Causal Inference in Observational Settings

7th Wellington Colloquium
Statistics NZ
30 August 2013

Professor Peter Davis
University of Auckland, New Zealand
and COMPASS Research Centre
www.compass.auckland.ac.nz
Outline

- Rationale, motivation

- Two background papers

- Handbook Outline
  - Volume I – Background
  - Volume II – Analytical techniques
  - Volume III – Temporal relations
  - Volume IV – Experimental analogues

- Two exemplar papers

- Implications for theory and practice
What’s at Issue

• **Fundamental issue of the “policy sciences”**
  – are Randomised Controlled Trials (RCTs) the only path to credible causal inference (see UK Cabinet paper)?
  – If not, how can we draw “credible” (causal?) inferences from observational data, particularly for policy?
Data Inference in Observational Settings

Four-Volume Set
SAGE Benchmarks in Social Research Methods

Edited by Peter Davis University of Auckland

Most social research is carried out in observational settings; that is, most social researchers collect information in the "real world" trying to do as little possible to alter the circumstances of study. However, there is a fundamental problem with this kind of research, in that it is very hard to draw "causal" conclusions, because of the complexity and obduracy of social reality. This is not just a problem for social scientists interested in policy or social action. It applies across the board, more generally, because it becomes difficult to know, without the conditions for credible inference, what conclusions can be drawn from any piece of empirical research that aspires to be anything more than descriptive of social phenomena.

Drawing from a variety of sources - from logicians and philosophers, to applied statisticians, computer scientists, econometricians, epidemiologists and social researchers - this collection provides an invaluable resource for scholars in the field.

Volume One: Background
Volume Two: Analytical Techniques
Volume Three: Temporal Relations
Volume Four: Experimental Analogues

December 2013 • 1628 pages
Cloth (978-1-4462-6650-2) Price £600.00

Special Introductory Offer • £550.00
(on print orders received before the end of month of publication)

---

Test, Learn, Adapt:
Developing Public Policy with Randomised Controlled Trials

Laura Haynes
Owain Service
Ben Goldacre
David Torgerson

Cabinet Office
Behavioural Insights Team
What’s at Issue

• Fundamental issue of the “policy sciences”
  – are Randomised Controlled Trials (RCTs) the only path to credible causal inference (see UK Cabinet paper)?
  – If not, how can we draw “credible” (causal?) inferences from observational data, particularly for policy?

• Causal identification via data analysis is problematic
  – often a form of speculative post-mortem
  – But, see my two health services research papers
Do Hospital Bed Reduction and Multiple System Reform Affect Patient Mortality? 
A Trend and Multilevel Analysis in New Zealand Over the Period 1988–2001

Peter Davis, PhD,* Ray Lay-Yee, MA,* Alastair Scott, PhD,† and Robin Gaud, PhD

Background: The impact of hospital and system restructuring on the quality and pattern of care is an important area of public policy concern.

Objectives: To assess the effect on patterns of care and patient outcomes of a substantial reduction in hospital bed availability and multiple reorganisations in New Zealand through the 1990s. Research Design: Trend analysis using both tabular and multilevel techniques.

Importance: Although the number of hospital beds in New Zealand declined by over 20% during the period and the national population grew by nearly 40%, discharge volumes increased significantly and rates of inpatient admission were maintained, as were access levels for vulnerable groups. These changes were accompanied by workload adjustments (a halving in length of stay and an increase in the number of non-elective cases). Yet perceived postdischarge patient mortality decreased by a quarter over the period of study, a rate of decline largely not expected by the major workload adjustments but not by reform phase.

Conclusion: Other things being equal, a substantial reduction in hospital bed availability can be effective in national public hospital systems, while largely maintaining access and quality of care. However, the workload adjustments that are required may slow improvements in patient outcomes.

Key Words: health system reform, patient outcomes, multilevel analysis

From the Departments of Medicine and Public Health, University of Auckland, New Zealand, and the Health Research Council of New Zealand, Auckland, New Zealand. Supported by the Health Research Council of New Zealand.

1186

Copyright 2007, Lippincott Williams & Wilkins. Unauthorized reproduction of this article is prohibited.
Davis et al (Lancet article)
What’s at Issue

• **Fundamental issue of the “policy sciences”**
  – are Randomised Controlled Trials (RCTs) the only path to credible causal inference (see UK Cabinet paper)?
  – If not, how can we draw “credible” (causal?) inferences from observational data, particularly for policy?

