Journal of Applied Psychology	Other
Journal of Business Ethics	Other
Journal of Business Venturing	Other
Journal of Consumer Psychology	Marketing
Journal of Consumer Research	Marketing
Journal of Finance	Finance
Journal of Financial and Quantitative Analysis	Finance
Journal of Financial Economics	Finance
Journal of International Business Studies	Other
Journal of Management Studies	Management
Journal of Marketing	Marketing
Journal of Marketing Research	Marketing
Journal of Operations Management	Other
Journal of Political Economy	Economics
Management Science	Other
Marketing Science	Marketing
MIS Quarterly	Other
Operations Research	Other
Organization Science	Other
Organization Studies	Other
Organizational Behavior and Human Decision Processes	Other
Quarterly Journal of Economics	Economics
RAND Journal of Economics	Economics
Review of Accounting Studies	Accounting
Review of Financial Studies	Finance
Strategic Management Journal	Management

Gassen, J., 2014. Causal Inference in Empirical Archival Financial Accounting Research. *Accounting, Organizations and Society*, *39*(7), pp.535-544.

Overview of Causal Inference for Business Research

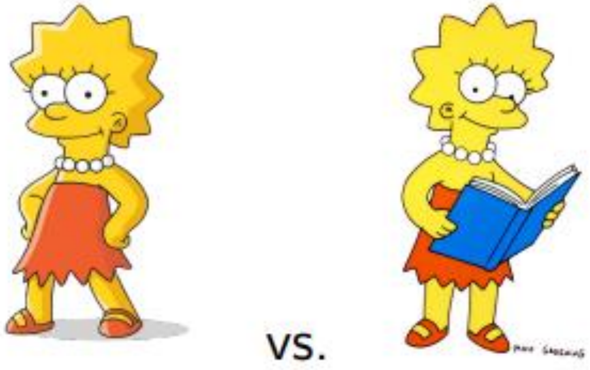
Potential Outcome Framework

아주대학교 AJOU UNIVERSITY UNC GREENSBORO

- Causation is defined as the difference in potential outcomes after the treatment.
  - *"What if the treatment was not applied?"*
  - Causal effect = (Actual outcome for treated if treated) – (Potential outcome for treated if not treated)

Counterfactual

Causal effect of reading on grades



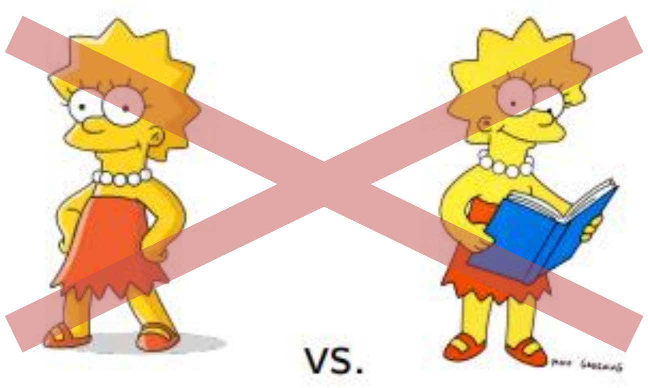
Counterfactual

Fundamental Problem of Causal Inference

- In reality, we do not observe both potential outcomes; we only observe one.
  - Causal effect = (Actual outcome for treated if treated) – $\boxed{\text{(Potential outcome for treated if not treated)}}$ **Counterfactual**
  - But, we can only observe:

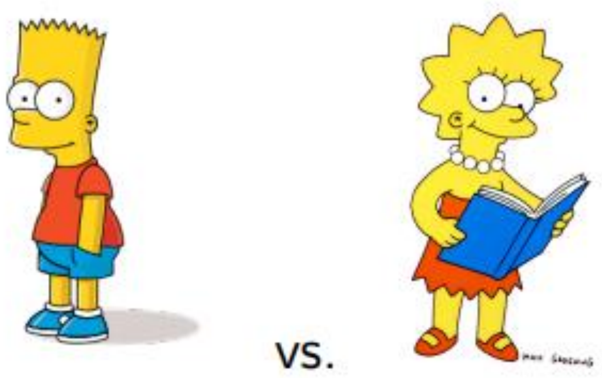
    = (Actual outcome for treated if treated) – $\boxed{\text{(Actual outcome for untreated if not treated)}}$ **Control Group**

