Introduction to Interrupted Time Series Designs

June 1, 2016

Steve Gregorich

Main Goals

- Brief(ish) introduction to the interrupted time series (ITS) design

- Comparing ITS to some other quasi-experimental designs
  that do not include control groups.

Outline

- First look at an Interrupted Time Series (ITS)
- Gold standard: The randomized controlled trial design
- Generic threats to internal validity: RCTs
- Some quasi-experimental designs (QED) with no control group
- Generic threats to internal validity: QEDs
- The ITS design; real examples of some archetypal outcomes
- Bolstering the ITS design
- ITS analysis, very briefly
- Summary

First look at an Interrupted Time Series Design

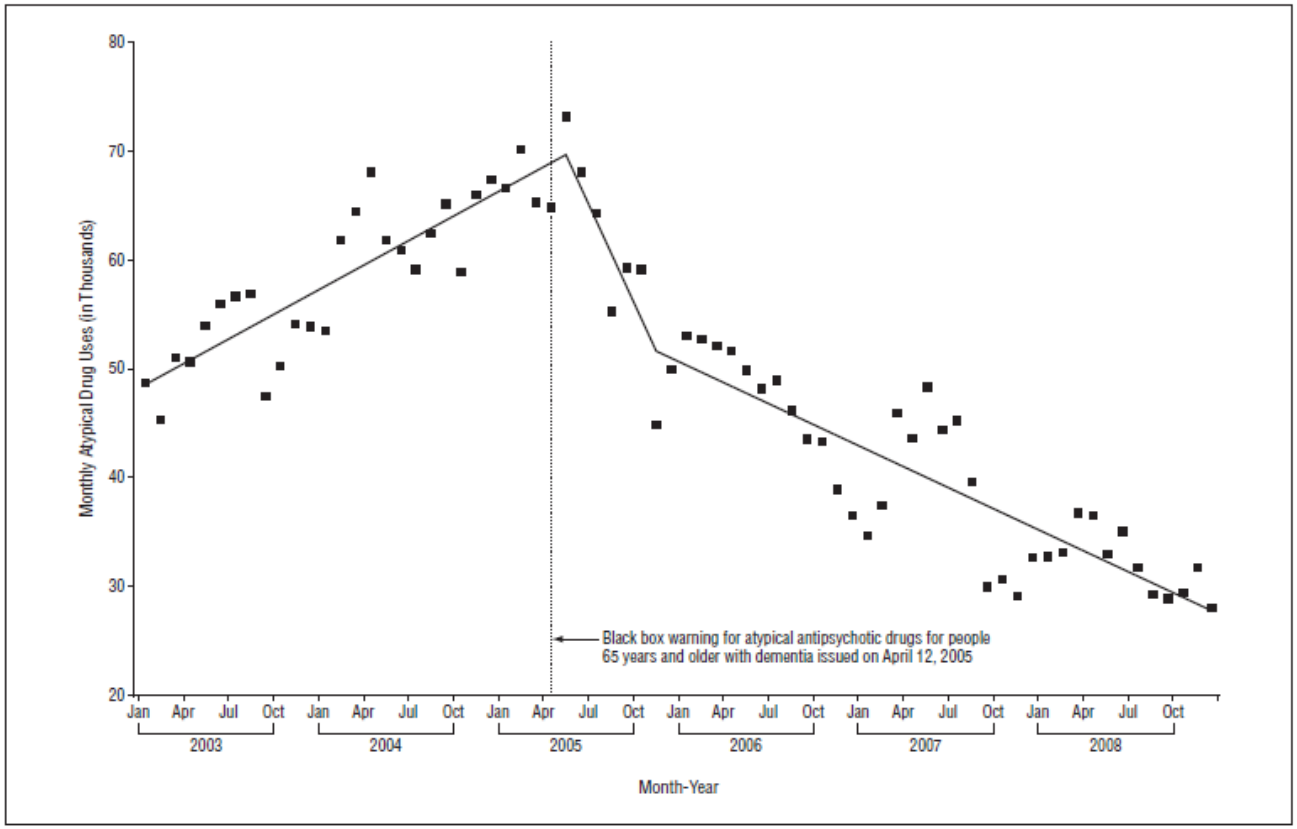


Figure 2. Joinpoint Regression Program analysis for atypical antipsychotics use among elderly patients with dementia. The data points represent patients 65 years and older with dementia (smoothed 6-month averages); the solid line, fitted joinpoint time series.

2010. Dorsey, RE et al. *Arch Intern Med*, 170, 96–103

April 2005: FDA issued an advisory and black box warning

Risks of ↑ mortality: atypical anti-psychotic use: elderly patients w/ dementia

At the time, the impact of the warning on atypical drug use was unknown

Gold Standard: The Randomized Controlled Trial Design

$$\text{Rnd} \begin{cases} \text{Intv:} & O_{t1} & \mathbf{Tx} & O_{t2} \\ \text{Ctrl:} & O_{t1} & & O_{t2} \end{cases}$$

- Rnd: Equivalent groups at t_1.

- If 'closed-system' maintained,
  then solid basis for causal inference about Tx effects

I.e., internal validity

Generic Threats to Internal Validity

Focal (for today)

- **Selection**: participant characteristics systematically differ across groups
- **History**: events acting upon population & co-occurring with Tx
- **Maturation**: natural changes in sampled Pts across time
- **Testing**: repeated exposure to a test may affect assessment

Generic Threats to Internal Validity

Others—almost universally problematic

- **Instrumentation**: the nature of a measure changes across time, such that the validity of repeated assessments may be questioned

- **Ambiguous temporal sequencing of variables**: $X \rightarrow Y$, or $Y \rightarrow X$?

- **Regression**: Pts with initial extreme values may 'regress'

- **Attrition**: if systematically correlated with Tx or outcomes

All threats (Focal and Others) can combine additively or interactively

RCT and 'Focal' Threats to Internal Validity

$$\text{Rnd} \begin{cases} \text{Intv:} & O_{t1} & \mathbf{Tx} & O_{t2} \\ \text{Ctrl:} & O_{t1} & & O_{t2} \end{cases}$$

Selection: randomization should address

History: synchronized assessments should address

Maturation: randomization & synchronized assessments should address

Testing: parallel assessment schedule should address

RCT and 'Other' Threats to Internal Validity

$$\text{Rnd} \begin{cases} \text{Intv:} & O_{t1} & \mathbf{Tx} & O_{t2} \\ \text{Ctrl:} & O_{t1} & & O_{t2} \end{cases}$$

Instrumentation: addressed, as long as measures are relevant to targeted constructs

Ambiguous temporal sequencing: longitudinal design addresses

Regression: randomization and parallel assessments should address, even if extreme groups are targeted for recruitment

