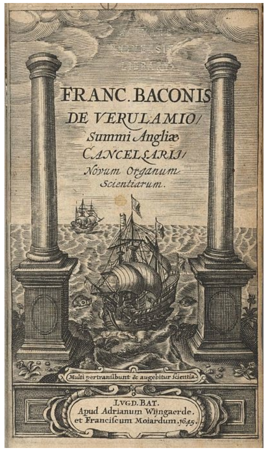
Data Analytics

#	Agenda
1	Causal inference
2	Project work
3	Presentations

Evidence-based decision making

Francis Bacon
(1561-1626)



1620

Causal inference in economics and marketing

Hal R. Varian

Varian, H. R. (2016). Causal inference in economics and marketing. Proceedings of the National Academy of Sciences of the United States of America, 113(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

Bacon suggests that one can draw up a list of all things in which the phenomenon to explain occurs, as well as a list of things in which it does not occur. Then one can rank the lists according to the degree in which the phenomenon occurs in each one. Then one should be able to deduce what factors match the occurrence of the phenomenon in one list and do not occur in the other list, and also what factors change in accordance with the way the data had been ranked.

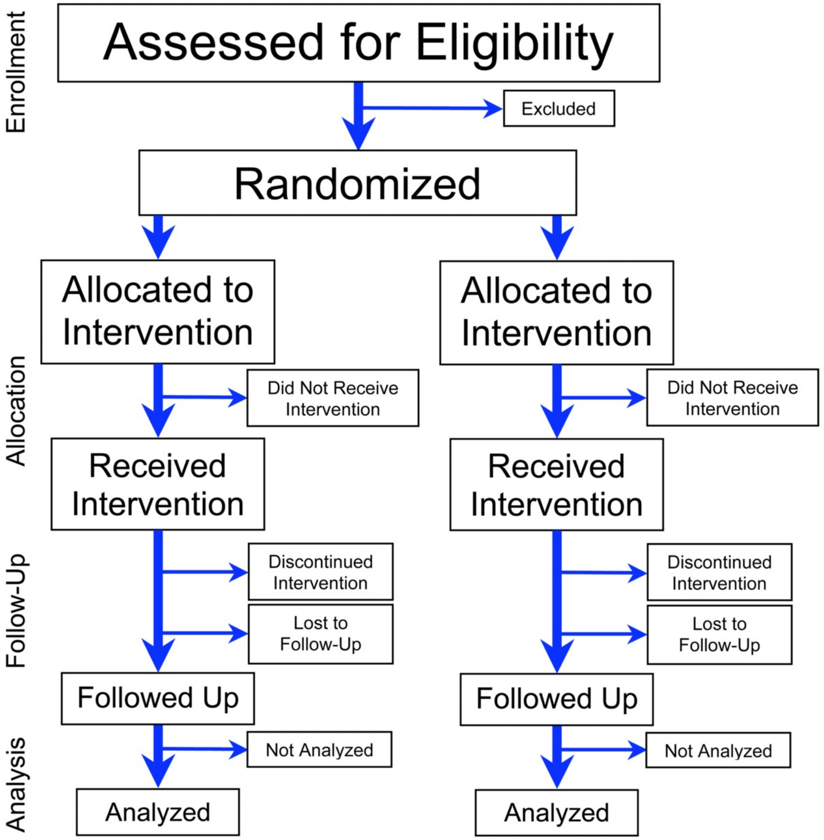
"The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment"

The gold standard…

Experiments/Randomised control trials (RCT)

A type of scientific experiment, where the people being studied are randomly allocated one or other of the different treatments under study. RCTs are considered the gold standard for a clinical trial. RCTs are often used to test the efficacy or effectiveness of various types of medical intervention and may provide information about adverse effects, such as drug reactions. Random assignment of intervention is done after subjects have been assessed for eligibility and recruited, but before the intervention to be studied begins.

$$Y = B_0 + B_1\text{group}$$



Shorter, E. (2011). A brief history of placebos and clinical trials in psychiatry. *Canadian Journal of Psychiatry, 56*(4), 193–197.

What are the limits of RCTs?

The Salk Polio Vaccine Trial & the Cutter Incident

- The 1954 Salk Polio vaccine trial was the largest RCT (a double-blind, randomized, and placebo-controlled study) ever conducted, involving over 1.8 million children, to test the safety and efficacy of a polio vaccine developed by Jonas Salk.
- The results showed that the vaccine was safe and effective in preventing polio.
- In 1955, shortly after the Salk polio vaccine was licensed, a manufacturing error at one of 5 licensed laboratories, Cutter Laboratories, resulted in the contamination of some batches of the vaccine with live polio virus, which led to an outbreak that affected a few hundred children, including some deaths and cases of permanent paralysis, known as the Cutter incident.
- The Cutter incident led to significant changes in vaccine regulation including the creation of oversight agencies and legislation.

The Cutter incident is an example of the problems that may arise from generalizing RCTs – and the continued need for evaluation…

A manufacturing error at Cutter Laboratories resulted in the contamination of some batches of the vaccine with live polio virus

Offit, P.A. (2005). The Cutter incident, 50 years later. *N Engl J Med*. 352, 1411-1412.

Dawson, L. (2004). The Salk polio vaccine trial of 1954: Risks, randomization and public involvement in research. *Clinical Trials, 1*, 122–130.

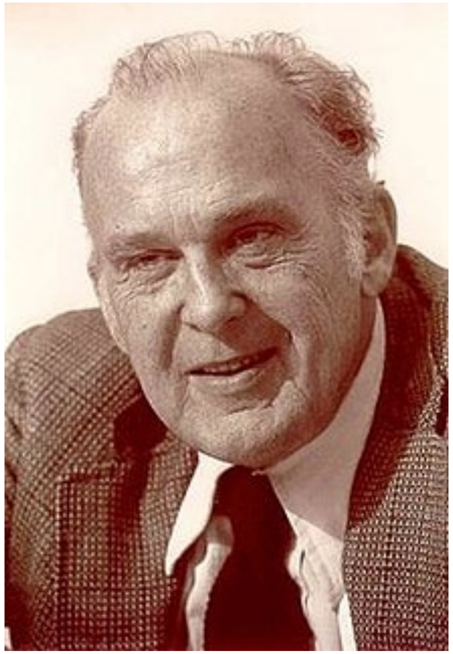
The gold standard is not always gold…

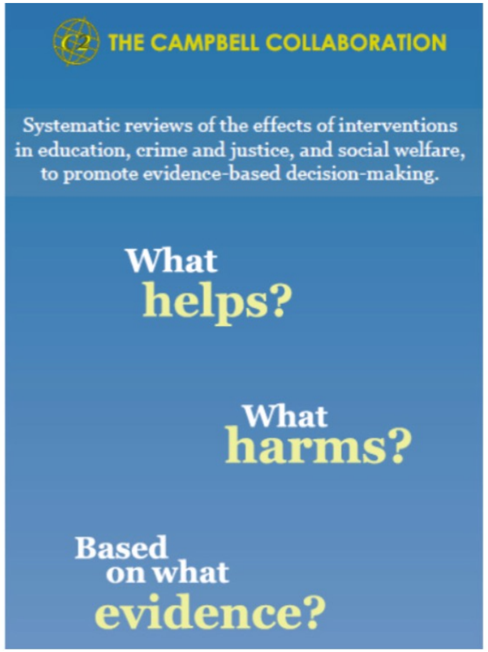
Table 1 Pros and cons of randomized control trials and observational studies

	Randomized control trials	Observational studies
Pros	• Random assignment makes study groups similar and comparable (no confounding at baseline) • Best fit to establish the efficacy of pharmacologic interventions • Currently considered as the gold standard for studying the effect of an intervention • Based on clear and well-established guidelines • Gives the true effect of an intervention under ideal conditions (internal validity)	• Useful to provide real-world evidence (external validity) • Relatively fast and inexpensive to conduct when data is already available • May take advantage of already available data like electronic health records • Suitable for studies where randomization is not ethical, or not feasible (e.g. rare diseases)
Cons	• Can be costly and take many years to conduct • Data collected may be biased due to non-compliance and drop-outs (post-randomization bias) • Possible to overlook biases • Generalizable only in simple systems, or when the conditions are exactly replicated	• Subject to outside factors that could distort the effect of the intervention (confounding) • Can be complex to design • Advanced analytical approaches are often required • Subject to limitations in the data available