Ideal Comparison



Counterfactual

Actual Comparison



Control Group

Fundamental Problem of Causal Inference

- In reality, we do not observe both potential outcomes; we only observe one.
  - Causal effect = (Actual outcome for treated if treated) – (Potential outcome for treated if not treated)
    Counterfactual
  - But, we can only observe:
    = (Actual outcome for treated if treated) – (Actual outcome for untreated if not treated)
    Control Group

Individual treatment effect (ITE) cannot be identified by definition.

	Treatment	Potential Outcomes		Causal Effect
Subject i	T_i	$Y_i(1)$	$Y_i(0)$	
1	1	3		ATE on the Treated (ATET)
2	1	1		
3	0		1	ATE on the Untreated (ATEU)
4	0		1	

Main Focus: Average Treatment Effect (ATE)

Fundamental Problem of Causal Inference

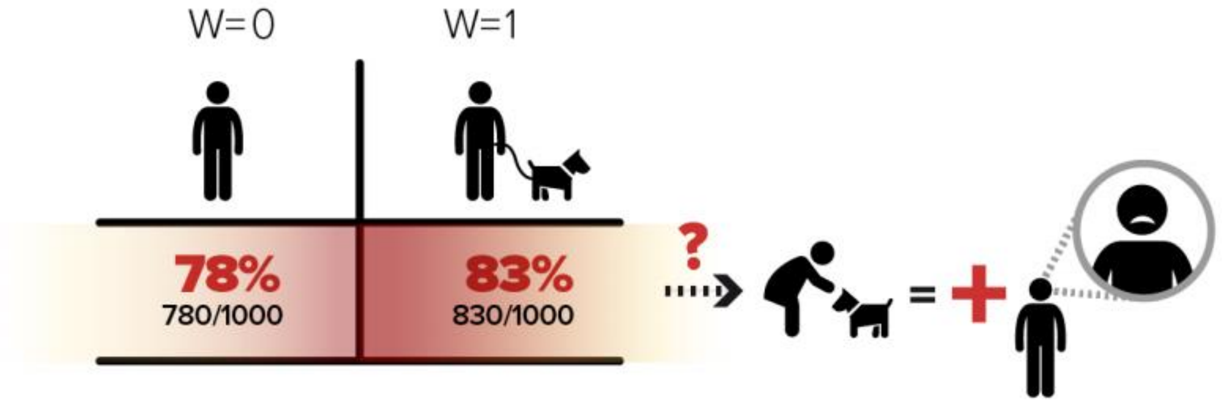
- In reality, we do not observe both potential outcomes; we only observe one.
  - Causal effect = (Actual outcome for treated if treated) – (Potential outcome for treated if not treated)

    Counterfactual
  - But, we can only observe:

    = (Actual outcome for treated if treated) – (Actual outcome for untreated if not treated)

