Evidence-based Decision Making
Experiments

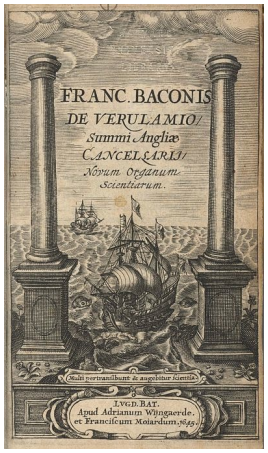
Rui Mata, FS 2023

Goals

- Understand the nature of causal inference as the comparison of treatment to some counterfactual
- Understand that RCTs/experiments have desirable properties for causal inference – but also have limitations…

Evidence-based decision making

Francis Bacon
(1561-1626)



1620

Causal inference in economics and marketing

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.

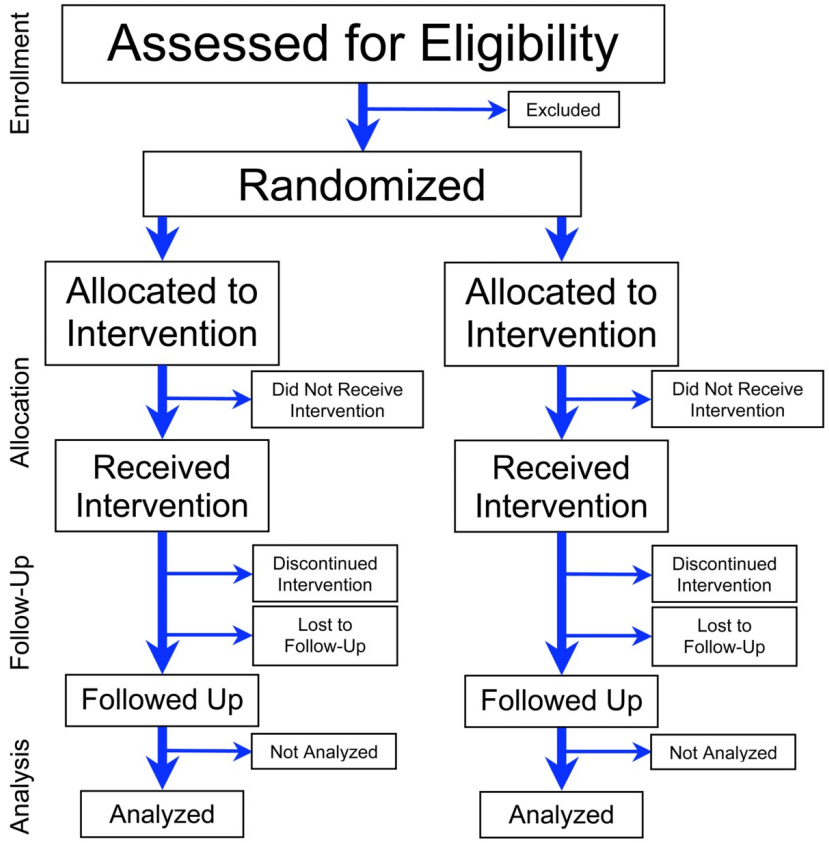
“The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment”

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.

$$Y = B_0 + B_1 group$$



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

The gold standard…

Experiments/Randomised control trials (RCT)

http://www.consort-statement.org

The gold standard is not always gold…

Experiments/Randomised control trials (RCT)

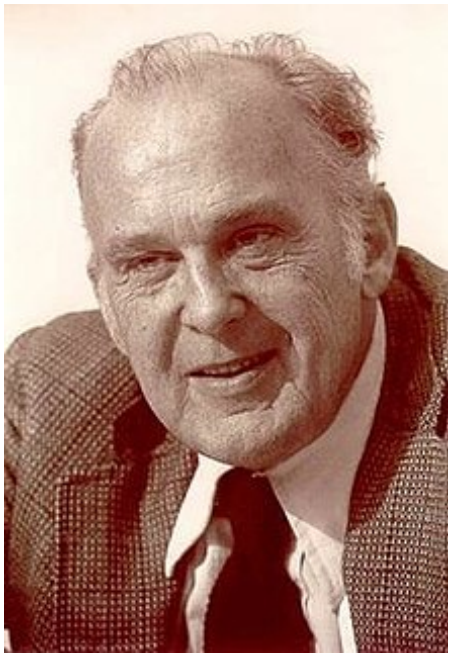
- Efficacy vs. effectiveness: Trials may not be widely applicable in real-world conditions….
- Generalizability: Results may not always generalize to other samples (e.g. inclusion /exclusion criteria)
- Ethical limitations: randomisation requires experimental equipoise: one cannot ethically randomise participants to some treatments (no-schooling condition)