Attrition: always a concern; to be dealt with in a principled fashion

Some Longitudinal QED designs w/ no Control Group

Often, QI study designs do not employ a control group

	One Sample, Longitudinal	Multiple-Cross Sections
pretest-posttest	O_{t1} **Tx** O_{t2}	O_{t1} **Tx** O_{t2}
pre-post w/ multi-pre	O_{t0} O_{t1} **Tx** O_{t2}	O_{t0} O_{t1} **Tx** O_{t2}
repeated Tx	O_{t1} **Tx** O_{t2} ~~**Tx**~~ O_{t3} **Tx** O_{t4}	O_{t1} **Tx** O_{t2} ~~**Tx**~~ O_{t3} **Tx** O_{t4}

- Many other designs exist

Summary: Internal Validity Threats w/ no Control Group

One-Sample, Longitudinal QEDs

	selection	history	maturation	testing
O_{t1} Tx O_{t2}		x	x	x
O_{t0} O_{t1} Tx O_{t2}		x	reduced	x
O_{t1} Tx O_{t2} ~~Tx~~ O_{t3} Tx O_{t4}		greatly reduced		x

Multiple-Cross Sectional, QEDs

	selection	history	maturation	testing
O_{t1} **Tx** O_{t2}	x	x	x	
O_{t0} O_{t1} **Tx** O_{t2}	x	x	reduced	
O_{t1} **Tx** O_{t2} ~~**Tx**~~ O_{t3} **Tx** O_{t4}	x	greatly reduced		

. Trade-off: Selection v History (and Power)

The Interrupted Time Series Design

- Longitudinal

O_{t1} O_{t2} O_{t3} O_{t4} O_{t5} **Tx** O_{t6} O_{t7} O_{t8} O_{t9} O_{t10}

- Multiple cross-section

O_{t1} O_{t2} O_{t3} O_{t4} O_{t5} **Tx** O_{t6} O_{t7} O_{t8} O_{t9} O_{t10}

Either way, it *can* be a strong design

ITS Example 1: Charging for directory assistance (DA)

- A change in level at intervention onset (March 1974). Y-axis: # calls

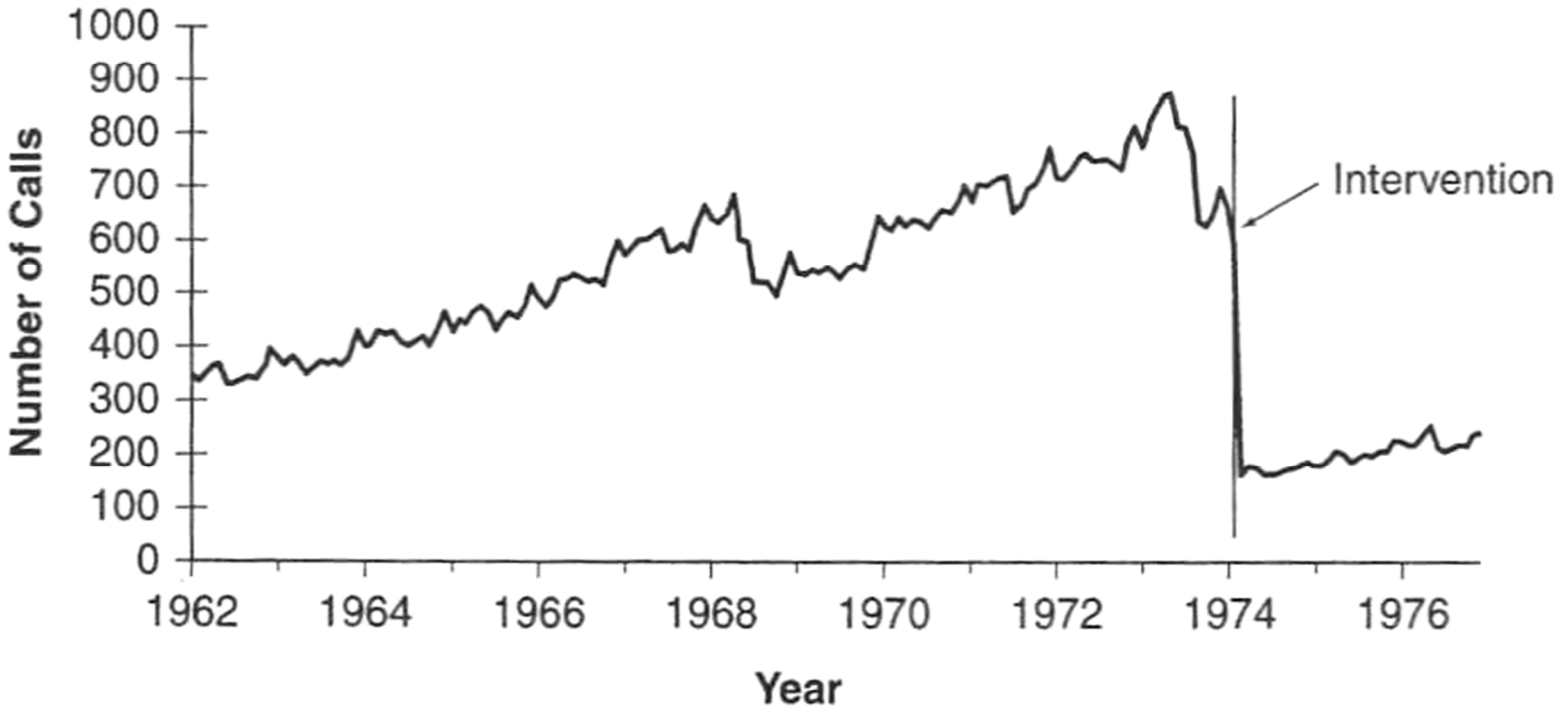


FIGURE 6.1 The effects of charging for directory assistance in Cincinnatii

[Figure from Cook & Campbell (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*]

ITS Example 1: Charging for directory assistance (DA)

- Immediate large drop in number of calls, March 1974

Selection implausible:
pre and post samples likely the same

Attrition implausible
New charges unlikely to prompt phone disconnections

Maturation implausible
no known maturation process could account for drop in calls

History implausible
unless another hypothetically causal event can be identified

Testing implausible
E.g., if phone co. changed salience of DA charges on phone bills

Regression to the mean implausible:
pre- trend suggested high call rates for many years