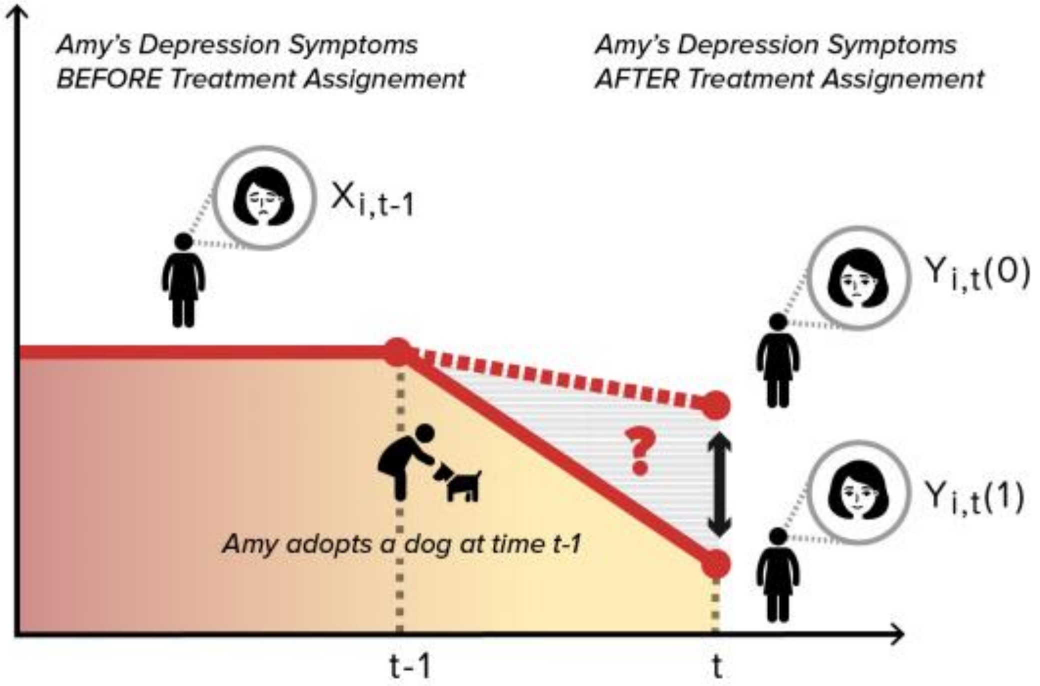
    Control Group

Actual Comparison



Dominici, F., Bargagli-Stoffi, F.J. and Mealli, F., 2020. From controlled to undisciplined data: estimating causal effects in the era of data science using a potential outcome framework. *arXiv preprint arXiv:2012.06865*.