Fernainy, P. et al. (2024). Rethinking the pros and cons of randomized controlled trials and observational studies in the era of big data and advanced methods: A panel discussion. BMC Proceedings, 18(Suppl 2):1 https://doi.org/10.1186/s12919-023-00285-8

There are alternatives…





Donald Campbell
1916-1996

THE CAMPBELL
COLLABORATION

Quasi-experimental designs

Before-and-after measures

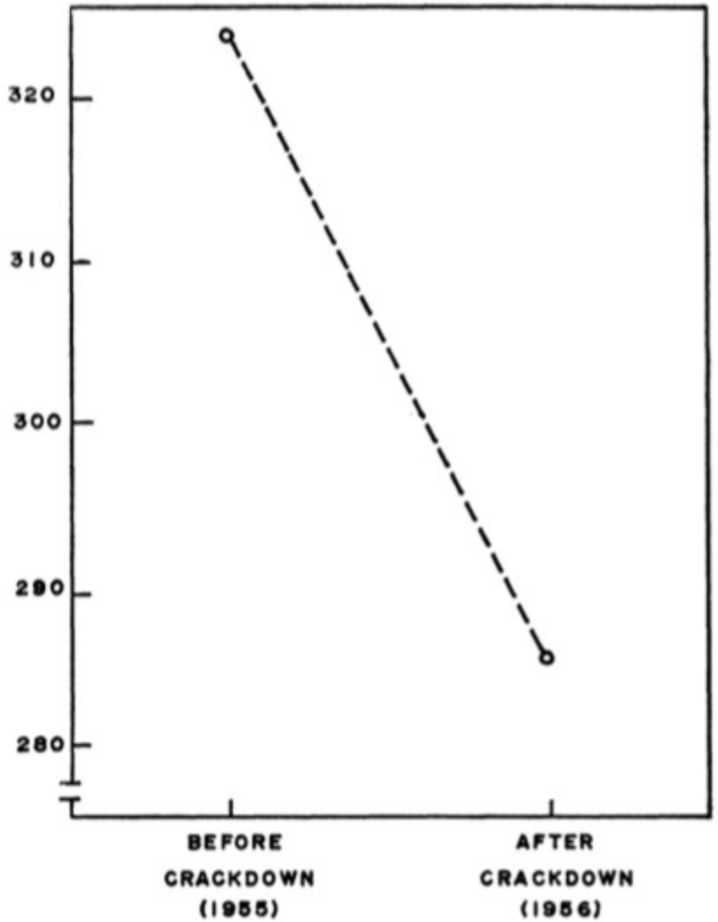


Figure 1. Connecticut Traffic Fatalities, 1955-1956

- was 1956 a dry year? (**history**: External events at the time of the intervention may explain changes rather than the intervention itself)
- overall trends in road safety? (**maturation**: Natural trends over time that could lead to change independent of the intervention)
- did publicising of death rates have an effect? (**testing**: The act of measuring or publicizing information may influence behavior)
- were fatalities counted differently? (**instrumentation**: Changes in how measurements are taken may affect the recorded outcomes)
- was this a big decrease? (**instability**: The observed effect may be due to random fluctuations rather than a systematic intervention)
- was 1995 an extreme year? (**regression**: Extreme values tend to move towards the average over time, independent of any intervention)

Campbell, D. T., Ross, H. L. (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 3(1), 33. http://doi.org/10.2307/3052794

Quasi-experimental designs

Interrupted time series

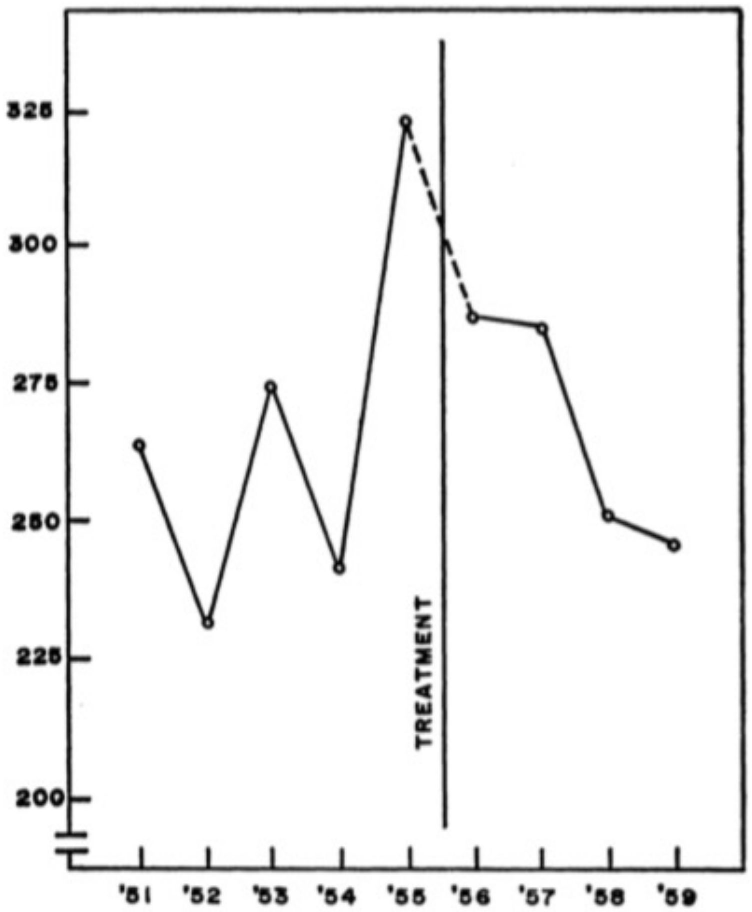


Figure 2. Connecticut Traffic Fatalities, 1951-1959

- was publicising of death rates similar across years? (**testing**: The act of measuring or publicizing information may influence behavior)
- were fatalities counted differently before and after the intervention? (**instrumentation**: Changes in how measurements are taken may affect the recorded outcomes)

Campbell, D. T., Ross, H. L. (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 3(1), 33. http://doi.org/10.2307/3052794

Quasi-experimental designs

Multiple time series

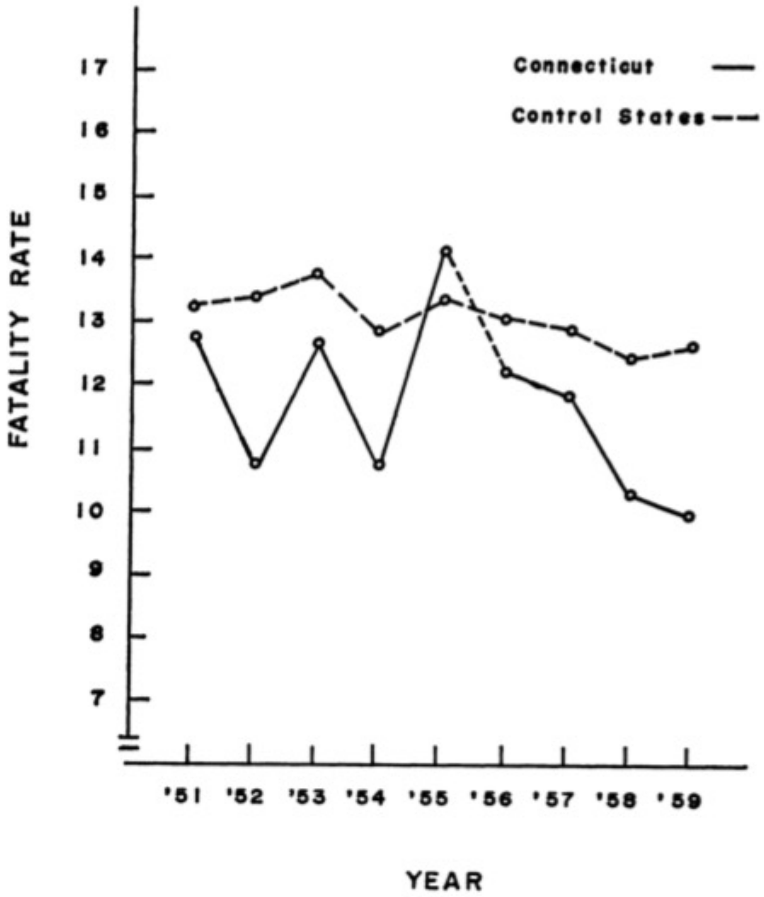


Figure 3. Connecticut and Control States Traffic Fatalities, 1951-1959 (per 100,000 population)