ITS Example 2: New Law Re. Sexual Assault Reporting

- Change in slope at intervention onset. Y-axis: # reported sexual assaults

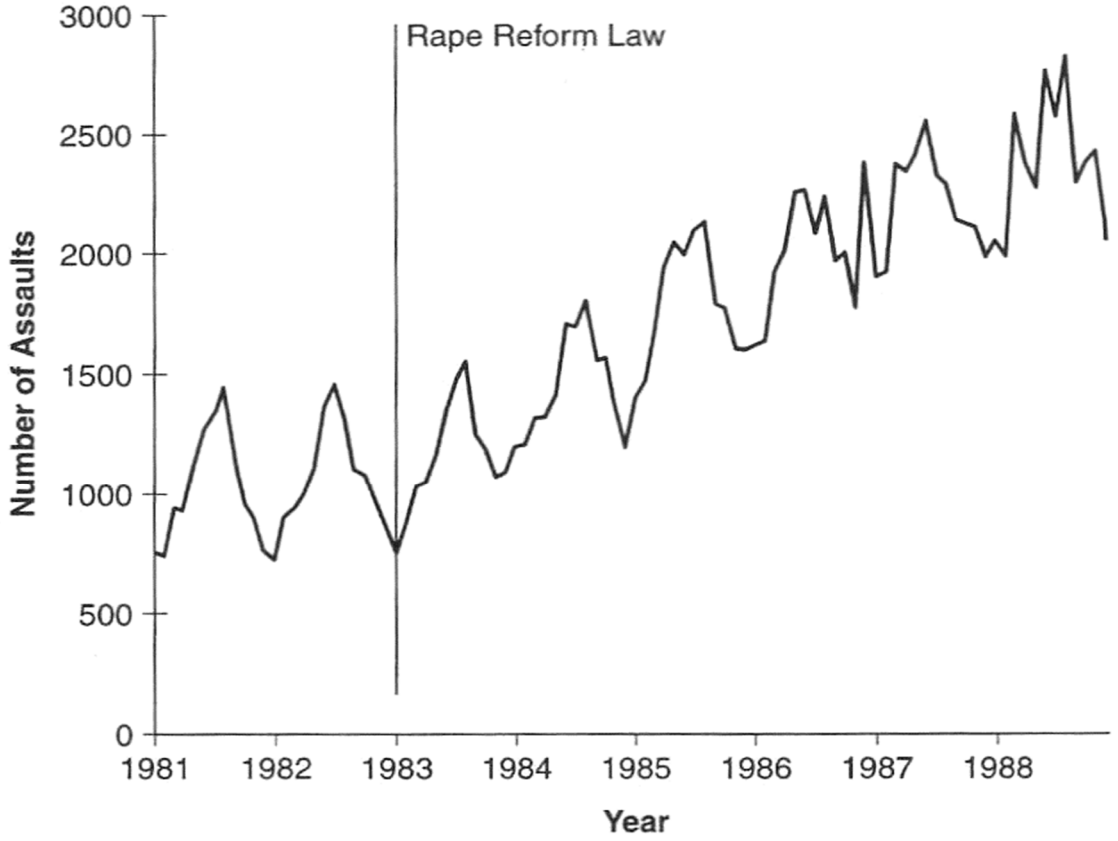


FIGURE 6.2 The effects of sexual assault law reform in Canada

From "Reforming rape laws: Effects of legislative change in Canada," by J. V. Roberts and R. J. Gebotys, 1992, *Law and Human Behavior, 16,* 555–573. Copyright 1992 by Kluwer Academic/Plenum Publishers.

[Figure from Cook & Campbell (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*]

ITS Example 2: New Law Re. Sexual Assault Reporting

- Immediate change in slope from flat to positive, 1983

Maturation implausible
no known maturation process could account for change in slope

History implausible
unless another hypothetically causal event can be identified

Instrumentation possible.
The new law changed the categories of reportable sexual assault
1. wives could charge husbands with sexual assault
2. included assaults against both males and females

Authors showed that, in the post-intervention period,
suspects who were women or husbands
did not increase sufficiently to explain the pattern of results.

ITS Example 3: Alcohol warning label re. prenatal drinking

- Weak, Delayed, Ambiguous Effects. Y-axis: Prenatal Drinking Score

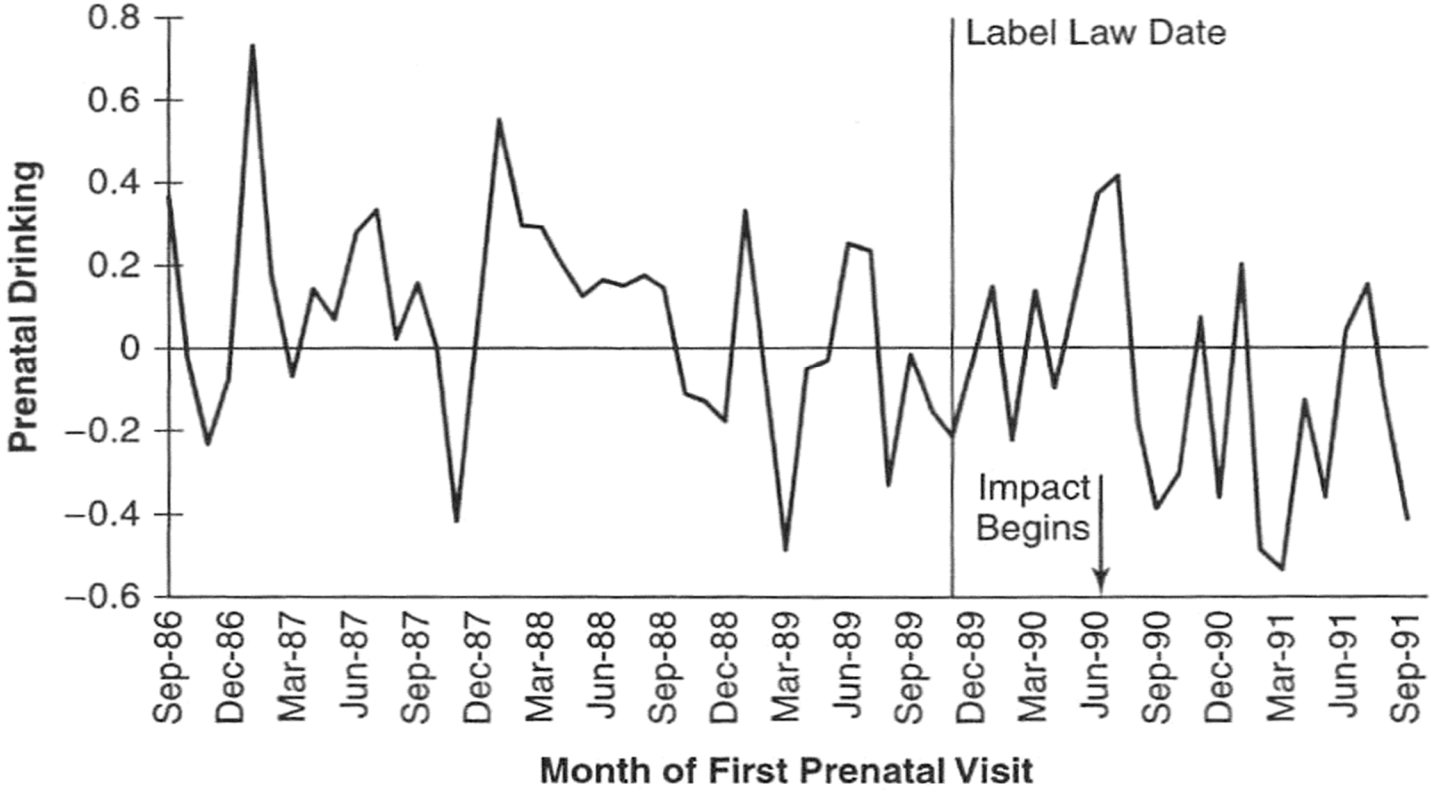


FIGURE 6.3 The effects of an alcohol warning label on prenatal drinking

From "A time series analysis of the impact of the alcohol warning label on antenatal drinking," by J. R. Hankin et al., 1993, *Alcoholism: Clinical and Experimental Research, 17,* pp. 284–289. Copyright 1993 by Lippincott, Williams &

