Quasi-Experimental Designs

Marek J. Drużdżel

Politechnika Białostocka
Wydział Informatyki

m.druzdzel@pb.edu.pl
http://www.wi.pb.edu.pl/~druzdzel/

Overview

- When experimentation is not feasible ...
- Internal and external validity of studies
- Threats to validity
- Quasi-experimental designs
- Concluding remarks

The necessity of empirical work

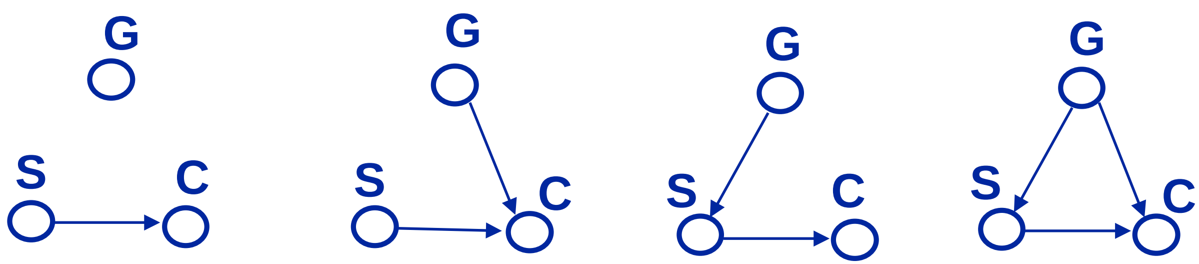
Smoking and lung cancer are correlated.
Can we reduce the incidence of lung cancer by reducing smoking?
In other words: Is smoking a cause of lung cancer?

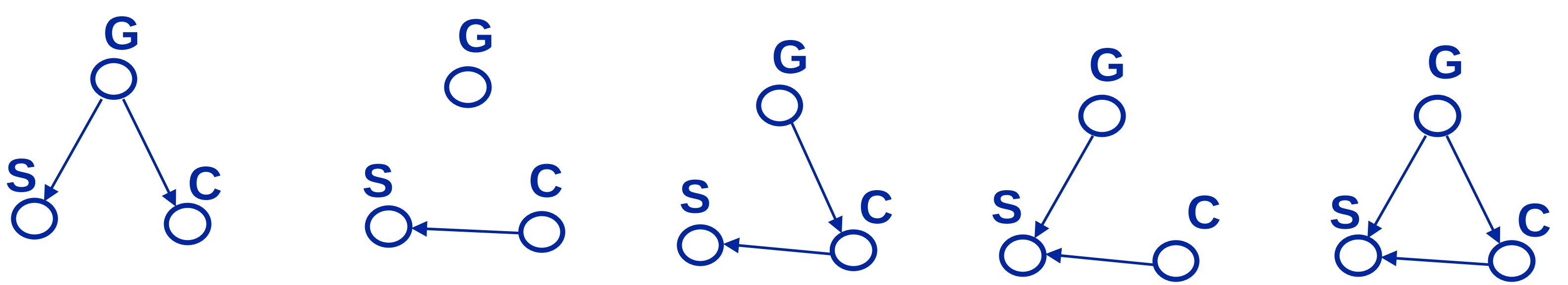
Each of the following causal structures is compatible with the observed correlation:

G = genetic factors

S = smoking

C = lung cancer





When experimentation is not feasible ...

Is an experiment testing the influence of smoking on incidence of lung cancer feasible?

No!!!

What should we do then?

Quasi-experimental designs

There are many natural settings in which the experimenter can introduce something like experimental design into her scheduling of data collection procedures, even though she lacks the full control over the scheduling of experimental stimuli which makes a true experiment possible.

Collectively, such situations can be regarded as quasi-experimental designs.
We should be thoroughly aware of situations where we lack full experimental control and learn to cope with these.