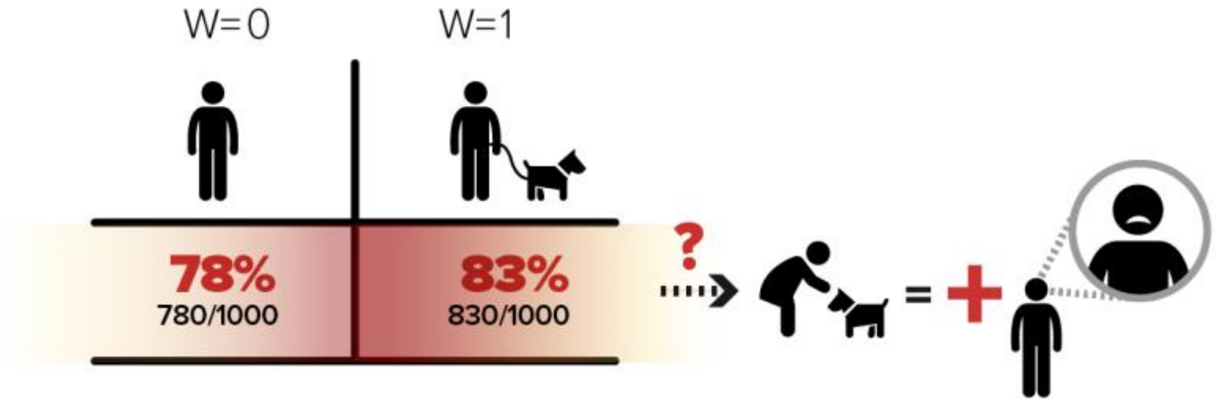
Ideal Comparison



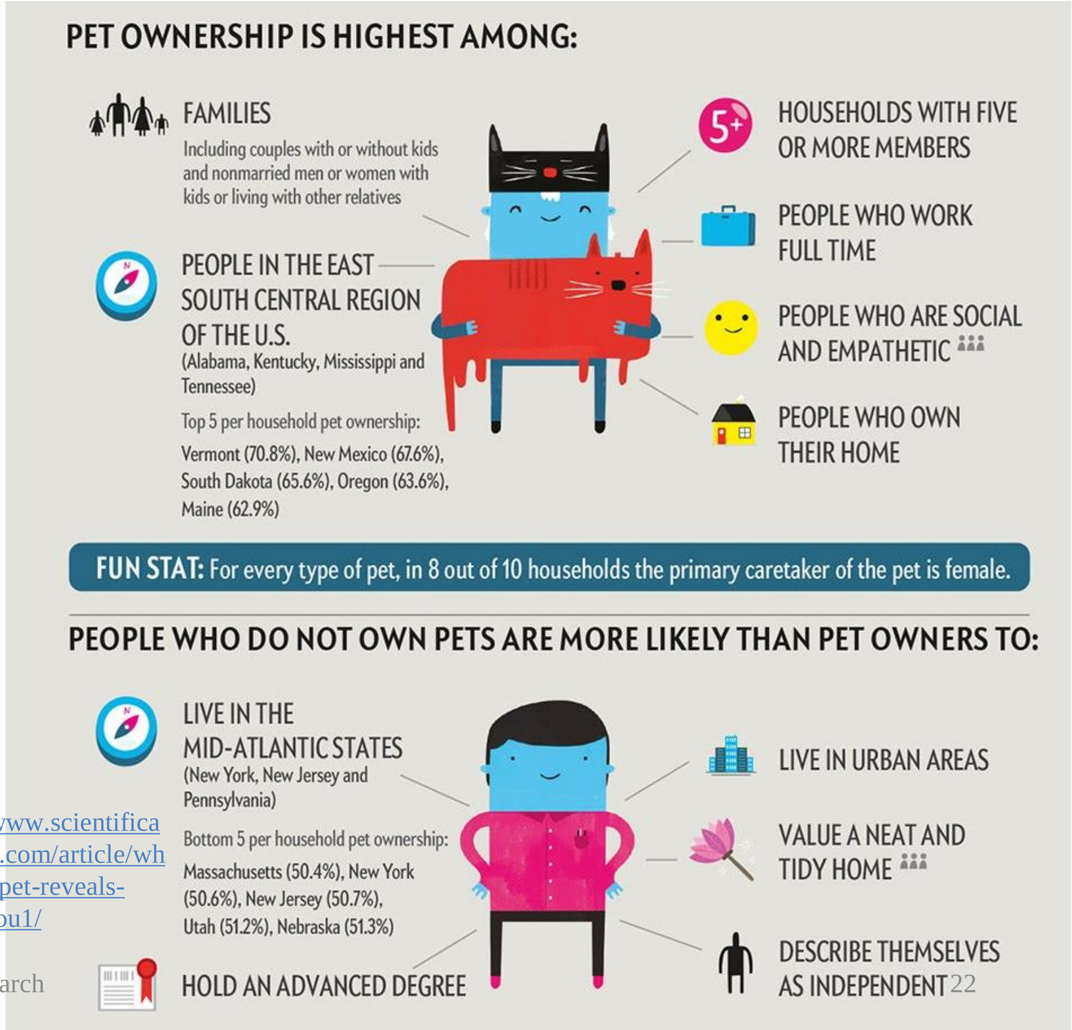
Selection Bias

- In the absence of experiments, treatments are not assigned randomly. Individuals select into the treatment.
- Treatment and control groups may be systematically different and not be comparable even without treatment.

Are they comparable,
except the adoption W?



Dominici, F., Bargagli-Stoffi, F.J. and Mealli, F., 2020. From controlled to undisciplined data: estimating causal effects in the era of data science using a potential outcome framework. *arXiv preprint arXiv:2012.06865*.



https://www.scientificamerican.com/article/what-your-pet-reveals-about-you1/

Selection Bias

- Selection bias is the systematic difference between the treatment group in the absence of treatment (i.e., counterfactual) and the control group.
  - Any differences between the two groups except the treatment are called confounders or confounding factors.