[Figure from Cook & Campbell (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*]

ITS Example 4: Pay-for-performance & BP control

- No effect observed

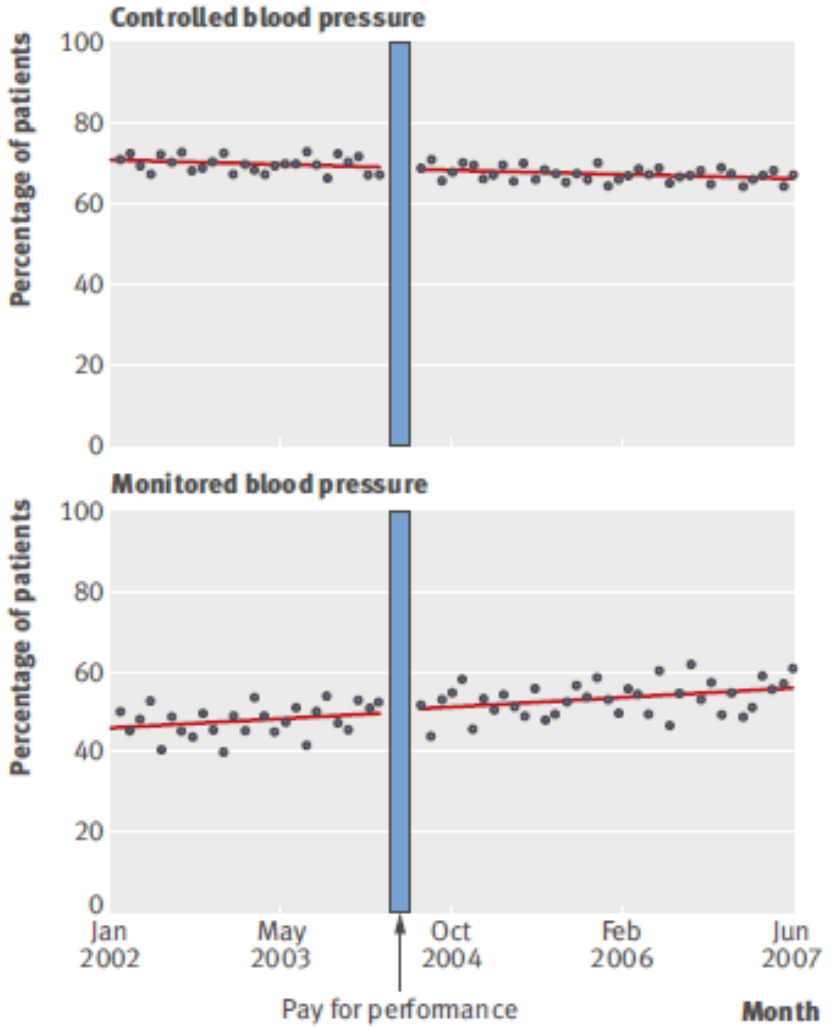


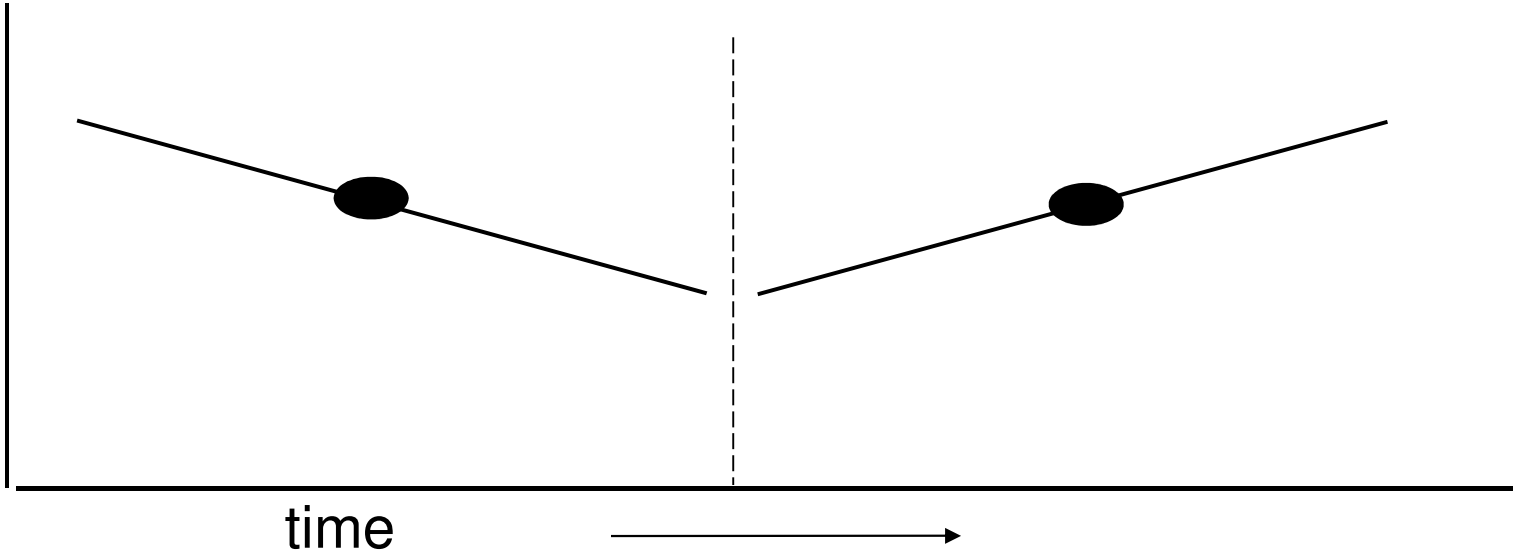
Fig 2| Effect of pay for performance on blood pressure control and monitoring in United Kingdom

ITS advantages over pre-test / post-test design: Simplified

Scenario #1: intervention effect observed: immediate change in slope

. ITS would identify the intervention effect

. A simple pre-post test design would not.
Comparing the pre- and post- means (black dots) suggests no overall pre-post difference