Quasi-experimental designs

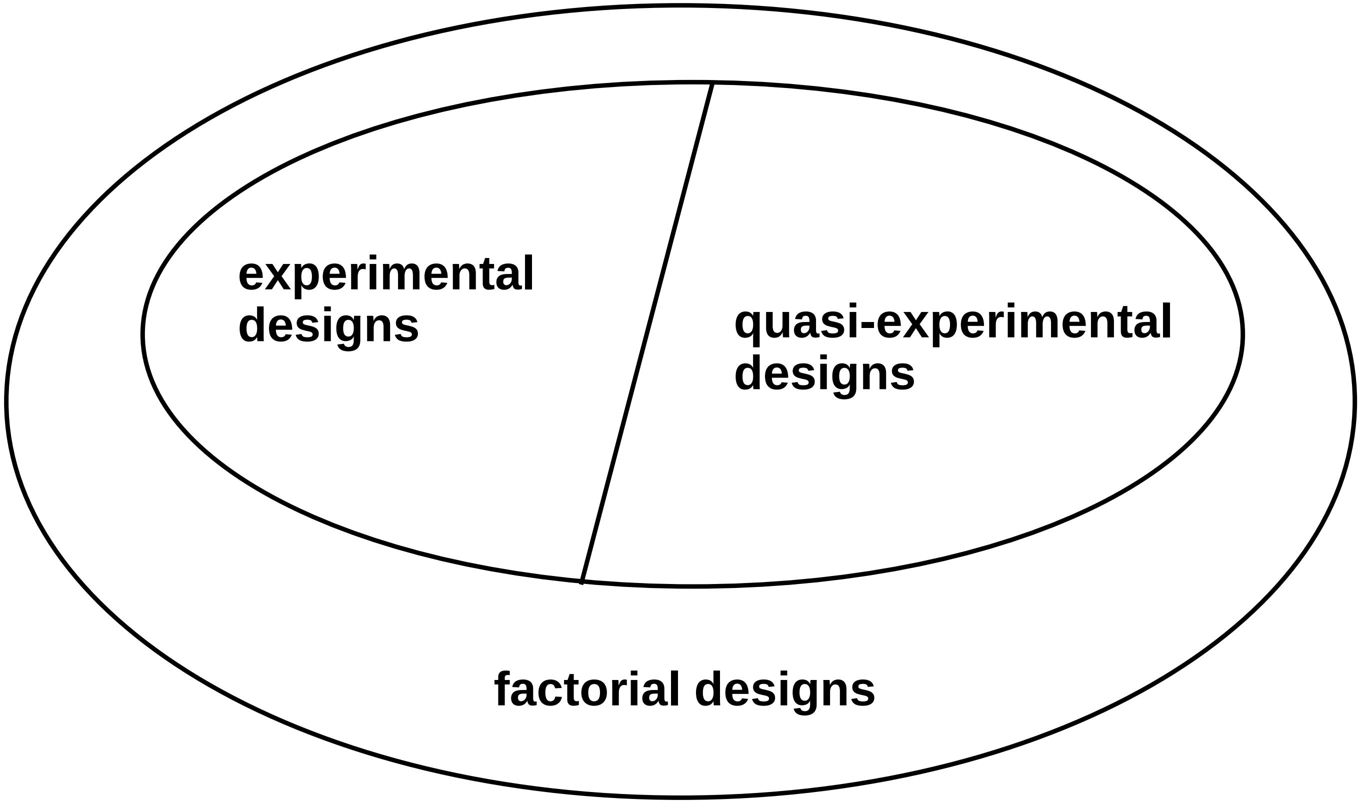
Two nice quotes from Campbell and Stanley:

> ***"The task of theory-testing data collection is ... predominantly one of rejecting inadequate hypotheses. In executing this task, any arrangement of observations for which certain outcomes would disconfirm theory will be useful, including quasi-experimental designs of less efficiency than true experiments."***

> ***"Varying degrees of 'confirmation' are conferred upon a theory through the number of plausible rival hypotheses available to account for the data. The fewer such plausible rival hypotheses remaining, the greater the degree of 'confirmation.'"***

Experimental designs roadmap

How are different experimental and quasi-experimental designs related?