- Decomposition of causal effect and selection bias
  - **Observed effect of the treatment** = (Outcome for treated if treated) – (Outcome for untreated if not treated)

    = (Outcome for treated if treated)

    – (Outcome for treated if not treated) + (Outcome for treated if not treated)

    – (Outcome for untreated if not treated)

    Causal effect

    = (Outcome for treated if treated) – (Outcome for treated if not treated)

    \+ (Outcome for treated if not treated) – (Outcome for untreated if not treated)

    Selection bias

    = **Causal effect + Selection bias**

Ceteris Paribus

- From the perspective of potential outcomes, causal inference is to remove the selection bias.

= (Outcome for treated if not treated) – (Outcome for untreated if not treated)

- How? The treatment group should be comparable to the control (untreated) group in the absence of treatment.

Counterfactual

Treatment Group w/o Treatment (Counterfactual)

Control Group

If the counterfactual is comparable to the control group,

	Treatment	Potential Outcomes	
Subject *i*	T_i	$Y_i(1)$	$Y_i(0)$
1	1	3	1
2	1	1	1
3	0	2	1
4	0	2	1

Ignorability ≃ Exchangeability ≃ Unconfoundedness ≃ Exogeneity

Ceteris Paribus

- From the perspective of potential outcomes, causal inference is to remove the <u>selection bias</u>.

= *(Outcome for treated if not treated) – (Outcome for untreated if not treated)*

- How? The treatment group should be comparable to the control (untreated) group in the absence of treatment.

Counterfactual

Causal inference can be achieved by taking advantage of **research designs** in which control groups are comparable to the treatment group in all aspects, on average, but the fact that they are treated.

Treatment Group w/o Treatment (Counterfactual)	**Control Group**	3	0	2	1
		4	0	2	1