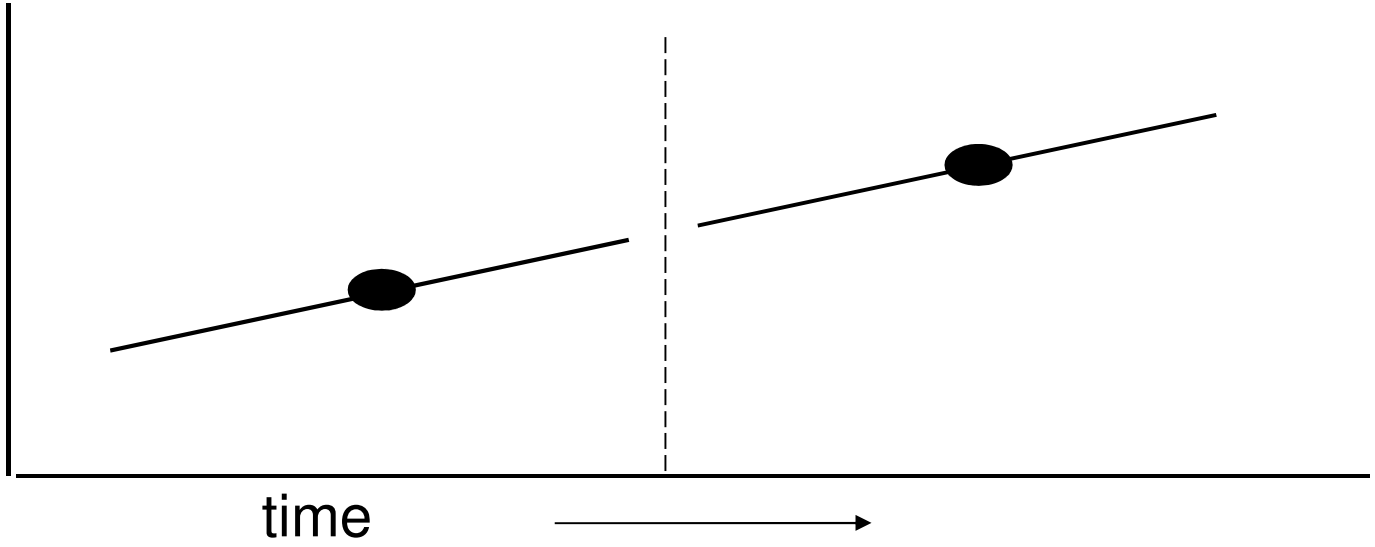


ITS advantages over pre-test / post-test design: Simplified

Scenario #2: no intervention effect

. ITS would identify the lack of intervention effect

. A simple pre-post test design would suggest an intervention effect.
Comparing pre- and post- means (black dots) suggests
a post-test increase in outcome level



Summary, So Far

- ITS design can provide a good basis for drawing causal inferences if…
  - . observed changes are well timed with intervention onset
  - . alternative explanations (threats to internal validity) are *implausible*
- However, even under those circumstances
  threats to internal validity may still operate,
  e.g., the seemingly implausible may obtain
- Next: ways to bolster the ITS design

Bolstering the ITS Design

- Non-equivalent no-treatment control group
- Non-equivalent dependent variables
- Removing a treatment at a known time
- Multiple replications
- Switching replications

Bolstering the ITS Design

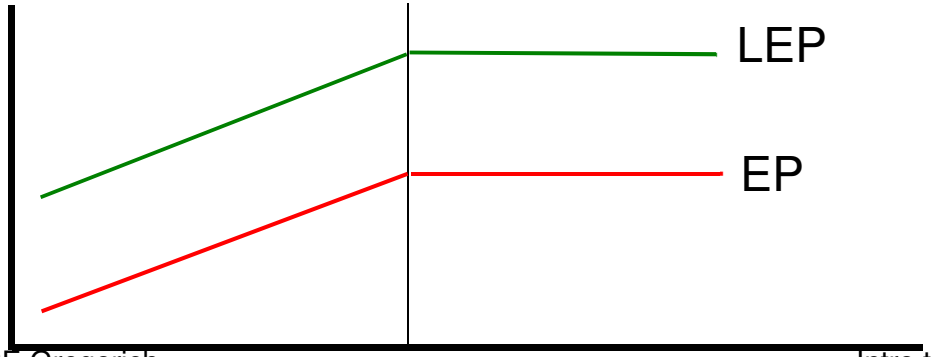
- **Non-equivalent no-treatment control group**

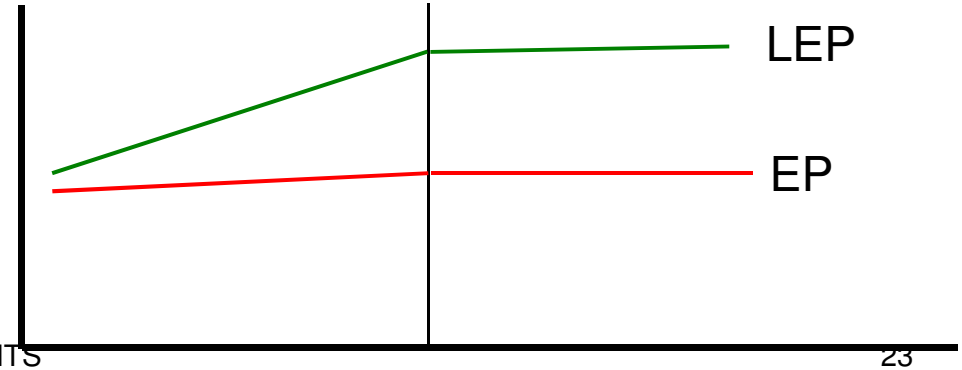
O_{t1}	O_{t2}	O_{t3}	O_{t4}	O_{t5}	**Tx**	O_{t6}	O_{t7}	O_{t8}	O_{t9}	O_{t10}
O_{t1}	O_{t2}	O_{t3}	O_{t4}	O_{t5}		O_{t6}	O_{t7}	O_{t8}	O_{t9}	O_{t10}

- I.e., add a group hypothetically unaffected by the intervention

- Example: L. Karliner (PI), in progress:
  Impact of hospital "bedside interpreter" on LEP patient outcomes
  (add a non-equivalent no-treatment control group of EP patients)

Most notably, this helps to diagnose **history** threats (made-up examples)





Bolstering the ITS Design

- **Non-equivalent dependent variables**

  Add an outcome hypothesized to be unaffected by the intervention, but that is hypothesized to be equally subject to validity threats