Internal and external validity (review)

Internal validity: Did the experimental treatments make a difference in this specific experimental instance?

External validity: To what populations, settings, treatment variables, and measurement variables can this effect be generalized?

Threats to internal validity

1. **History**, the specific events occurring between the first and second measurement in addition to the experimental variable.
2. **Maturation**, processes within the respondents operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing more tired, and the like.
3. **Testing**, the effects of taking a test upon the scores of a second testing.
4. **Instrumentation**, in which changes in the calibration of a measuring instrument of changes in the observers or scorers used may produce changes in the obtained measurements.
5. **Statistical regression**, operating where groups have been selected on the basis of their extreme scores.
6. Biases resulting in differential **selection** of respondents from the comparison groups.
7. **Experimental mortality**, or differential loss of respondents from the comparison groups.
8. **Selection-maturation interaction**, etc., which in certain of the multiple-group quasi-experimental designs, such as Design 10, is confounded with, i.e., might be mistaken for, the effect of the experimental variable.

Threats to external validity

9. The reactive or interaction effect of testing, in which a pretest might increase or decrease the respondent's sensitivity or responsiveness to the experimental variable and thus make the results obtained for a pre-tested population unrepresentative of the effect of the experimental variable for the untested universe from which the experimental respondents were selected.
10. The interaction effects of selection biases and the experimental variable.
11. Reactive effects of experimental arrangements, which should preclude generalization about the effect of the experimental variable upon persons being exposed to it in non-experimental designs.
12. Multiple-treatment interference, likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable. This is a particular problem for one-group designs or type 8 or 9.

Symbolic notation (Campbell & Stanley)

Two-dimensional encoding of the elements of the design.

Symbols used:

X	Treatment
O	Measurement (testing)
R	Randomization

Time precedence:

Things happen from left to right (e.g., "X O" means that treatment precedes testing).

Multiple groups:

Each group is represented by a different line. Time precedence holds across lines.