Summary

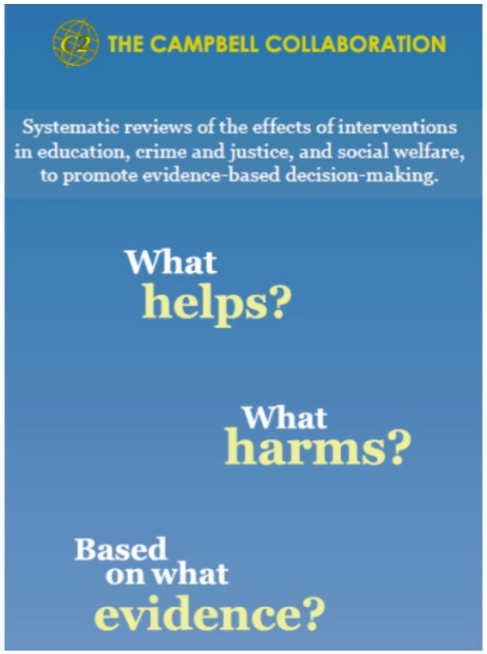
"The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment"

RCTs are great **but** do not guarantee effectiveness, generalizability, or ethical treatment of participants…

There are alternatives…



Donald Campbell
1916-1996



THE CAMPBELL COLLABORATION

Quasi-experimental designs

Before-and-after measures

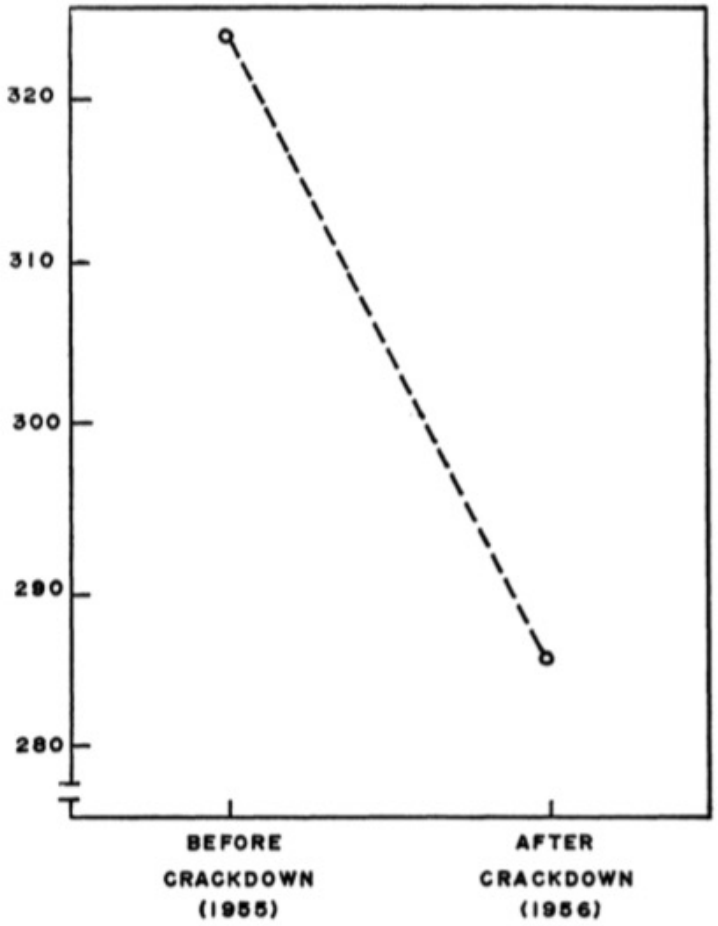


Figure 1. Connecticut Traffic Fatalities, 1955-1956

- was 1956 a dry year? (history)
- overall trends in road safety? (maturation)
- did publicising of death rates have an effect? (testing)
- were fatalities counted differently? (instrumentation)
- was this a big decrease? (instability)
- was 1995 an extreme year? (regression)