Ignorability ≃ Exchangeability ≃ Unconfoundedness ≃ Exogeneity

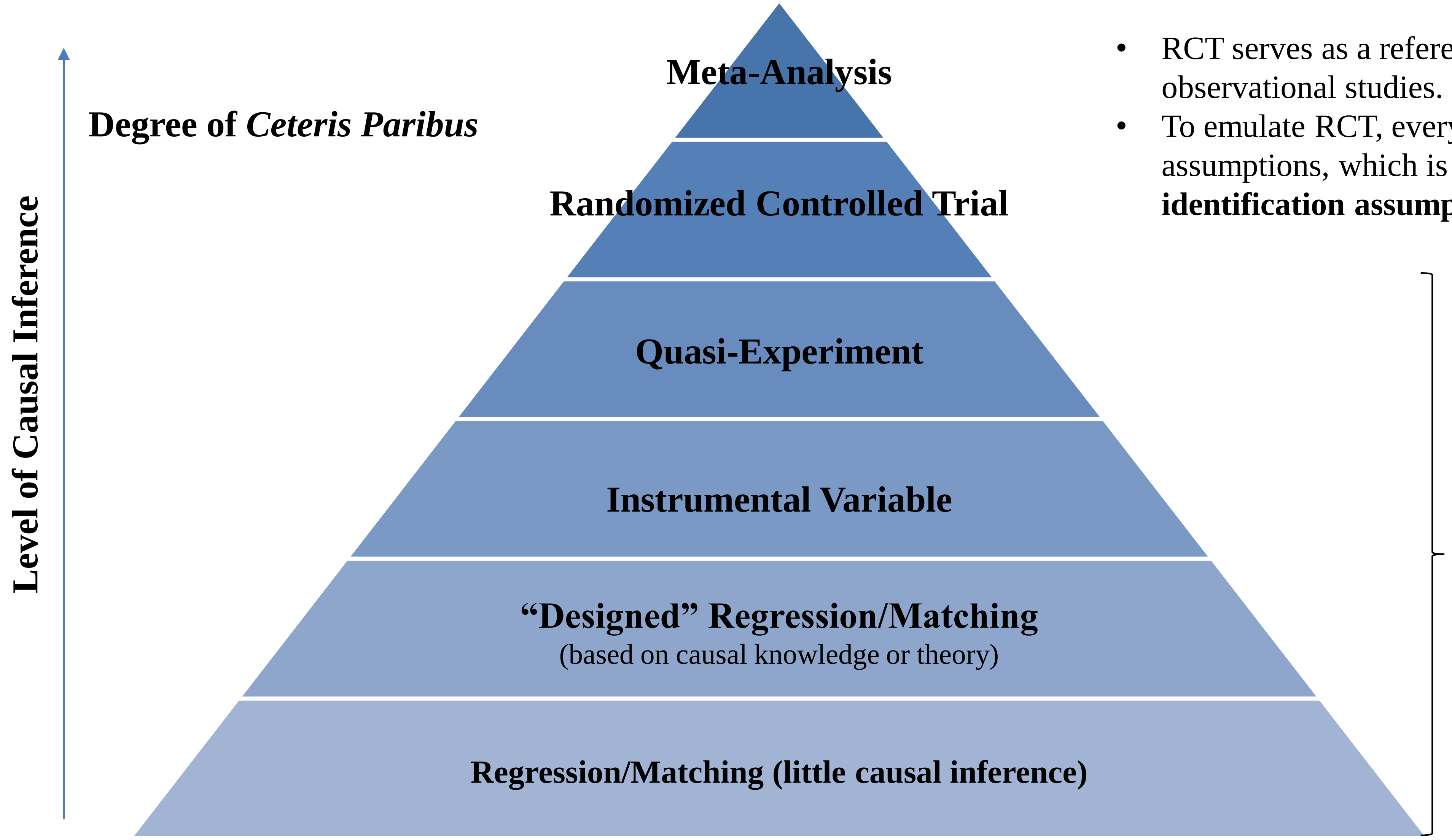
Research Design *Really* Matters

“We call this framework the **design-based approach** to econometrics because the skills and strategies required to use it successfully are related to research design.” (Angrist and Pischke 2017, p. 126)

“The analyst is successful at identifying the causal effect not because of the complex statistical methods that are applied to the data, but **due to the effort in developing a design before data are collected**.” (Keele 2015, p. 331)

Angrist, J.D. and Pischke, J.S., 2017. Undergraduate econometrics instruction: through our classes, darkly. Journal of Economic Perspectives, 31(2), pp.125-144.
Keele, L., 2015. The Statistics of Causal Inference: A View from Political Methodology. *Political Analysis*, 23(3), pp.313-335.

Causal Hierarchy from the Potential Outcome Perspective



Model-Free Descriptive Statistics (no causal inference)

- RCT serves as a reference in observational studies.
- To emulate RCT, every method needs assumptions, which is called **identification assumption**.

The level of causal inference is about the level of difficulty in validating identification assumptions.

Steps for Causal Inference

Research Question | Identification | Estimation

- ✓ Research questions that aim at intervention and manipulation of causes, such as how an outcome would respond to a contemplated change in a cause.

"*Without treatment definitions that specify actions to be performed on experimental units, we cannot unambiguously define causal effects of treatments.*" (Rubin 1978, p. 39)

"*No Causation without Manipulation*" (Holland 1986, p. 959)

Rubin, D.B., 1978. Bayesian inference for causal effects: The role of randomization. *The Annals of Statistics*, pp.34-58.
Holland, P.W., 1986. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396), pp.945-960.

Steps for Causal Inference

Research Question → **Identification** → Estimation

✓ Conditions that permit to estimate a causal effect from observed data

Selection on Observables
→ Theories about outcome and its confounders

Selection on Unobservables
→ Research designs (where data generation process that guarantees *Ceteris Paribus* is known to researchers)

Steps for Causal Inference

Research Question → Identification → Estimation

Identification

✓ Conditions that permit to estimate a causal effect from observed data

Selection on Observables
→ Theories about outcome and its confounders

Selection on Unobservables
→ Research designs (where data generation process that guarantees *Ceteris Paribus* is known to researchers)