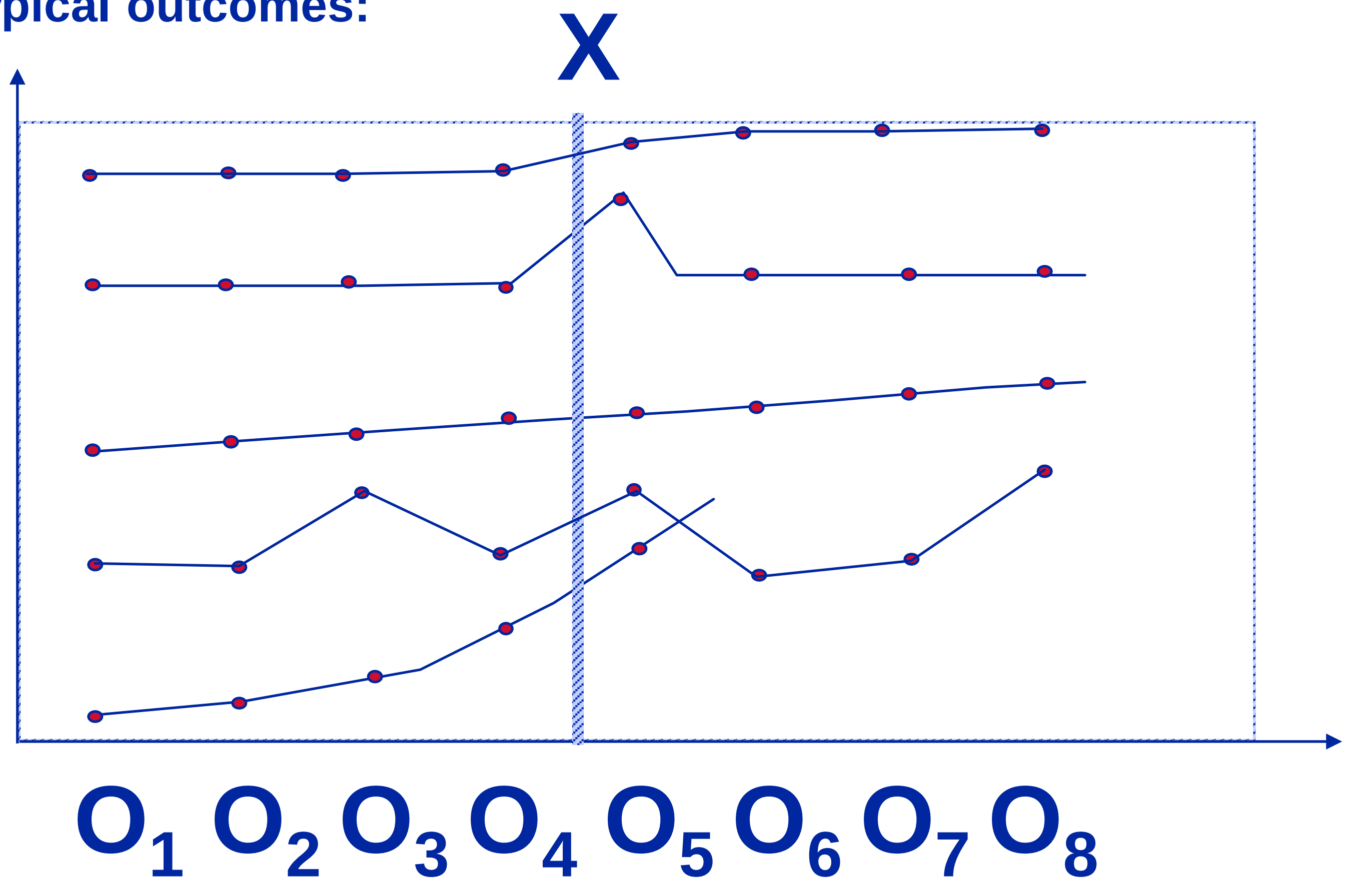
Design 7: The Time-Series Experiment

$$O_1 \; O_2 \; O_3 \; O_4 \; X \; O_5 \; O_6 \; O_7 \; O_8$$

The essence of the time-series design is the presence of a periodic measurement process on some group or individual and the introduction of an experimental change into this time series of measurements, the result of which are indicated by a discontinuity in the measurements recorded in the time series.

Design 7: The Time-Series Experiment

Typical outcomes:



Design 7: The Time-Series Experiment

$$O_1 \; O_2 \; O_3 \; O_4 \; X \; O_5 \; O_6 \; O_7 \; O_8$$

Main weaknesses:

- **History (can remedy somewhat by experimental isolation). (Note that maturation is pretty ruled out – why would it cause a shift exactly between O_4 and O_5. The same holds for instrumentation, regression, selection. Mortality only if looking at individuals and not at collective data.)**
- **External validity – the results may be specific to the population under study.**
- **Need to watch out the time between manipulation and effect (e.g., the effect of INFSCI 2040 on your academic careers ☺).**

Design 8: The Equivalent Time-Samples Design

$$X_1O \quad X_0O \quad X_1O \quad X_0O$$

A form of time-series experiment with the repeated introduction of the experimental variable (the alternation is intended to be random).

Most useful when the effect of the experimental variable is anticipated to be of transient or reversible character (e.g., influence of beer/wine/no alcohol on performance in theorem proving ☺).

Design 8: The Equivalent Time-Samples Design

$$X_1O \quad X_0O \quad X_1O \quad X_0O$$

History is controlled by presenting X on numerous separate occasions, rendering extremely unlikely any rival hypothesis based on coincidence of extraneous events.
Example: Influence of type of music that cows are exposed to on the amount of milk.