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

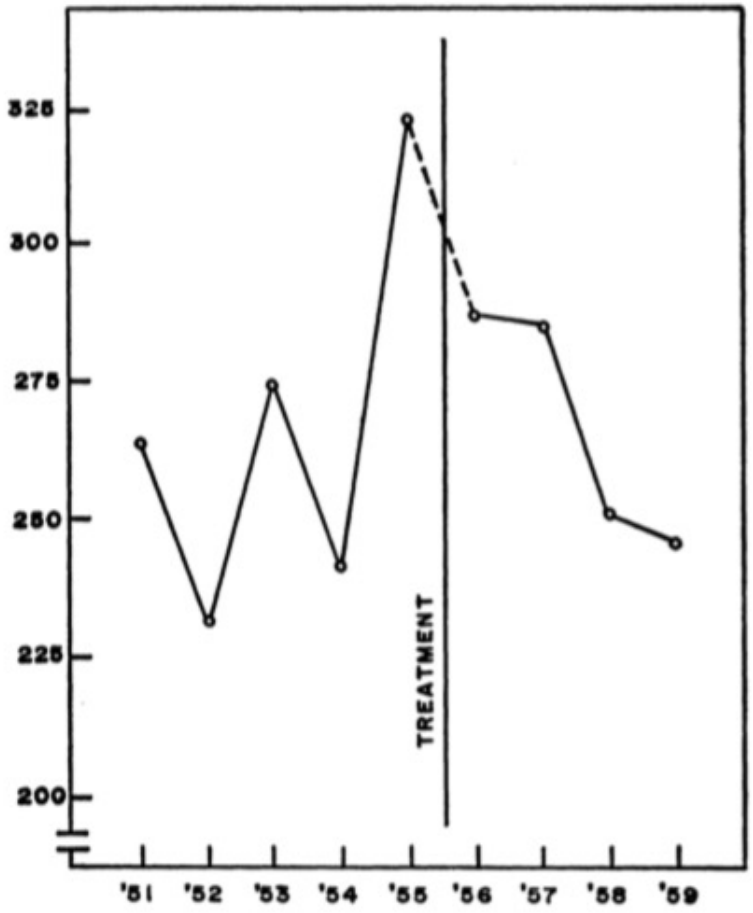


Figure 2. Connecticut Traffic Fatalities, 1951-1959

- was publicising of death rates similar across years? (testing)
- were fatalities counted differently before and after the intervention? (instrumentation)

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

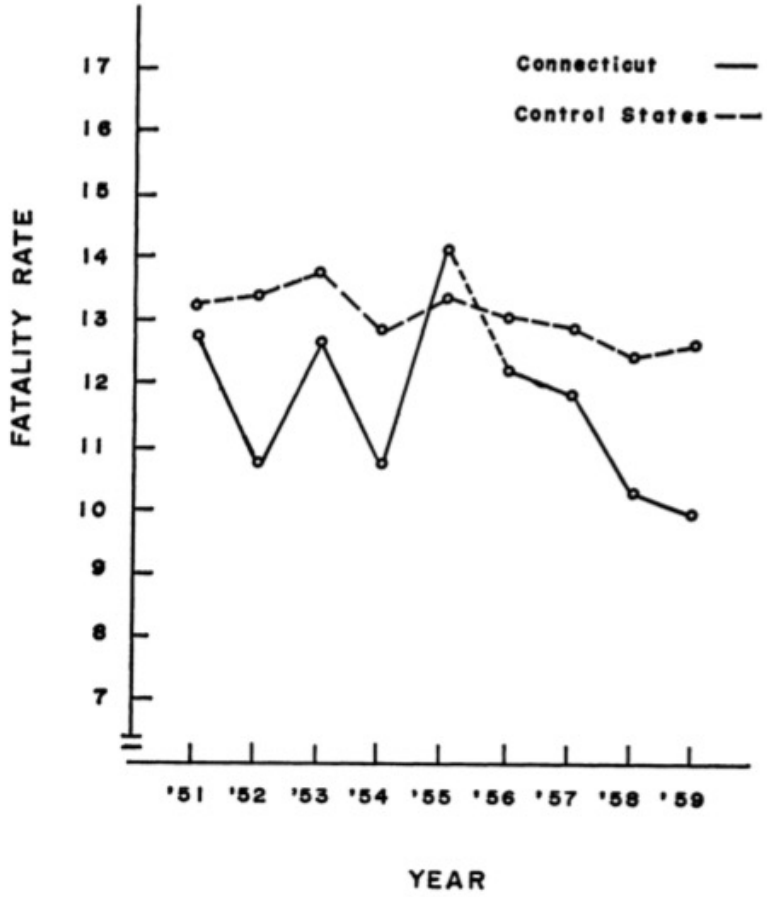


Figure 3. Connecticut and Control States Traffic Fatalities, 1951-1959 (per 100,000 population)