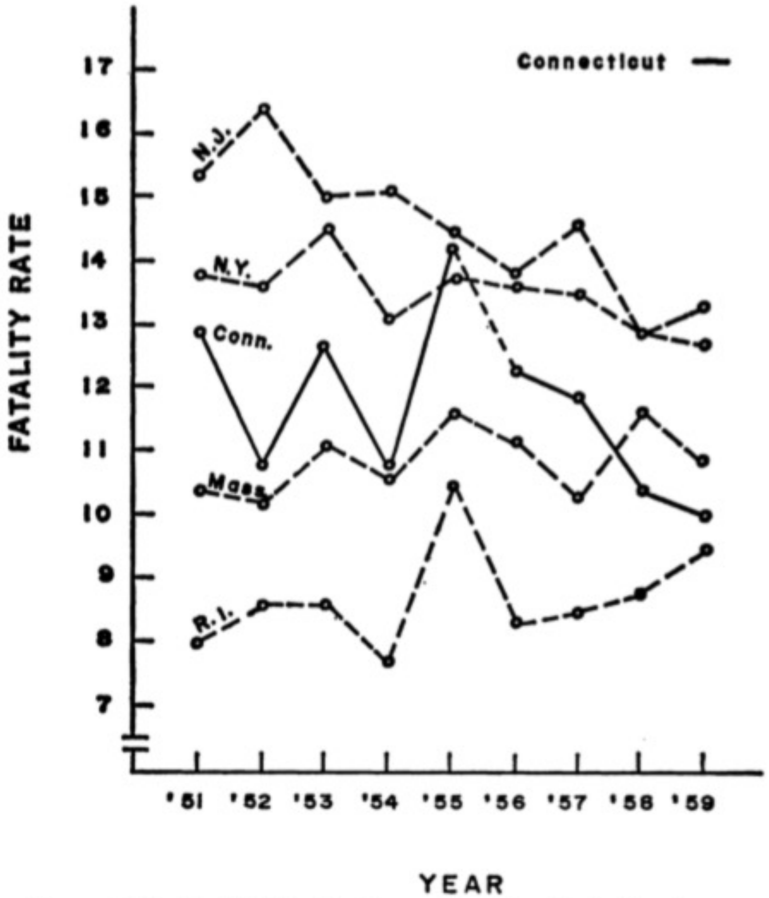
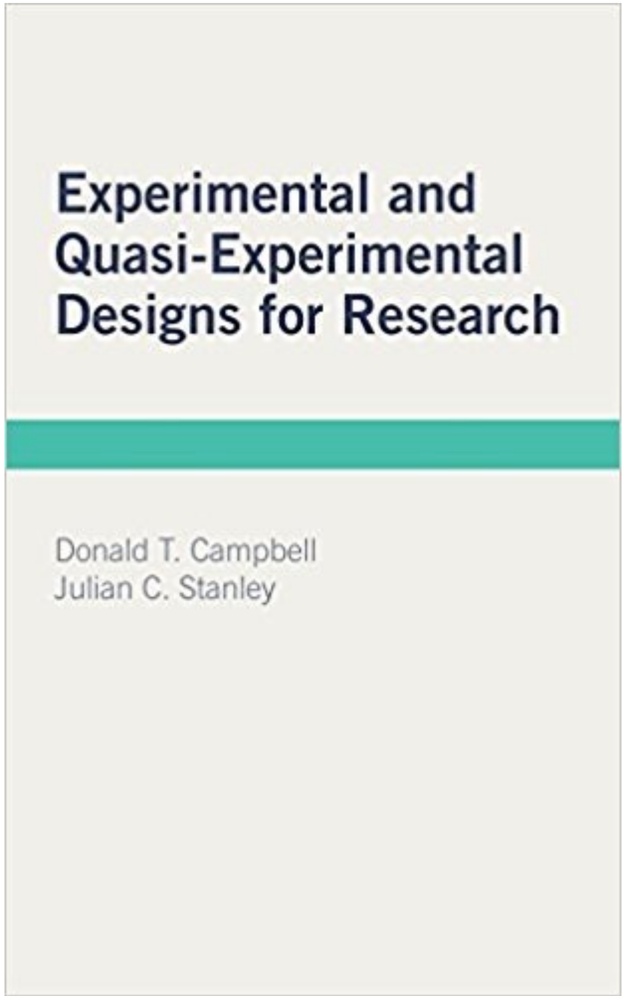


Figure 4. Traffic Fatalities for Connecticut, New York, New Jersey, Rhode Island, and Massachusetts (per 100,000 persons)

Campbell, D. T., Ross, H. L. (1968). The Connecticut crackdown on speeding: Time-series data in quasi-experimental analysis. Law and Society Review, 3(1), 33. http://doi.org/10.2307/3052794


Experimental and Quasi-Experimental Designs for Research
Donald T. Campbell
Julian C. Stanley

1963

TABLE 1

SOURCES OF INVALIDITY FOR DESIGNS 1 THROUGH 6

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Pre-Experimental Designs:												
1. One-Shot Case Study *X O*	−	−				−	−			−		
2. One-Group Pretest-Posttest Design *O X O*	−	−	−	−	?	+	+	−	−	−	?	
3. Static-Group Comparison *X O* - - - - - - *O*	+	?	+	+	+	−	−	−		−		
True Experimental Designs:												
4. Pretest-Posttest Control Group Design *R O X O* *R O O*	+	+	+	+	+	+	+	+	−	?	?	
5. Solomon Four-Group Design *R O X O* *R O O* *R X O* *R O*	+	+	+	+	+	+	+	+	+	?	?	
6. Posttest-Only Control Group Design *R X O* *R O*	+	+	+	+	+	+	+	+	+	?	?	

Note: In the tables, a minus indicates a definite weakness, a plus indicates that the factor is controlled, a question mark indicates a possible source of concern, and a blank indicates that the factor is not relevant.

It is with extreme reluctance that these summary tables are presented because they are apt to be "too helpful," and to be depended upon in place of the more complex and qualified presentation in the text. No + or − indicator should be respected unless the reader comprehends why it is placed there. In particular, it is against the spirit of this presentation to create uncomprehended fears of, or confidence in, specific designs.