Estimation

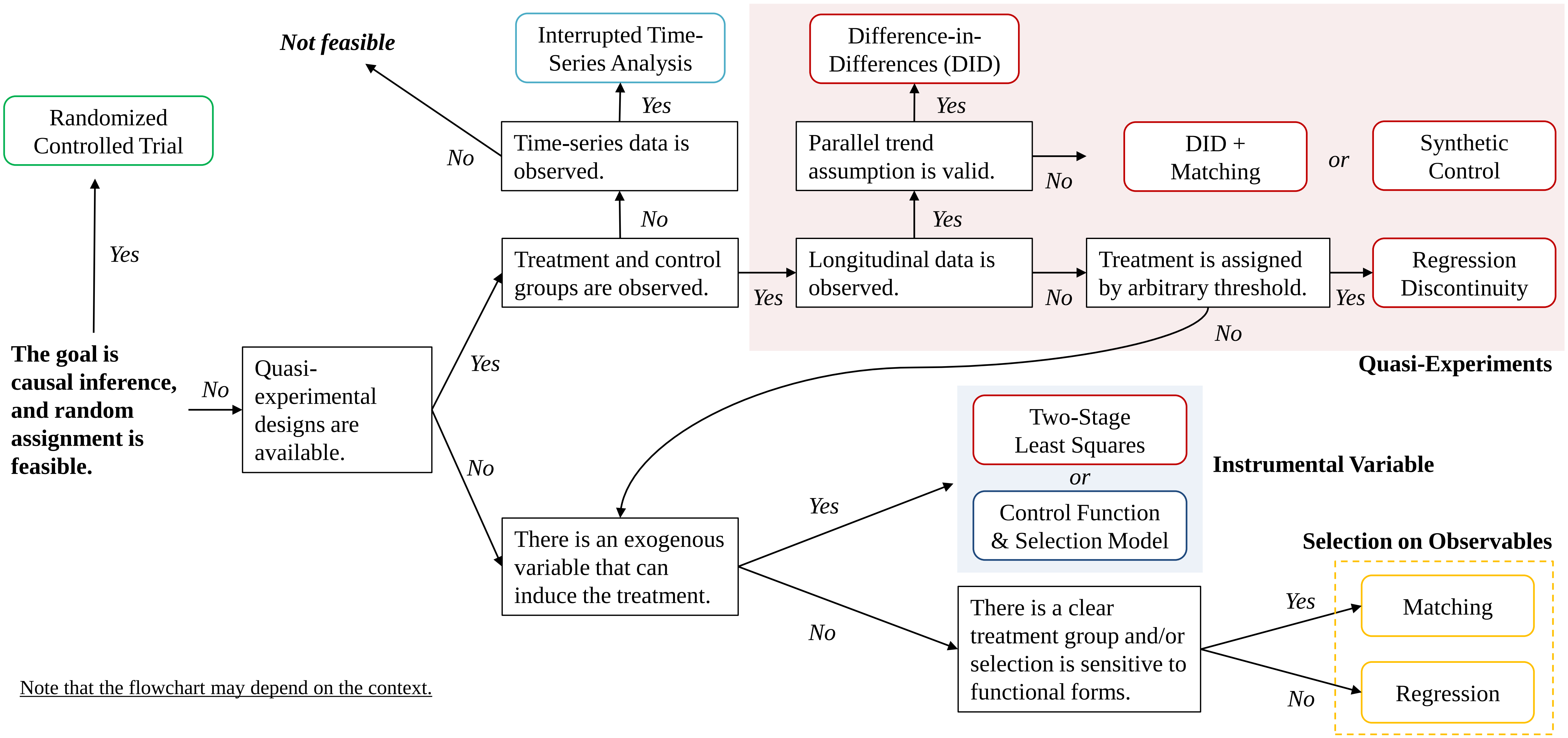
✓ Methods that estimate a causal effect using observed data

Estimation Methods
→ Regression, matching, etc.