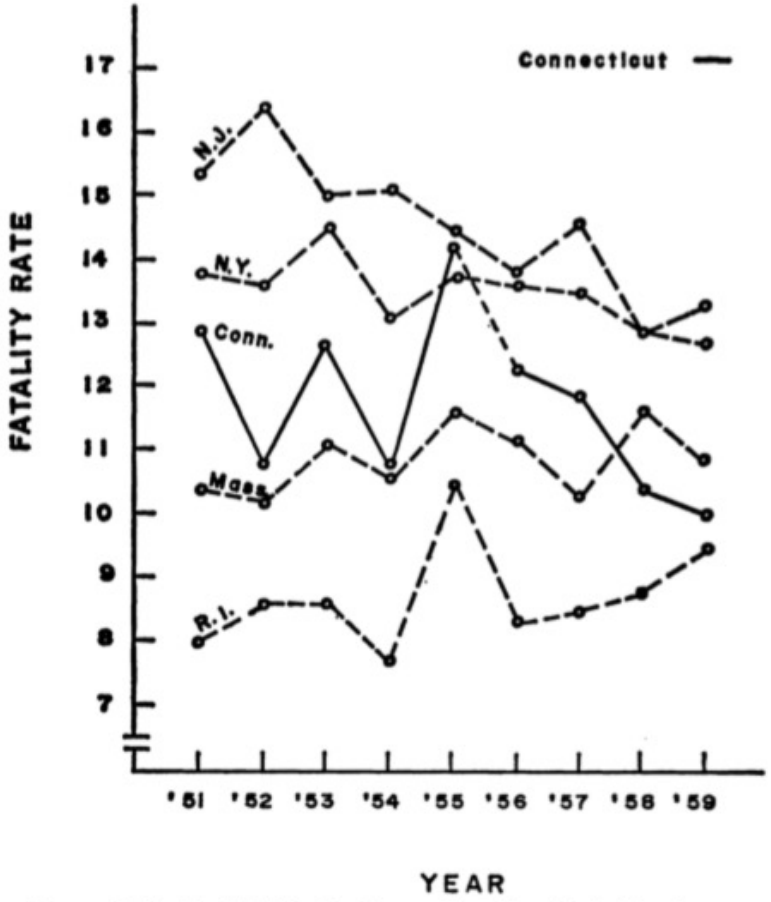
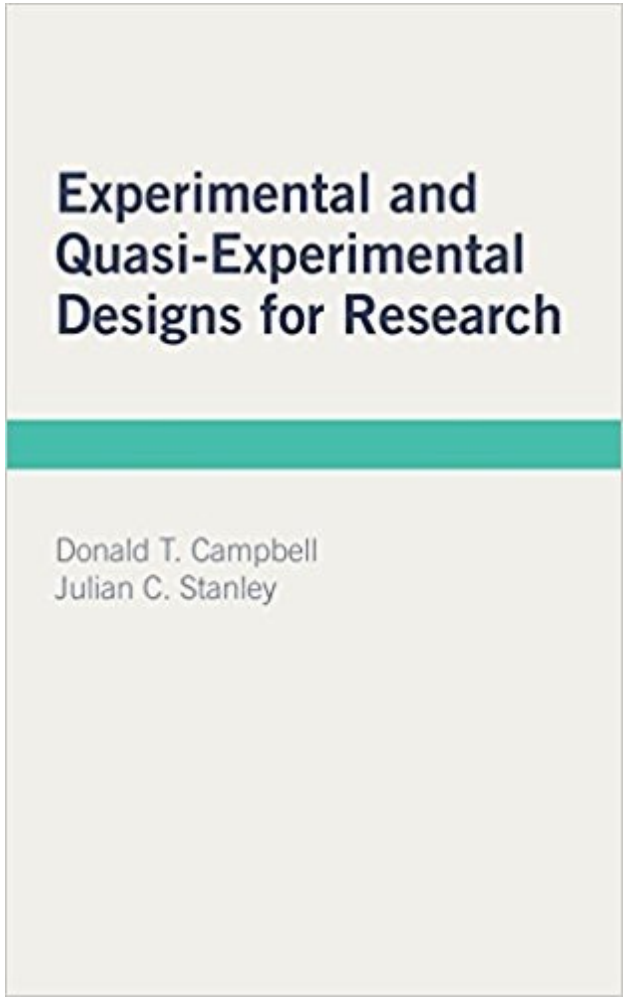


Figure 4. Traffic Fatalities for Connecticut, New York, New Jersey, Rhode Island, and Massachusetts (per 100,000 persons)

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.