TABLE 2

SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 7 THROUGH 12

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Quasi-Experimental Designs:												
7. Time Series $O\ O\ O\ OXO\ O\ O\ O$	−	+	+	?	+	+	+	+	−	?	?	
8. Equivalent Time Samples Design $X_1O\ X_0O\ X_1O\ X_0O$, etc.	+	+	+	+	+	+	+	+	−	?	−	−
9. Equivalent Materials Samples Design $M_aX_1O\ M_bX_0O\ M_cX_1O\ M_dX_0O$, etc.	+	+	+	+	+	+	+	+	−	?	?	−
10. Nonequivalent Control Group Design $O\ X\ O$ - - - - - - - $O\ \ \ \ O$	+	+	+	+	?	+	+	−	−	?	?	
11. Counterbalanced Designs $X_1O\ X_2O\ X_3O\ X_4O$ - - - - - - - $X_2O\ X_4O\ X_1O\ X_3O$ - - - - - - - $X_3O\ X_1O\ X_4O\ X_2O$ - - - - - - - $X_4O\ X_3O\ X_2O\ X_1O$	+	+	+	+	+	+	+	?	?	?	?	−
12. Separate-Sample Pretest-Posttest Design $R\ O\ (X)$ $R\ \ \ \ X\ O$	−	−	+	?	+	+	−	−	+	+	+	
12*a*. $R\ O\ (X)$ $R\ \ \ \ X\ O$ - - - - - - - $R\ \ \ \ O\ (X)$ $R\ \ \ \ \ \ \ X\ O$	+	−	+	?	+	+	−	+	+	+	+	
12*b*. $R\ O_1\ \ \ (X)$ $R\ \ \ \ O_2\ (X)$ $R\ \ \ \ \ \ \ X\ O_3$	−	+	+	?	+	+	−	?	+	+	+	
12*c*. $R\ O_1\ X\ O_2$ $R\ \ \ \ X\ O_3$	−	−	+	?	+	+	+	−	+	+	+	

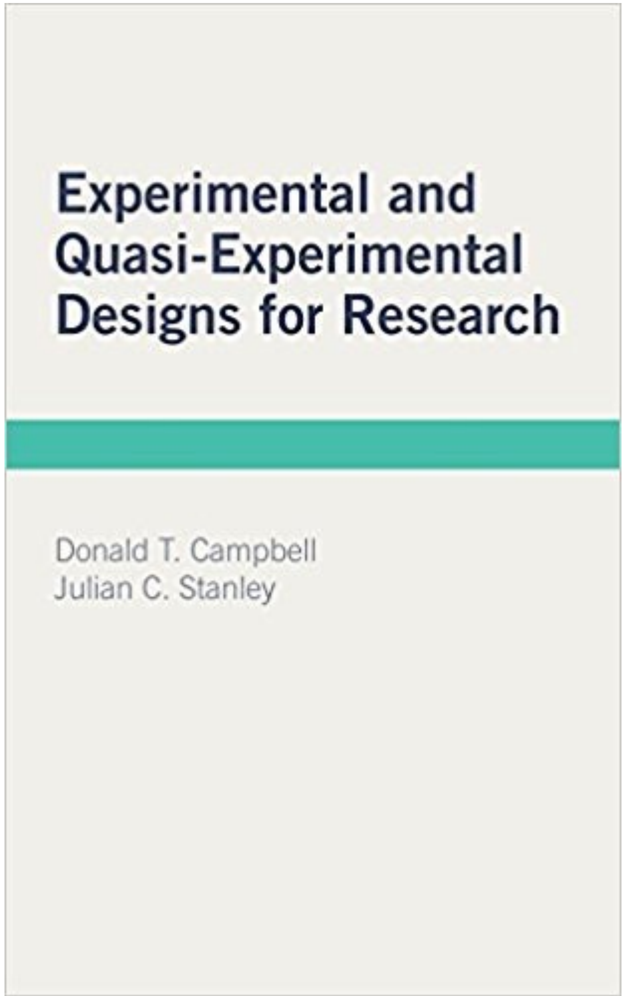
TABLE 3

SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 13 THROUGH 16

	Sources of Invalidity											
	Internal								External			
	History	Maturation	Testing	Instrumentation	Regression	Selection	Mortality	Interaction of Selection and Maturation, etc.	Interaction of Testing and X	Interaction of Selection and X	Reactive Arrangements	Multiple-X Interference
Quasi-Experimental Designs Continued:												
13. Separate-Sample Pretest-Posttest Control Group Design R O (X) R X O R O R O	+	+	+	+	+	+	+	−	+	+	+	
13a. R' { R O (X) / R X O / R O (X) / R X O / R O (X) / R X O } R' { R O / R O / R O / R O / R O / R O }	+	+	+	+	+	+	+	+	+	+	+	
14. Multiple Time-Series O O OXO O O O O O O O O	+	+	+	+	+	+	+	+	−	−	?	
15. Institutional Cycle Design Class A X O_1 Class B_1 RO_2 X O_3 Class B_2 R X O_4 Class C O_5 X [a]Gen. Pop. Con. Cl. B O_6 [a]Gen. Pop. Con. Cl. C O_7												
$O_2 < O_1$, $O_5 < O_4$	+	−	+	+	?	−	?		+	?	+	
$O_2 < O_3$	−	−	−	?	?	+	+		−	?	+	
$O_2 < O_4$	−	−	+	?	?	+	?		+	?	?	
$O_6 = O_7$, $O_{2y} = O_{2o}$		+						−				
16. Regression Discontinuity	+	+	+	?	+	+	?	+	+	−	+	+

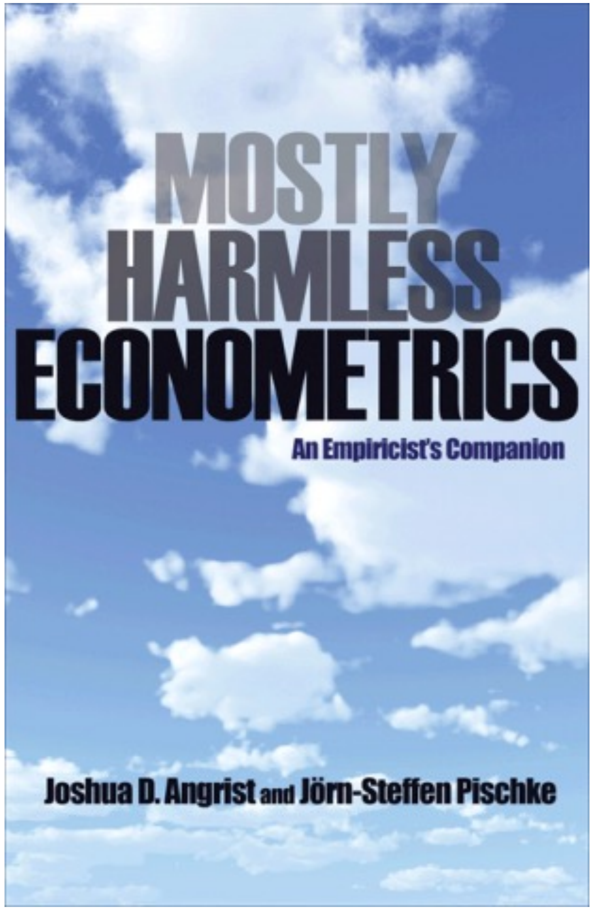
[a] General Population Controls for Class B, etc.

Experimental and Quasi-experimental Designs



"In conclusion, in this chapter we have discussed alternatives in the arrangement or design of experiments, with particular regard to the problems of control of extraneous variables and threats to validity. (…) Through out, attention has been called to the possibility of **creatively** utilizing the idiosyncratic features of any specific research situation in designing unique tests of causal hypotheses.