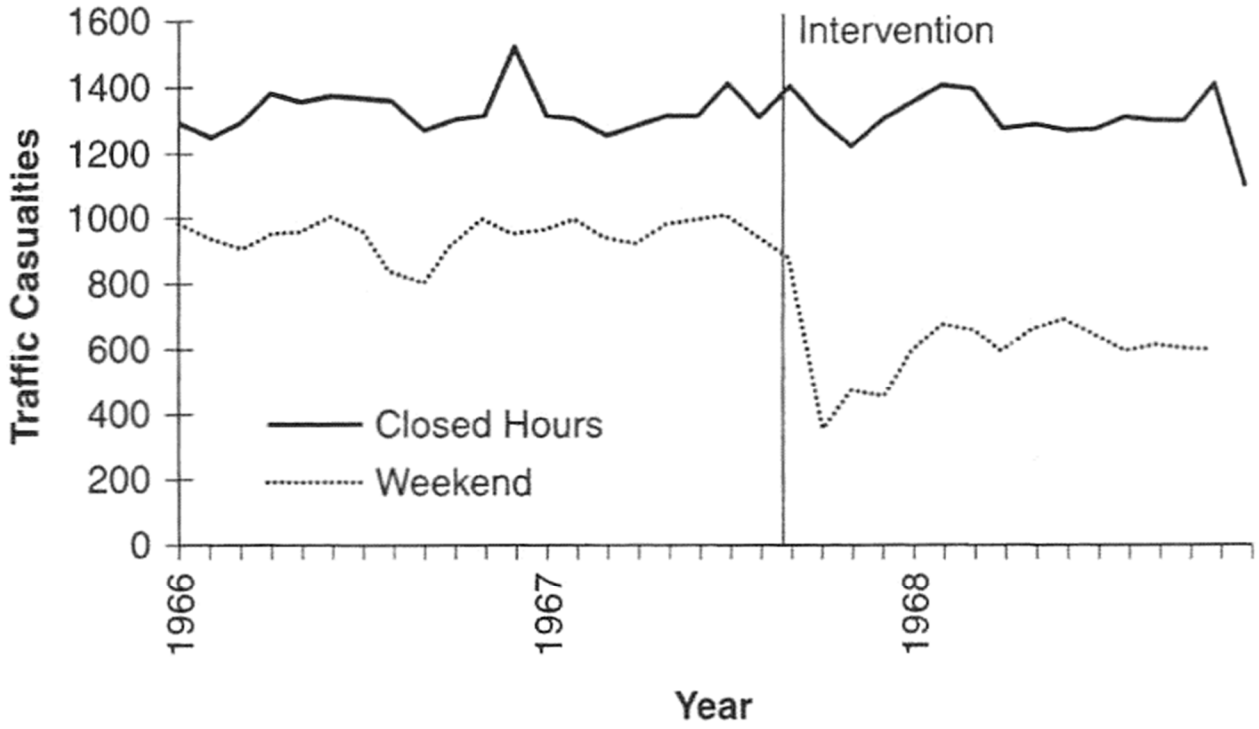


FIGURE 6.6 The effects of the British Breathalyzer crackdown on traffic casualties during weekend nights when pubs are open, compared with times when pubs were closed

From "Determining the social effects of a legal reform: The British 'breathalyser' crackdown of 1967," by H. L. Ross, D. T. Campbell, and G. V. Glass, 1970, *American Behavioral Scientist, 13,* pp. 493–509. Copyright 1970 by Sage

Bolstering the ITS Design

- **Removing a treatment at a known time**

O_{t1} O_{t2} O_{t3} O_{t4} **Tx** O_{t5} O_{t6} O_{t7} ~~**Tx**~~ O_{t8} O_{t9} O_{t10}

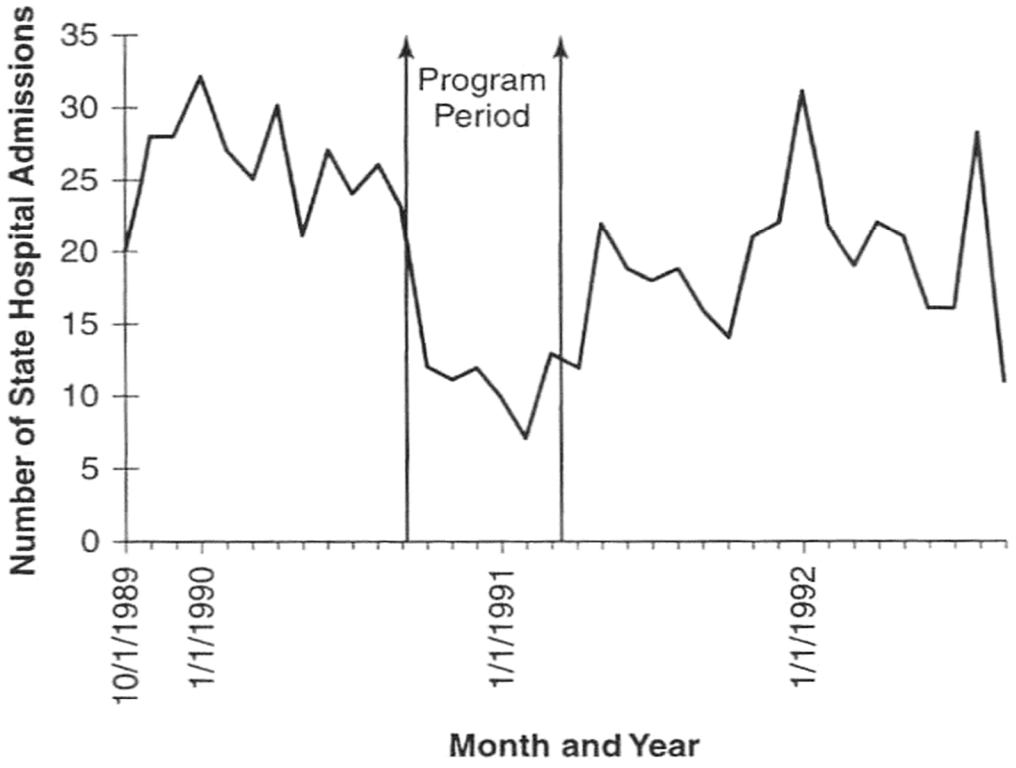


FIGURE 6.8 The effects of psychiatric crisis intervention on hospitalization

From "Around-the-clock mobile psychiatric crisis intervention: Another effective alternative to psychiatric hospitalization," by G. R. Reding and M. Raphelson, 1995, *Community Mental Health Journal, 31,* pp. 179–187.