→ Randomized controlled trial, difference-in-differences, synthetic control, regression discontinuity, instrumental variable, etc.

Data

What's Your Research Design and Data Structure?



Note that the flowchart may depend on the context.

Concluding Remarks

- No research methodology can save any wrong research questions.
  - Methodology cannot be a sufficient condition for good research.

"Type III errors occur when a researcher answers the wrong question using the right methods. A lot of effort may be expended, a great deal of rigor may be applied, but **coming up with the right answer to the wrong question does not create value**." (p. iii)

"An incomplete or imprecise answer to the right question can be a significant advance, while **a complete and precise answer to the wrong question does not create value**." (p. vii)

- Arun Rai, Former Editor-in-Chief of *MIS Quarterly*

Rai, A., 2017. Editor's Comments: Avoiding Type III Errors: Formulating IS Research Problems that Matter. *MIS Quarterly*, 41(2), pp.iii-vii.

Concluding Remarks

- Wrong methodology can ruin good research questions (even if you have good data).
  - Methodology cannot be a sufficient condition, but can be a necessary condition for good research.

"The group that is introducing new ideas and new topics to IS research that faces the most significant of challenges in publishing in top outlets (journals, conferences, etc.) because **the review process is designed to avoid type I errors (with the null hypothesis that the paper has no value or is wrong)** and in the process we make more type II errors where we fail to assess research that does make significant contributions as significant." (p. 2)

- Alok Gupta, Editor-in-Chief of *Information Systems Research*

Gupta, A., 2017. Editorial Thoughts: What and How ISR Publishes. *Information Systems Research*, 28(1), pp.1-4.

Concluding Remarks

- There is no silver bullet. There is only the right tool for a specific question.

Abraham Maslow (1966), *Psychology of Science*

Concluding Remarks

- Recommended reading

EDITOR'S COMMENTS

Causality Meets Diversity in Information Systems Research

By: **Sunil Mithas, Senior Editor**
Ling Xue, Associate Editor
Ni Huang, Associate Editor
Andrew Burton-Jones, Editor-in-Chief

MIS Quarterly Vol. 46 No. 3 / September 2022

Table 1. Interplay between Different Views of Causation and Types of Research Questions

Causal view	Main focus, and typical research questions	Key concepts and references
Path analytic	• What are the causal paths that explain the effect of a treatment on an outcome, and is it mediated, moderated, or both? If so, how? • Example: ○ *How does* IT capability affect firm performance via other organizational capabilities?	Concomitant variation, structural equation modeling (SEM), LISREL / PLS, directed acyclic graphs, measurement and structural models, recursive and nonrecursive models, Granger causality[5] (Bollen, 1989; Duncan, 1966; Goldberger, 1972; Jöreskog, 1978; Pearl, 1998; Wright, 1921)
Potential outcomes	• How much does the outcome change when the treatment changes? • Example: ○ *Does* IT capability increase firm profits or shareholder value compared to if the firm did not possess IT capability?	Rubin's causal model (RCM), matching, Instrumental Variables, Regression Discontinuity Design (RDD), fixed effects, difference-in-differences (DID), mechanism-based causal inference, sensitivity analysis (Angrist & Pischke, 2009; Fisher, 1935; Imbens, 2020; Imbens & Rubin, 2015; Splawa-Neyman, 1923/1990; Pearl, 2000; Rubin, 1974)
Configurational	• What sets of conditions lead to an outcome of interest and in what configurations does a particular factor of interest lead to the outcome? • Example: ○ *What configurations* of IT capabilities and other organizational capabilities create high (vs. low) performance?	Conjunctural causation, regularity, constant conjunction, complex causality (necessary/ sufficient conditions, INUS), sets and Boolean Algebra (Mackie, 1965; Mill, 1843/1882; Ragin, 1987)

This editorial can be downloaded at https://misq.umn.edu/contents-46-3.

Thank You!

Email: jiyong.park@uncg.edu

More information (including ppt files for educational purpose) can be found at https://sites.google.com/view/causal-inference2022.