“Furious Five” statistical methods for causal inference

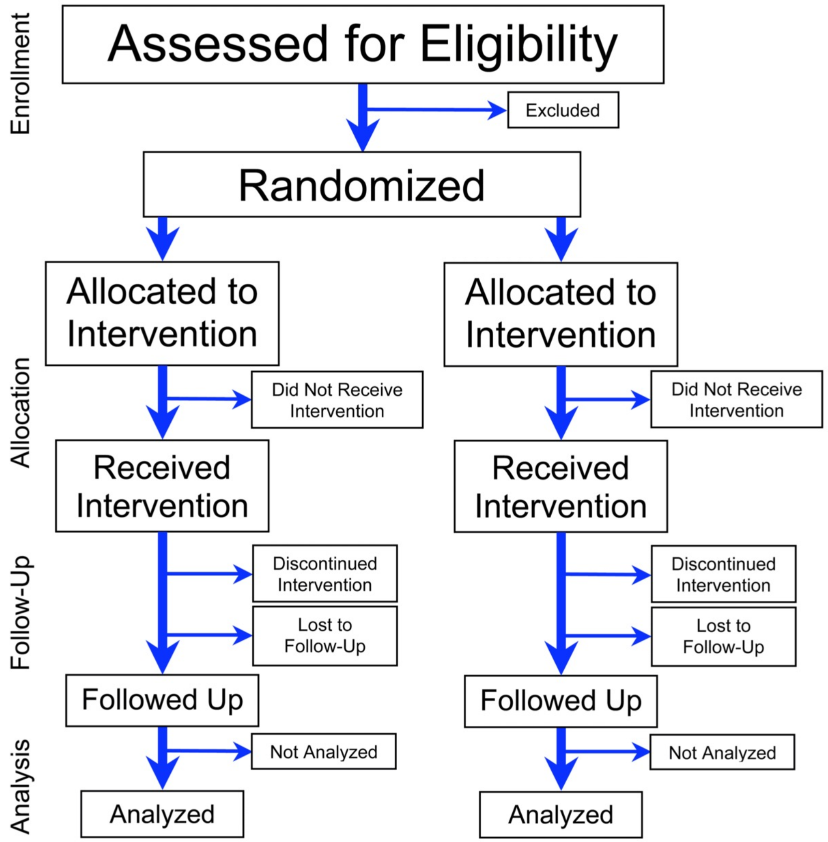


- Randomisation
- Regression
- Difference in differences
- Regression discontinuity
- Instrumental variables

Angrist, J. D., & 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), 3–30. http://doi.org/10.1257/jep.24.2.3

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

Randomisation



The "ideal" data, from the viewpoint of the analyst, would be data from an incompetent advertiser who allocated expenditures randomly across cities. If ad expenditure is truly random, then we do not have to worry about confounding variables because the predictors will automatically be orthogonal to the error term. However, statisticians are seldom lucky enough to have a totally incompetent client.

Problem: Randomization is not always possible or desirable

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

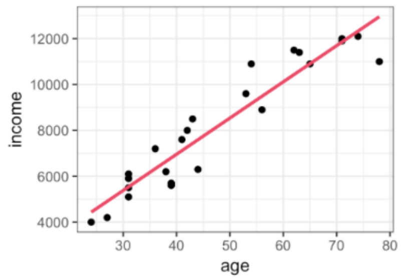
Regression

Regression analysis is a set of statistical processes for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable (criterion) and one or more independent variables (predictors). More specifically, regression analysis helps one understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are fixed.

Definition: Simple linear regression is a linear model with one predictor x, and where the error term ϵ is Normally distributed.

$$y = \beta_0 + \beta_1 x + \epsilon$$

Parameter	Description	In words
β_0	Intercept	When x = 0, what is the predicted value for y?
β_1	Coefficient for x	For every increase of 1 in x, how does y change?



Formula

$$income = 885 + 149.3 \times age + \epsilon$$

Coefficients

$$\beta_0 = 885, \beta_{age} = 149.3$$

Problem: Correlation is not causation…

Regression: Nested structures

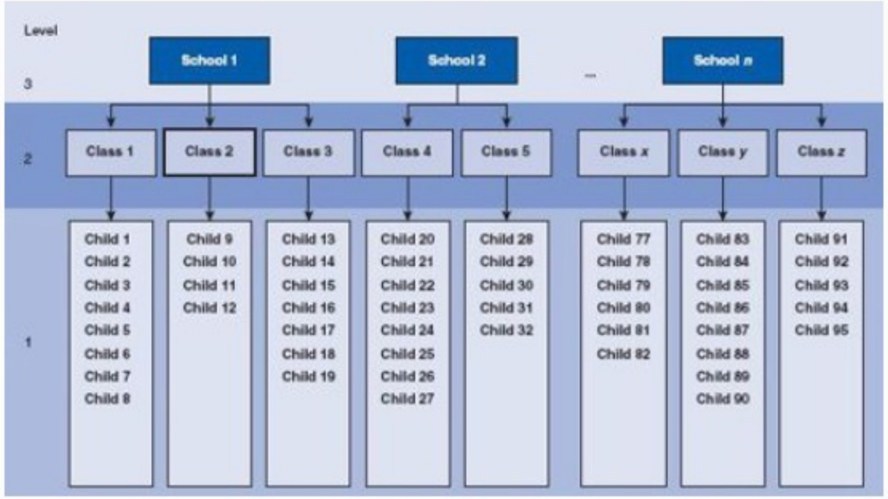


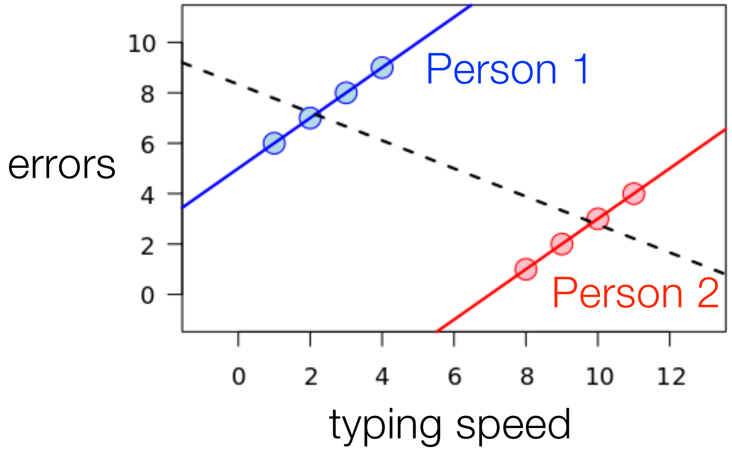
FIGURE 19.3 An example of a three-level hierarchical data structure

A mixed-effects regression model is a statistical model containing both **fixed** effects and **random** effects. These models are useful in a wide variety of disciplines in the physical, biological and social sciences. They are particularly useful in settings where repeated measurements are made on the same statistical units (longitudinal study), or where measurements are made on clusters of related statistical units. Because of their advantage in dealing with missing values, mixed effects models are often preferred over more traditional approaches.

Field, A., Miles, J., & Field, Z (2012). *Discovering statistics using R*. SAGE Publications.

Regression: Simpson’s paradox

Mixed-effects models can help deal with Simpson’s paradox in which a trend that appears in groups of data disappears when these groups are combined and the reverse trend appears for the aggregate data.



speed-accuracy trade-off

people who are faster are also more accurate (between)

when people are faster they become less accurate (within)

exercise and fatigue

people who exercise more are less fatigued (between)

when people exercise more they are more fatigued (within)

Fitting Linear Mixed-Effects Models Using **lme4**

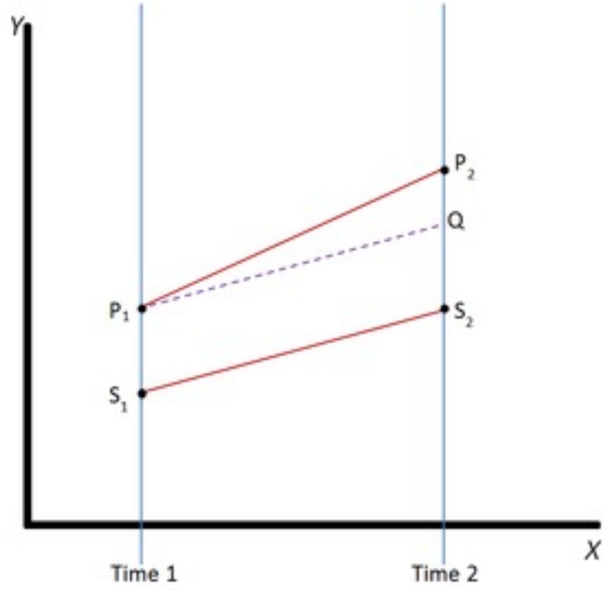
model <- lmer(Y ~ x + (x | g), data)