Main weakness:

- **External validity threat: multiple-X (treatment) interference**

Threats to external validity

9. The reactive or interaction effect of testing, in which a pretest might increase or decrease the respondent's sensitivity or responsiveness to the experimental variable and thus make the results obtained for a pre-tested population unrepresentative of the effect of the experimental variable for the untested universe from which the experimental respondents were selected.
10. The interaction effects of selection biases and the experimental variable.
11. Reactive effects of experimental arrangements, which should preclude generalization about the effect of the experimental variable upon persons being exposed to it in non-experimental designs.
12. **Multiple-treatment interference**, likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable. This is a particular problem for one-group designs or type 8 or 9.

Design 8: Multiple-X interference

$$X_1O \quad X_0O \quad X_1O \quad X_0O$$

The effect of X_1, in the simplest situation in which it is being compared with X_0, can be generalized only to conditions of repetitious and spaced presentations of X_1. No sound basis is provided for generalization to possible situations in which X_1 is continually present, or to the condition in which it is introduced once and once only.
e.g., classical music leads to producing more milk only when it is interspersed with hard rock ☺.

Design 9: The equivalent Materials Design

$$M_a X_1 O \quad M_b X_0 O \quad M_c X_1 O \quad M_d X_0 O \; \ldots$$

A refinement on the equivalent time-samples design: A single group of subjects is subjected to treatments that are equivalent to one another. This is used when the effect of the treatment is lasting and the stimulus should not be repeated.

Examples:

- **Learning nonsense syllables (need to change the stimulus at each stage)**
- **Testing the influence of group opinion (Xs) on voting schemes (Os) (testing conformity of the subjects). Ms are equivalent questionnaires. In-between the subjects get (falsified) feedback about group opinion.**

Design 10: The Nonequivalent Control Group Design

O	X	O
O		O

Two groups of subjects: both are given a pretest and a posttest, but the control group and the experimental group do not have pre-experimental sampling equivalence. Rather the groups constitute naturally assembled collectives such as classrooms, as similar as availability permits but yet not so similar that one can dispense with the pretest. The assignment of X to one group or the other is assumed to be random and under the experimenter's control.

Do not confuse this design with Design 4, the Pretest-Postest Control Group Design in which the subjects are assigned randomly from a common population to the experimental and the control groups!

Design 10: The Nonequivalent Control Group Design

O	X	O
O		O

Main weaknesses:

- **Regression - when one group has been chosen based on extreme scores.**
- **Maturation (especially interaction between selection and maturation, e.g., psychotherapy patients vs. normal population – their remission may be simply due to regression that happens to an extreme group).**
- **Interaction of selection and X is a possible threat to external validity.**

Design 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

A class of designs in which experimental control is achieved or precision enhanced by entering all respondents (or settings) into all treatments. Also called "rotation experiments," "cross-over designs," and "switch-over designs."

Design 11: Counterbalanced Designs

$$
\begin{matrix}
X_1O & X_2O & X_3O & X_4O \\
X_2O & X_4O & X_1O & X_3O \\
X_3O & X_1O & X_4O & X_2O \\
X_4O & X_3O & X_2O & X_1O
\end{matrix}
$$

Main weaknesses:

- **The effect can be possibly attributed to complex interactions among treatments.**
- **Possible effects of repeated testing, history, maturation, practice, and cumulative carryovers or transfer.**

Design 12: The Separate-Sample Pretest-Posttest Designs

R	O	(X)	
R		X	O

Sometimes called "simulated before-and-after design"

Used when we cannot randomly segregate groups, but we can fully control over the "when" and "to whom" of the O, employing random assignment procedures. (X) stands for a presentation of X irrelevant to the argument. One sample is measured prior to the X, an equivalent one subsequent to X.

Used typically for large populations, such as cities, factories, schools, and military units.