[Figure from Shadish, Cook & Campbell (2001). *Experimental & Quasi-Experimental Design for Generalized Causal Inference*]

Bolstering the ITS Design

- **Adding multiple replications**

O_{t1} O_{t2} **Tx** O_{t3} O_{t4} ~~**Tx**~~ O_{t5} O_{t6} **Tx** O_{t7} O_{t8} ~~**Tx**~~ O_{t9} O_{t10}

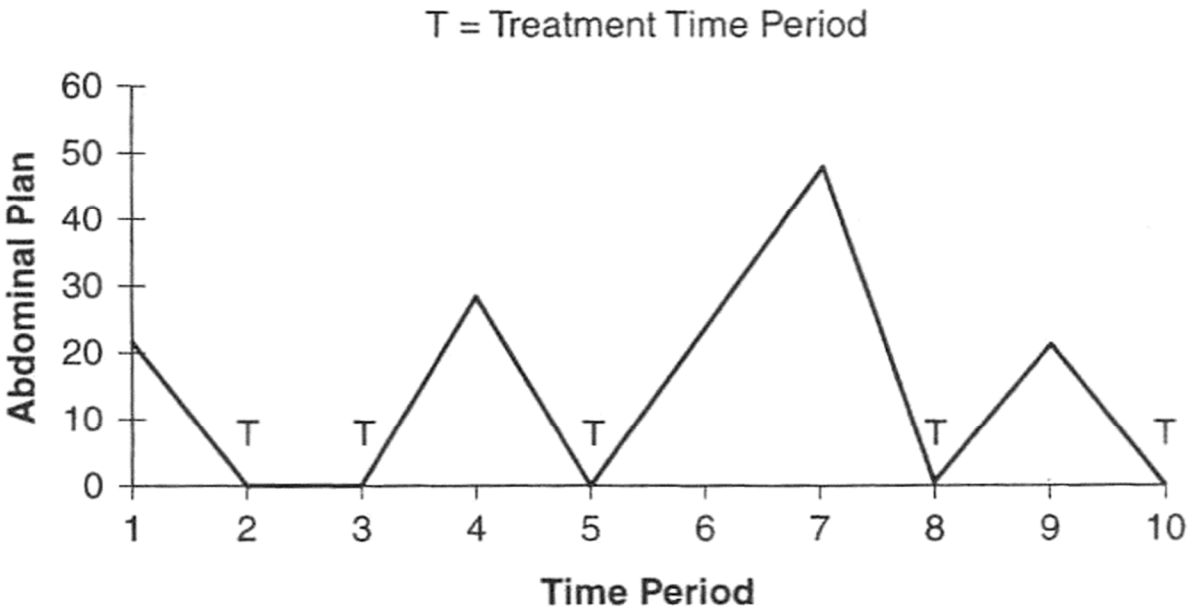


FIGURE 6.9 The effects of treatment for inflammation of continent ileostomy. In the graphs, the letter *T* indicates the time period during which treatment occurred

From "Single patient randomized clinical trial: Its use in determining optimal treatment for patient with inflammation of a Kock continent ileostomy reservoir," by R. S. McLeod et al., 1986, *Lancet, 1,* pp. 726–728.

[Figure from Cook & Campbell (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*]

Bolstering the ITS Design

- **Adding switching replications**
  2 or more nonequivalent groups w/ staggered intervention introduction

O_{t1} O_{t2} O_{t3} **Tx** O_{t4} O_{t5} O_{t6} O_{t7} O_{t8} O_{t9} O_{t10}

O_{t1} O_{t2} O_{t3} O_{t4} O_{t5} O_{t6} O_{t7} **Tx** O_{t8} O_{t9} O_{t10}

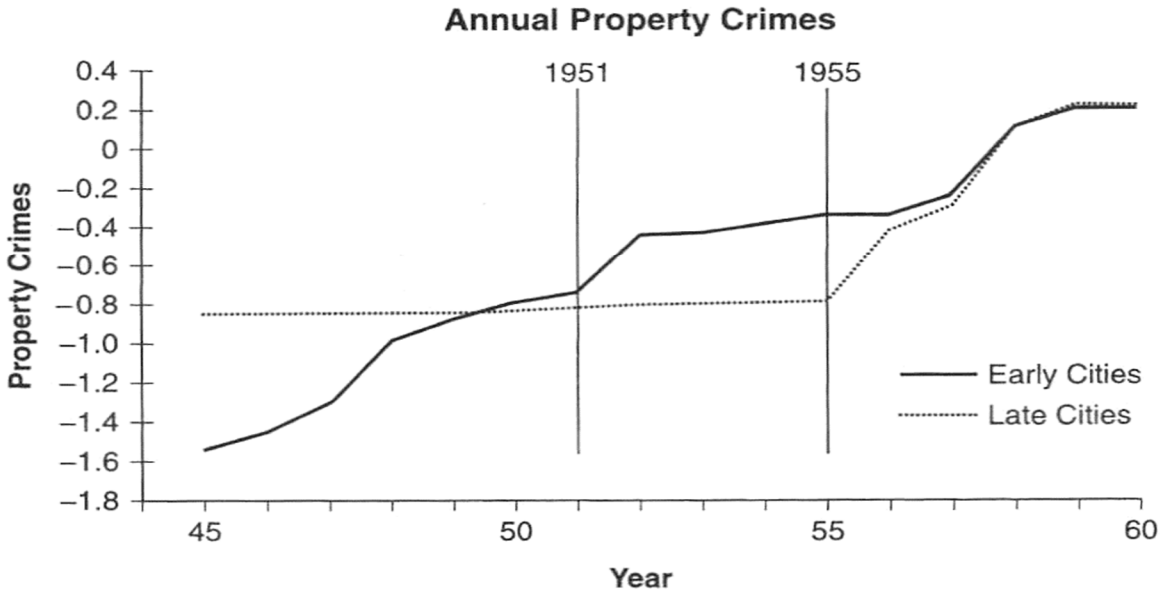


FIGURE 6.10 The effects of the introduction of television on property crime rates in cities in which television was introduced in 1951 versus 1955

From "The evolution of the time series experiment," by R. D. McCleary, 2000, *Research design: Donald Campbell's legacy, Vol. 2,* edited by L. Bickman, Thousand Oaks, CA: Sage. Copyright 2000 by Sage Publications.