<table>
<tr><th>Formula</th><th>Alternative</th><th>Meaning</th></tr>
<tr><td><code>(1 | g)</code></td><td><code>1 + (1 | g)</code></td><td>Random intercept with fixed mean.</td></tr>
<tr><td><code>0 + offset(o) + (1 | g)</code></td><td><code>-1 + offset(o) + (1 | g)</code></td><td>Random intercept with a priori means.</td></tr>
<tr><td><code>(1 | g1/g2)</code></td><td><code>(1 | g1)+(1 | g1:g2)</code></td><td>Intercept varying among <code>g1</code> and <code>g2</code> within <code>g1</code>.</td></tr>
<tr><td><code>(1 | g1) + (1 | g2)</code></td><td><code>1 + (1 | g1) + (1 | g2).</code></td><td>Intercept varying among <code>g1</code> and <code>g2</code>.</td></tr>
<tr><td><code>x + (x | g)</code></td><td><code>1 + x + (1 + x | g)</code></td><td>Correlated random intercept and slope.</td></tr>
<tr><td><code>x + (x || g)</code></td><td><code>1 + x + (1 | g) + (0 + x | g)</code></td><td>Uncorrelated random intercept and slope.</td></tr>
</table>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–51. http://doi.org/10.18637/jss.v067.i01

Difference in differences

Difference in differences (DID or DD) is a statistical technique used in the social sciences that attempts to mimic an experimental research design using observational study data, by studying the differential effect of a treatment on a 'treatment group' versus a 'control group' in a natural experiment. It calculates the effect of a treatment on an outcome by comparing the average change over time in the outcome variable for the treatment group, compared to the average change over time for the control group. Although it is intended to mitigate the effects of extraneous factors and selection bias, depending on how the treatment group is chosen, this method may still be subject to certain biases (e.g., omitted variable bias).

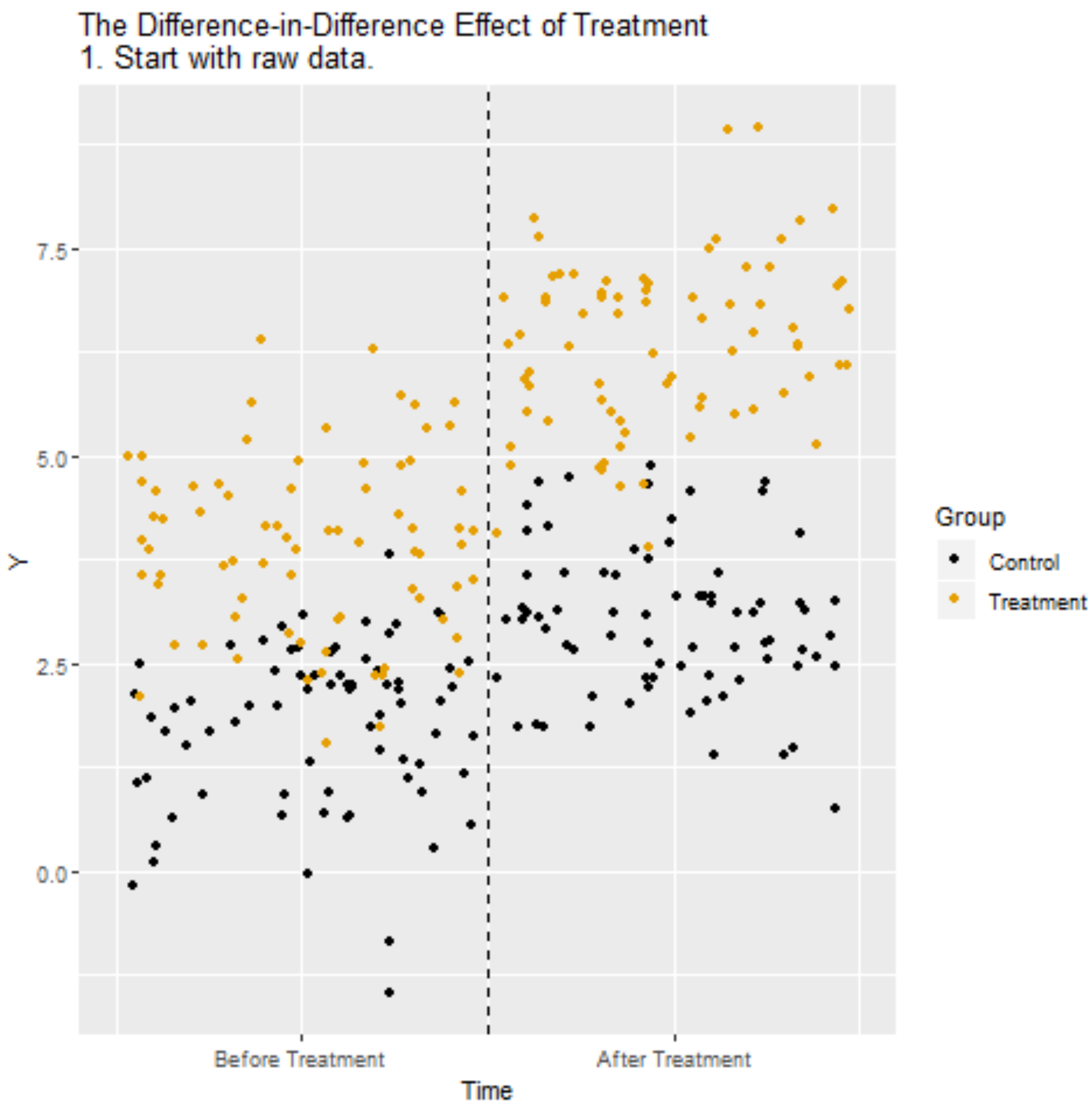


$$Y = B_0 + B_1 Group + B_2 Time + B_3\, Group{*}Time$$

Problem: Assumption that the change in outcomes from pre- to post-intervention in the control group (S) is a good proxy for the (counterfactual) change in untreated potential outcomes in the treated group (P) may not be warranted; choice of treatment/control groups is crucial (an additional trick may be *matching* on observables)…

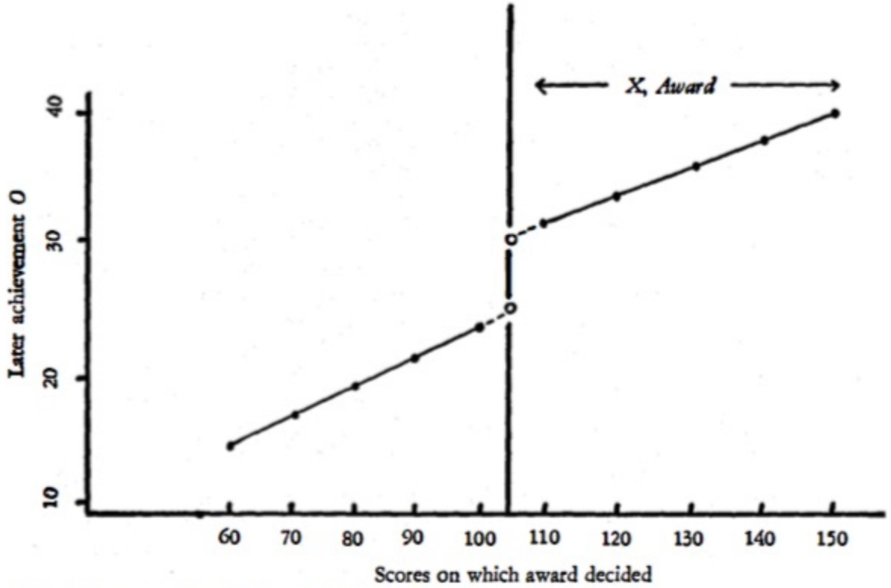
Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? The Quarterly Journal of Economics, 119(1), 249–275.

Difference in differences



Regression discontinuity

A regression discontinuity design (RDD) is a quasi-experimental pre-posttest design that elicits the causal effects of interventions by assigning a cutoff or threshold above or below which an intervention is assigned. By comparing observations lying closely on either side of the threshold, it estimates the average treatment effect in environments in which randomization is unfeasible. RDD was first applied by Thistlethwaite and Campbell to the evaluation of the effects of scholarship programs on (future) academic performance.