Design 12: The Separate-Sample Pretest-Posttest Designs

R	O	(X)	
R		X	O

Example application:

- Cincinnati publicity campaign for the United Nations and UNESCO (Star & Hughes, 1950)

Main strength:

- A great improvement on Design 2 (controls both the effect of testing and interaction of testing with X).

Main weaknesses:

- History, instrumentation, mortality (if the pretest and posttest are separated in time by several months).

Design 13: The Separate-Sample Pretest-Posttest Control Group Design

R	O	(X)	
R		X	O
---	---	---	---
R	O		
R			O

An improvement on the previous design that allows to compare the treatment group to a control group. Similar to Design 10, but the same specific subjects are not re-tested and thus the possible interaction of testing and X is avoided.
Probably never used (according to Campbell & Stanley).

Design 14: The Multiple Time-Series Design

O O O O X O O O O

O O O O   O O O O

A control group (not randomized) helps to fix some problems of the time-series design.

Alternative Designs

Let us not forget about the plentitude of other creative possibilities ...

- **Archimedes (king's golden crown).**

http://azriazroie.blogspot.com/

Alternative Designs

Let us not forget about the plentitude of other creative possibilities ...

- **Witch trials (trial by water, trial by fire).**

Alternative Designs

Let us not forget about the plentitude of other creative possibilities ...

King Solomon's judgment

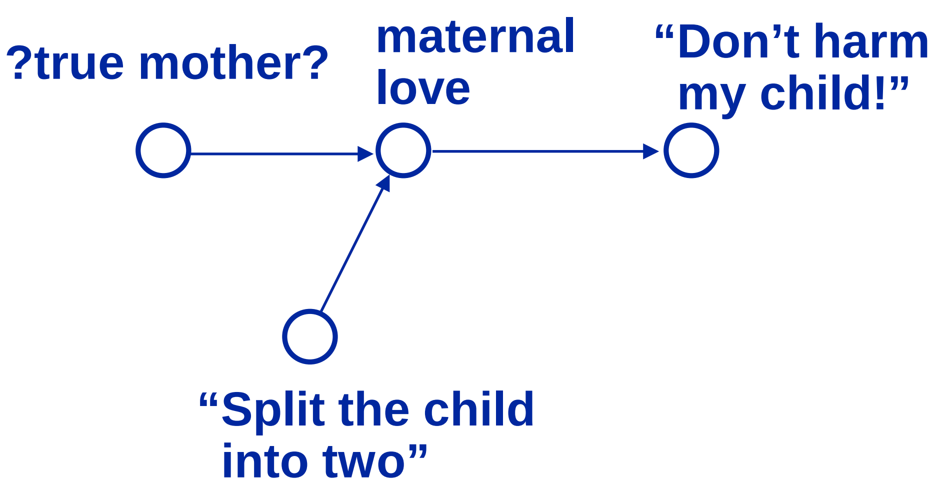
http://callofthepatriot.blogspot.com/2010_12_01_archive.html

Alternative Designs

Let us not forget about the plentitude of other creative possibilities ...

King Solomon's judgment

"Split it into three to include my consultancy fee."



Concluding Remarks

- The experimental and quasi-experimental are two sides of the same coin. Of course, experimental are stronger, but they are after the same thing, which should be clear from studying causal graphs. Spirtes et al., for example, discover the causal graph purely from observations. Factorial designs just add one more dimension and can be both experimental and quasi-experimental (i.e., one can apply ANOVA to both).
- It is a good idea to replicate studies to control for such factors as history.
- Use common sense in addition to your knowledge of statistics and experimental design.
- It is a good idea not to feel bound by the existing experimental approaches. Be creative, it pays. As long as you make a good argument, people will buy your design. Each of these designs was introduced at some point and there is no guarantee that they are all that there is.
- The main purpose of this course is to make you a critical recipient of research results and to make you see what is going on in a design.