Campbell & Stanley (1963)

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	−	−	+	?	+	+	−	−	+	+	+	
12a. R O (X) R X O --- R O (X) R X O	+	−	+	?	+	+	−	+	+	+	+	
12b. R O_1 (X) R O_2 (X) R X O_3	−	+	+	?	+	+	−	?	+	+	+	
12c. R O_1 X O_2 R X O_3	−	−	+	?	+	+	+	−	+	+	+	

Campbell & Stanley (1963)

TABLE 3

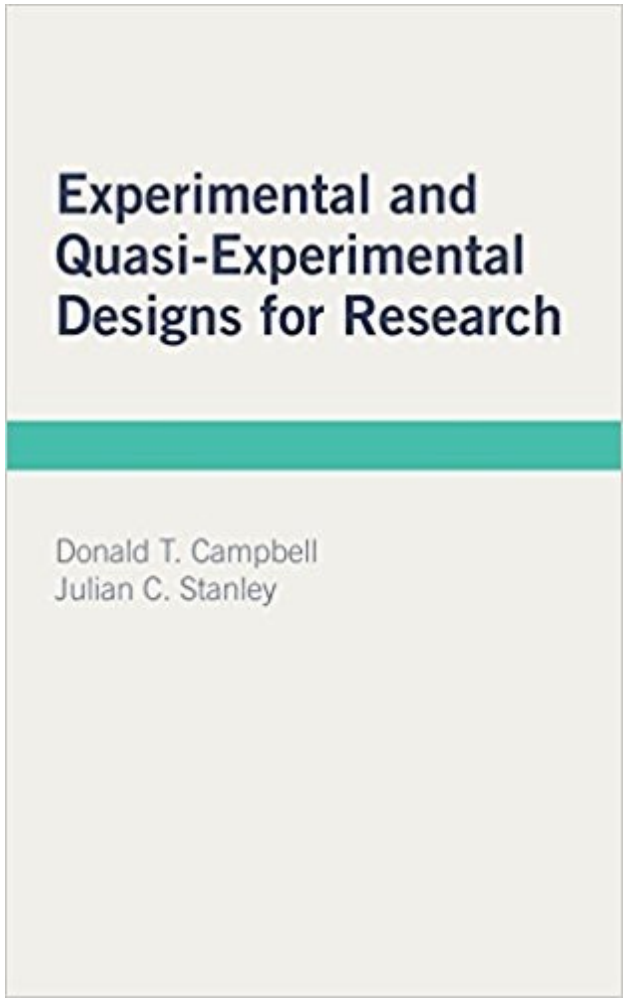
SOURCES OF INVALIDITY FOR QUASI-EXPERIMENTAL DESIGNS 13 THROUGH 16

	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 O (X) R X O --- R O (X) R X O --- R O (X) R X O 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 ===												
•Gen. Pop. Con. Cl. B O_6												
•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	+	+	+	?	+	+	?	+	+	−	+	+

• General Population Controls for Class B, etc.

Campbell & Stanley (1963)

Experimental and Quasi-experimental Designs



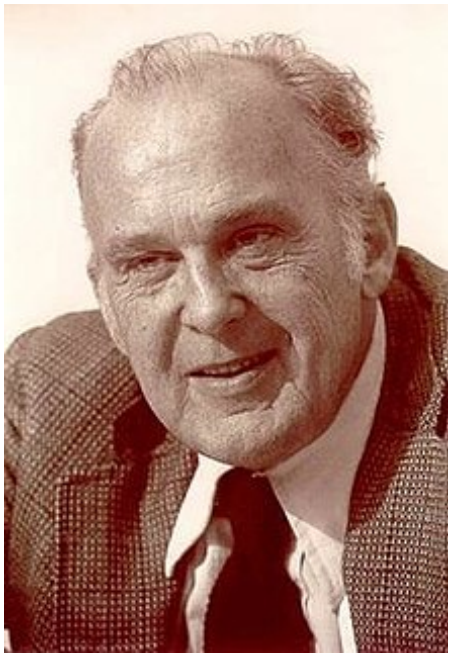
"In conclusion, in this chapter we have discussed alternatives in the arrangement or design of experiments, with particular regard to the problems of control of extraneous variables and threats to validity. (…) Through out, attention has been called to the possibility of **creatively** utilizing the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