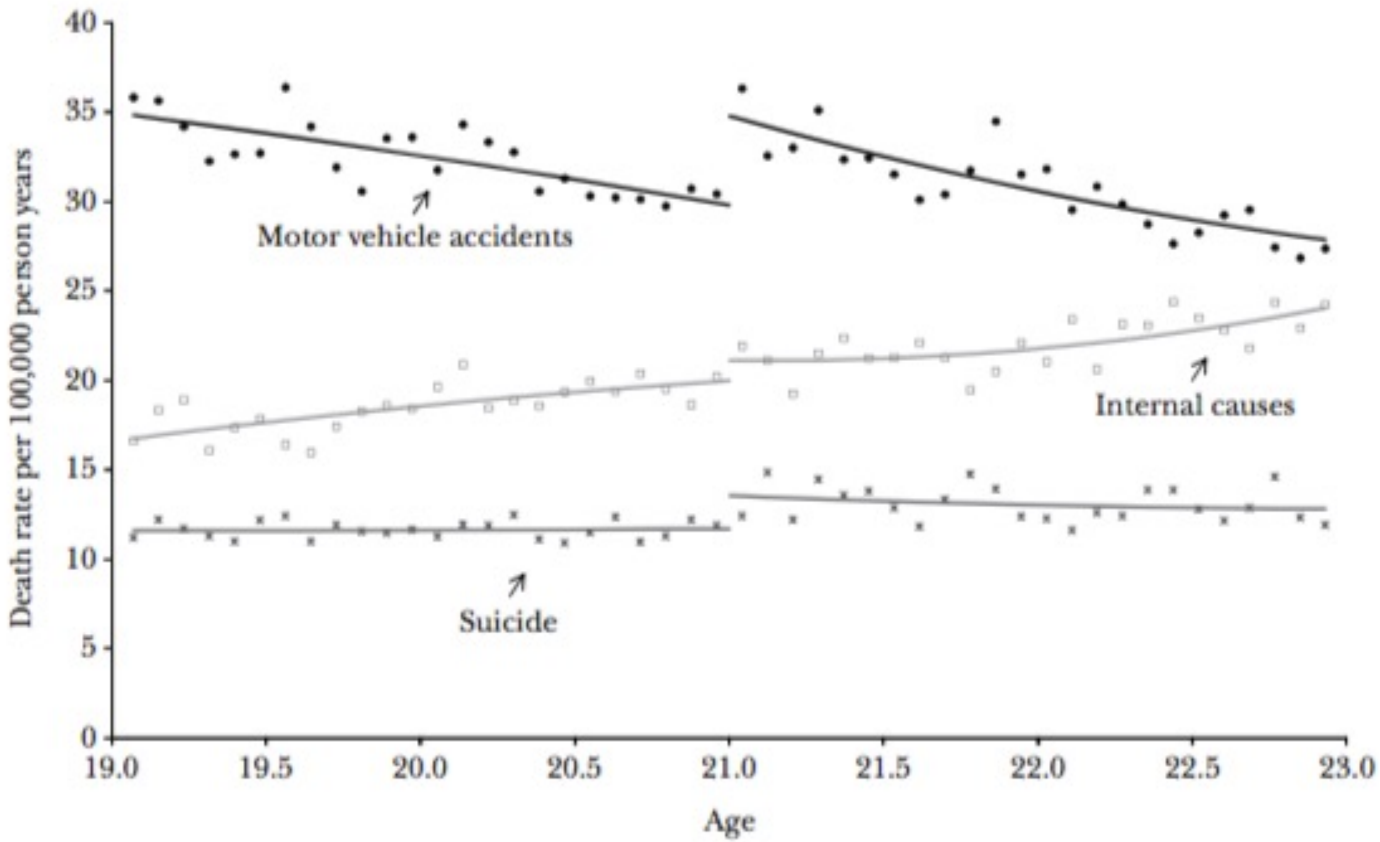
$$Y = B_0 + B_1 \text{Score} + B_2 \text{ Award}$$

Problem: Assumption that the individuals just below the cutoff are not systematically different from those just above can be wrong (e.g., individuals just above the threshold could try harder); the estimation may not generalise to observations away from the cutoff (e.g., awards could have different results at different levels of ability).

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature, 48*(2), 281–355. http://doi.org/10.1257/jel.48.2.281

Regression discontinuity

Figure 2
Age Profiles for Death Rates in the United States

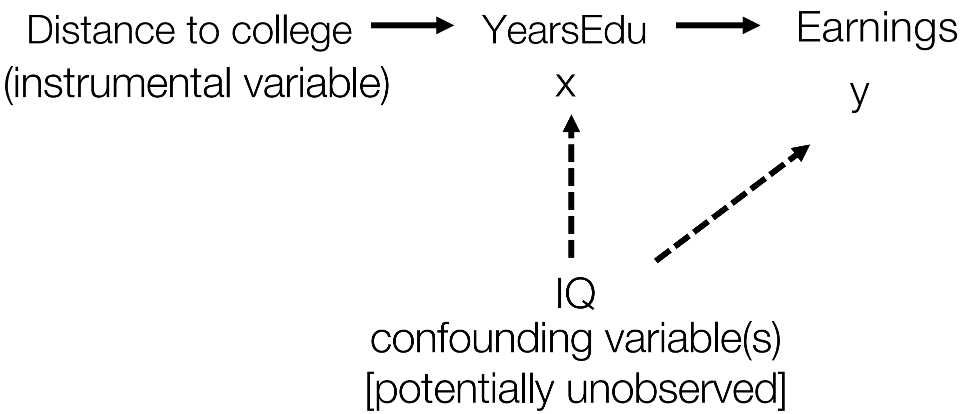


Notes: The death rates are estimated by combining the National Vital Statistics records with population estimates from the U.S. Census.

Carpenter, C., & Dobkin, C. (2011). The Minimum Legal Drinking Age and Public Health. *Journal of Economic Perspectives*, *25*(2), 133–156.

Instrumental variables

The method of instrumental variables (IV) is used to estimate causal relationships when controlled experiments are not feasible or when a treatment is not successfully delivered to every unit in a randomized experiment. Intuitively, the method is used when an explanatory variable of interest is correlated with the error term, in which case ordinary least squares gives biased results. A valid instrument (instrumental variable) induces changes in the explanatory variable (x) but has no independent effect on the dependent variable (y), allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.



Estimation though two-stage least squares.

Stage 1: generate predictions of YearsEdu:

YearsEdu_pred = B_0 + B_1 DisttoCollege + Error

Stage 2: test whether YearsEdu_pred is significantly associated with earnings:

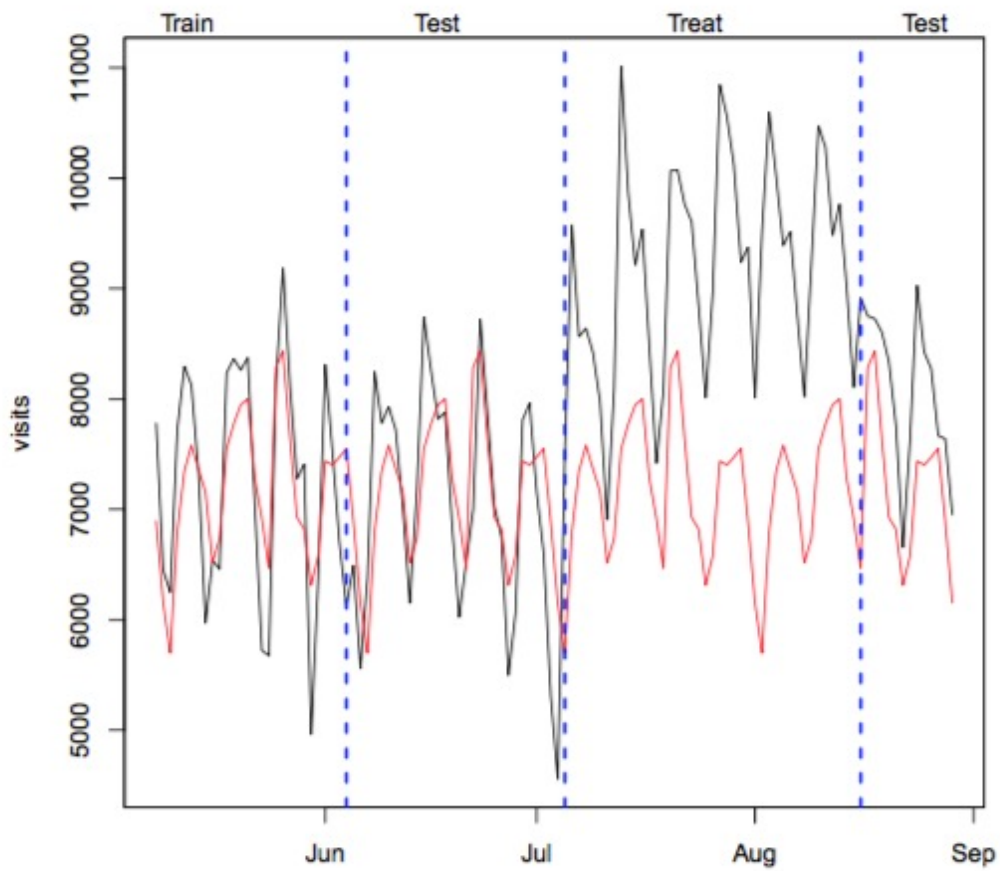
Earnings = B_0 + B_1 YearsEdu_pred + Error