[Figure from Cook & Campbell (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*]

Annual admissions for Phenylketonuria (PKU)-caused retardation as a function of PKU screening onset

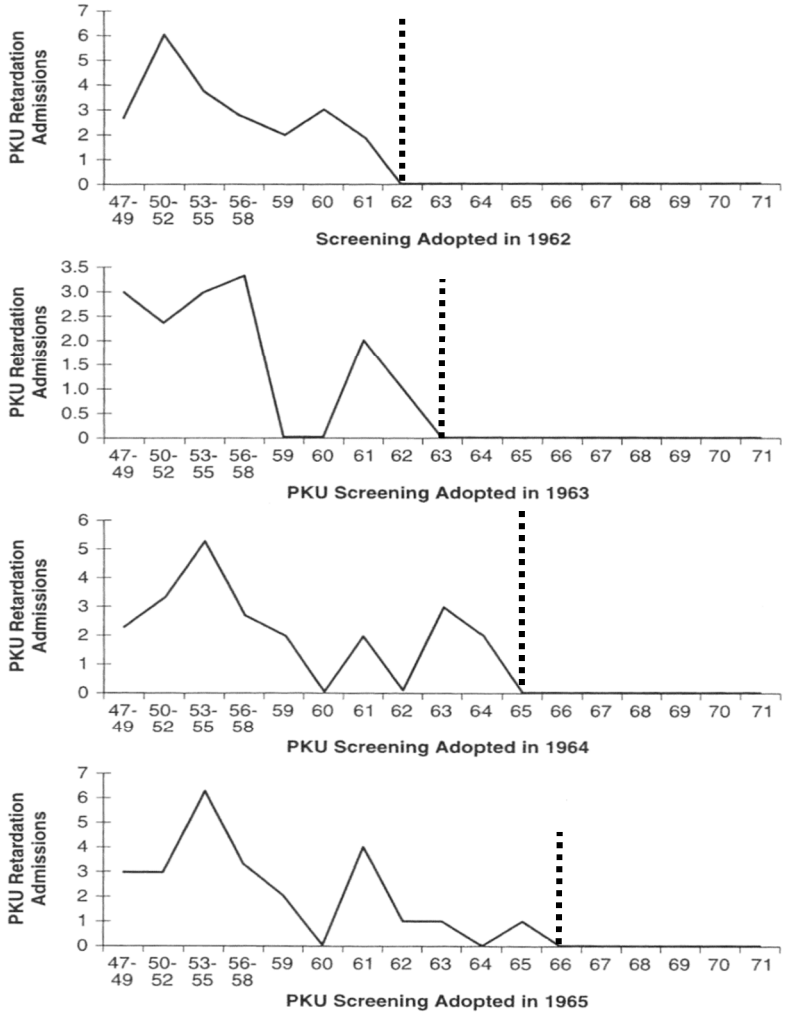


FIGURE 6.11 The effects of screening for phenylketonuria (PKU) on admissions for retardation due to PKU, with the implementation of screening staggered over 4 years in different locales

[Figure from Cook & Campbell (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*]

Analysis of data from ITS designs

Originally, time-series analysis, a modeling framework from econometrics, was used almost exclusively

Alternative: Segmented linear regression, a form of spline-regression. Plus explore and model correlation among residuals (easy)

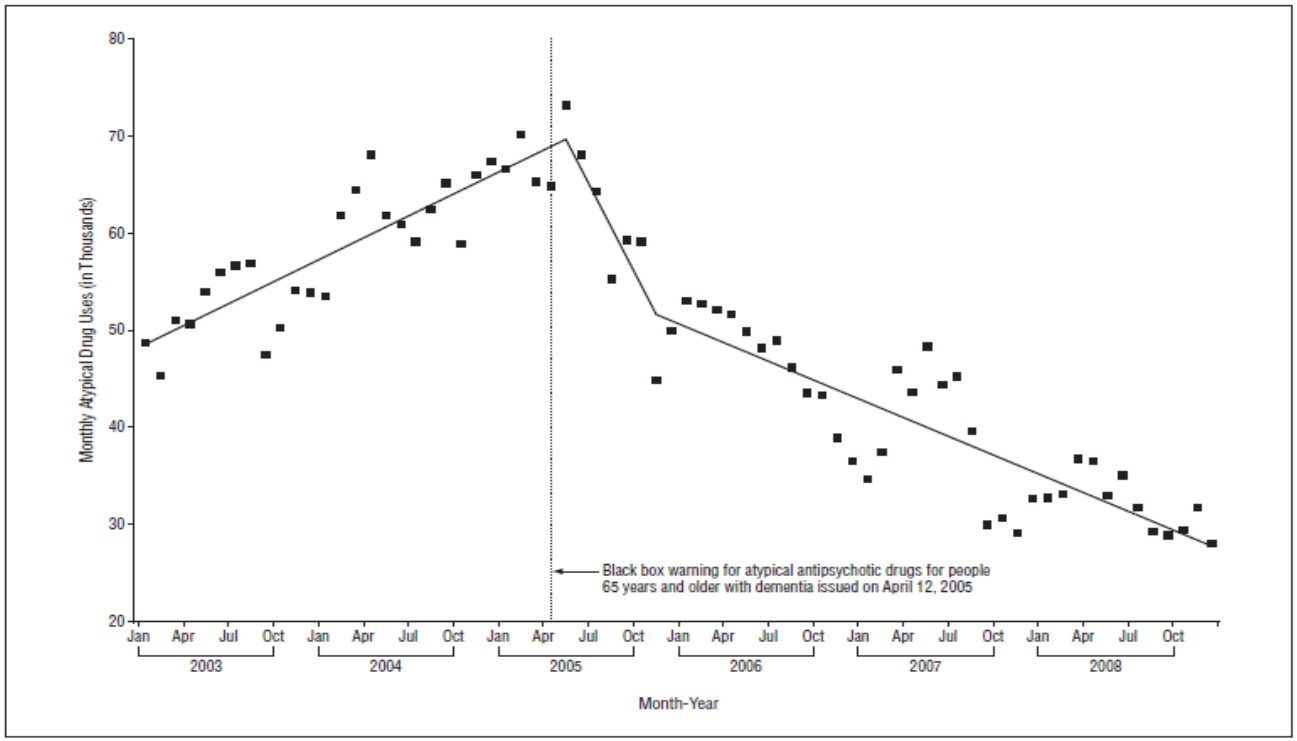


Figure 2. Joinpoint Regression Program analysis for atypical antipsychotics use among elderly patients with dementia. The data points represent patients 65 years and older with dementia (smoothed 6-month averages); the solid line, fitted joinpoint time series.

Summary

- ITS vs. other QED wrt threats to internal validity
  - . ITS far superior to pre-/post-test type designs with no control group
  - . ITS better than pre-post designs with an unmatched, non-randomized control group
  - . ITS can be better than pre-post designs with a matched non-randomized control group sample

- A suggested 'minimal' ITS design
  - . intervention onset at a single point in time
  - . intervention delivered to one population
  - . add a non-equivalent control group
  - . add non-equivalent outcomes

- Often attainable advanced design element: Switching replications
  - . A natural addition when working in multiple practices within a system, multiple hospital systems, etc.

Summary

- Units of analysis
  Outcomes often aggregated monthly, quarterly, or annual summaries
    e.g., annual incidence of a specific condition,
      total quarterly costs, average (or median) monthly LOS

- Trade-off between length of observation, level of aggregation, noise, and statistical power

- APeX data
  Opportunities to evaluate clinical policy changes, either
    . Truly retrospectively
    . Semi-prospectively, with the aid of retrospective pre- data

- Equally suitable for QI or research,
  but admin data are often more suitable for QI/policy evaluation work

END