The gold standard…

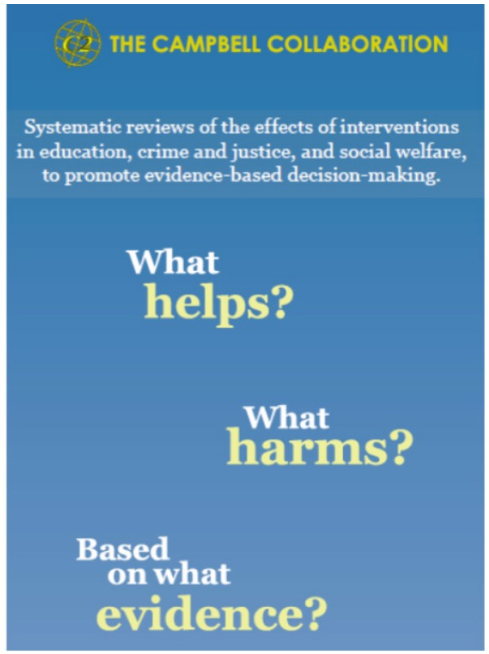
Experiments/Randomised control trials (RCT)

http://www.consort-statement.org

Alternatives…



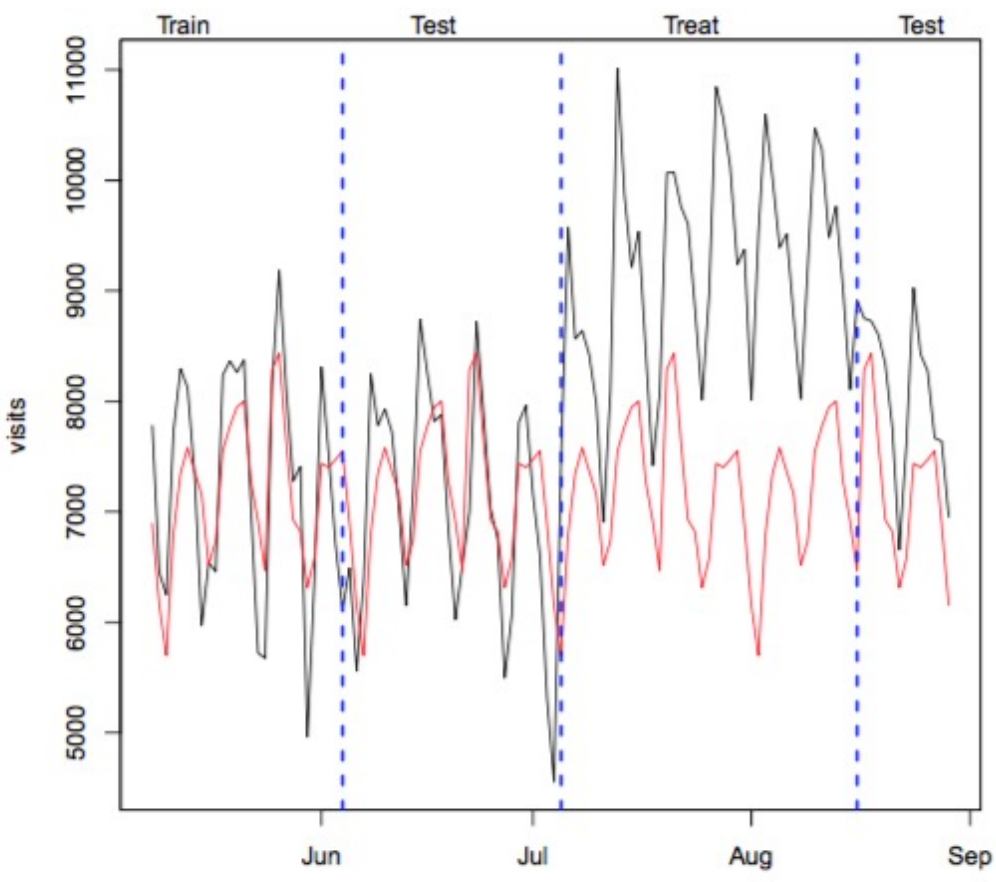
Donald Campbell
1916-1996



THE CAMPBELL COLLABORATION

New developments…

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