Problem: Good instrumental variables (i.e., that are correlated with x but not any confounding variables) are hard to find…

Angrist, J. D., & Krueger, A. B. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives, 15*(4), 69–85.

New developments: Synthetic control methods

Using models as the control group (Train-test-treat-compare)



An online advertiser might ask "if I increase my ad expenditure by some amount, how many extra sales do I generate?"

A predictive statistical model (based on number of "searches" about topics related to the subject matter of the website) is estimated during the training period and its predictive performance is assessed during the test period. The extrapolation of the model during the treat period (red line) serves as a counterfactual. This counterfactual is compared with the actual outcome (black line), and the difference is the estimated treatment effect. When the treatment is ended, the outcome returns to something close to the original level.

Varian, H. R. (2016). Causal inference in economics and marketing. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

Use of approaches in economics vs psychlogy…

Table 1. Review of a Random Sample of Economics and Psychology Articles

	Economics	Psychology
Number of articles reviewed	108	108
Number of articles containing . . .		
An empirical study	88	96
A randomized experiment	17	42
A natural experiment	36[a]	0
A standard natural experiment with true randomization	1	0
A standard natural experiment with as-if randomization	19	0
An instrumental-variable design using a natural experiment with true randomization	1	0
An instrumental-variable design using a natural experiment with as-if randomization	18	0
A sharp regression-discontinuity design	3	0
A fuzzy regression-discontinuity design	2	0

Note: We reviewed a random sample of 216 articles published in eight flagship journals from psychology and economics in the year 2019. We sampled articles from four empirical psychology journals that had a relatively high impact according to the 2021 SCImago Journal Rank (*Journal of Applied Psychology*, *Journal of Personality and Social Psychology*, *Psychological Science*, and *Clinical Psychological Science*) and the four of the top five economic journals (e.g., Heckman & Moktan, 2020) that publish largely empirical studies (*American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*). A student assistant coded basic information (e.g., authors, year, DOI, and whether the article contains an empirical study or not). The study designs of each article were independently coded by either two authors of the current work or by one of the authors and the student assistant (for coding manual, data, and interrater agreements, see Tables S1 to S3 on the OSF at https://osf.io/a5nxm). Disagreements and uncertainties were resolved by the first author. [a]Of the 36 natural experiments, 11 were borderline cases (for details, see Table S2). The total number of articles using natural experiments (36) is smaller than the sum of articles using specific types of natural experiments because some articles used more than one type.

Grosz, M. P., Ayaita, A., Arslan, R. C., Buecker, S., Ebert, T., Hünermund, P., Müller, S. R., Rieger, S., Zapko-Willmes, A., & Rohrer, J. M. (2024). Natural Experiments: Missed Opportunities for Causal Inference in Psychology. Advances in Methods and Practices in Psychological Science, 7,1, 1-15. https://doi.org/10.1177/25152459231218610

Which approach, if any, could you use in your project?

Does education improve intelligence?

Quasi-Experimental Designs: Educational effects on intelligence

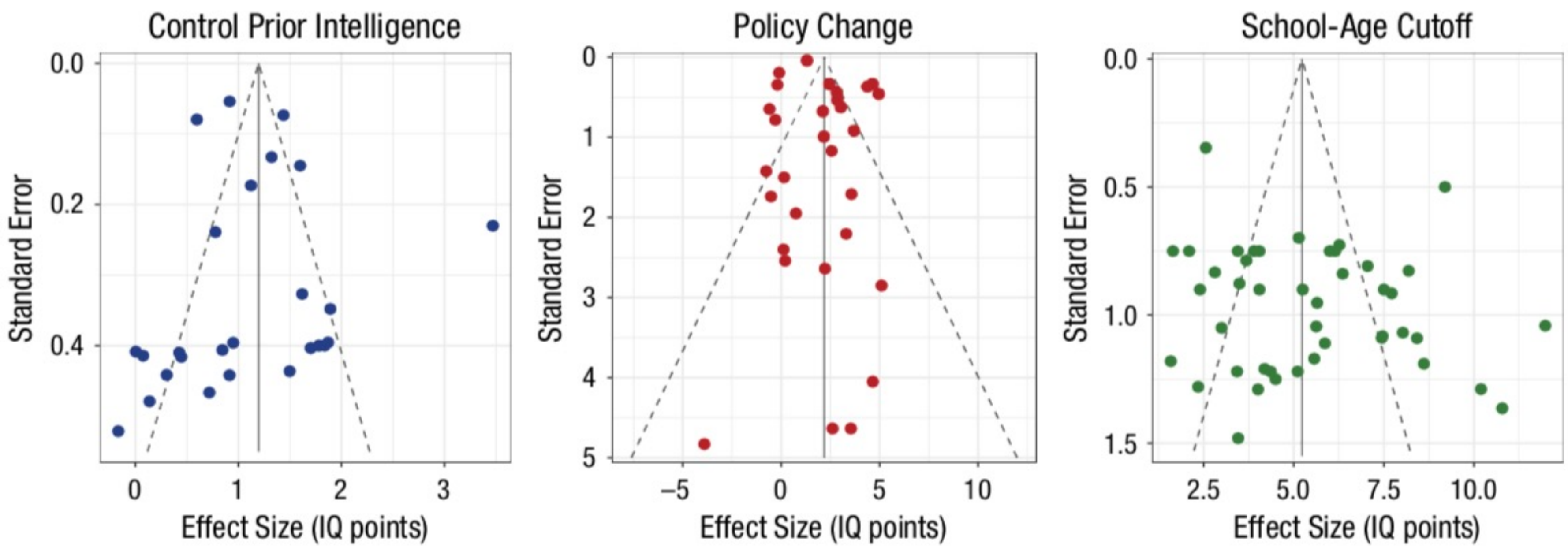


Fig. 2. Funnel plots showing standard error as a function of effect size, separately for each of the three study designs. The dotted lines form a triangular region (with a central vertical line showing the mean effect size) where 95% of estimates should lie in the case of zero within-group heterogeneity in population effect sizes. Note that 42 of the total 86 standard errors reported as approximate or as averages in the original studies were not included for the school-age-cutoff design.

control prior intelligence = longitudinal studies in which cognitive testing data were collected before and after variation in the duration of education (e.g., before and after university vs. no university)

policy change = study of the effects of a change in educational duration (e.g., increase of compulsory education by 1 year) on mental testing

school-age cutoff = studies use regression-discontinuity analysis to leverage the fact that school districts implement a date-of-birth cutoff for school entry (example: compare 3.9-year olds that did not attend "Kindsgi" vs. 4.0 year-olds that did)

Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*(8), 1358–1369. http://doi.org/10.1177/0956797618774253