• **Causal identification via data analysis is problematic**
  – often a form of speculative post-mortem
  – But, see my two health services research papers

• **Basic conundrum of causal reasoning**
  – impossible to observe unit response under alternative
  – So, how do we know what “works”, what is evidence-based?
What Works: evidence centres for social policy

March 2013
Rationale of Handbook

1. **Traditional statistical theory**
   mainly about representation not causation (i.e. sampling)

2. **Statistical inference**\(\rightarrow\)** causal inference
   random assignment and manipulation of treatment conditions

3. **Counterfactual/potential outcomes**
   conceptually bridges experimental/observational settings

4. **Forward causation only**
   cause-to-effect (e.g. impact of policy intervention)

5. **Econometrics**
   a parallel community of policy practice (e.g. to public health)
Outline

Rationale, motivation

- Two background papers

- Handbook Outline
  - Volume I – Background
  - Volume II – Analytical techniques
  - Volume III – Temporal relations
  - Volume IV – Experimental analogues

- Two exemplar papers

- Implications for theory and practice
Two Background Papers

1. **Counterfactual thinking**

2. **Statistical reasoning**
   – Fisher never related his work on likelihoods and models to his work on experimental design

3. **Causal diagrams**
   – Using diagrams to clarify causal relationships

4. **The econometric paradigm**
   – Relying on research design rather than questionable statistical assumptions

5. **Within-study comparisons**
Ahern et al.

American Journal of Epidemiology
Volume 193, Number 8, April 15, 2006

Practice of Epidemiology


Jennifer Ahern, Alan Hubbard, and Sandra Galea

Initially submitted March 30, 2006; accepted for publication January 15, 2007

Causal inference methods allow estimation of the effects of potential public health interventions on the population burden of disease. Motivated by data for epidemiologic research to be presented in ways that are more informative for intervention, the authors present a practical discussion of the steps required to estimate the population effect of a potential intervention using an intervention-based causal inference method and discuss the assumptions of and limitations to its use. An analysis of neighborhood smoking norms and individual smoking behavior is used as an illustration. The implementation steps include the following: 1) modeling the adjusted exposure and outcome association, 2) imputing the outcome probability for individuals while maintaining the exposure by “tracing” it to different values, 3) averaging those probabilities across the population, and 4) bootstrapping confidence intervals. Imputed probabilities represent counterfactual estimates of the population smoking prevalence. Neighborhood smoking norms might be manipulated through intervention. The degree to which temporal ordering, randomization, stability, and experimental treatment assignment assumptions are met in the illustrative example is discussed, along with ways that future studies could be designed to better meet the assumptions. With this approach, the potential effects of an intervention targeting neighborhoods, individuals, or other units can be estimated.

causality; intervention studies; methods; population; residence characteristics; smoking; social environment

Abbreviations: GEE, generalized estimating equation; CRI, crude risk

Most analyses of epidemiologic data apply a regression model such as linear or logistic regression. These models have in common that they estimate differences (relative or absolute) between outcomes (in terms of rates, risks, odds, or prevalence) associated with variations in exposure, while holding constant a set of covariates (1-3). These models estimate differences in outcomes that are not specific, because they are estimated within strata of the covariates specified in the model. Although such findings constitute the backbone of modern epidemiologic research (2), they represent only the first step in exploring the association between exposure and an outcome. This approach tells us little about population disease burden or about how the disease burden might change if the exposure were modified. For example, a recent study (4) suggested that reducing neighborhood smoking norms might reduce disease burden. However, the study did not quantify the potential intervention on the population level. Given that the exposure being studied might reduce disease burden across the population, more methods can estimate population parameters under hypothetical interventions. In some situations, stratification can estimate a population-level causal effect (5, 6). Certain causal inference methods generalize stratification to situations with covariates that are continuous as well as categorical, covariates that are time dependent, models that include multiplicative confounders, and nonlinear models for binary responses (7-10). Although many causal inference methods were developed to control confounding, the machinery allows the estimation of population parameters under hypothetical interventions for cross-sectional studies. Causal inference analysis of epidemiologic data with the specification of a causal effect is made. This approach provides the population average causal effect specified in the outcome (e.g., the

Correspondence to Dr. Jennifer Ahern, School of Public Health, University of California, Berkeley, (1) 10 Harvord Hall, Berkeley, CA 94720-7365 (e-mail: jahern@berkeley.edu).

1440
Am. J. Epidemiol. 2007;166:1440-1447
Counterfactual – Neighbourhood Norms

• Population average causal effect
  • difference under one intervention vs. another (or none) by estimating counterfactual exposures->outcomes

• Epidemiological association smoking/norms
  • estimate counterfactual - impute new pattern of neighbourhood smoking norms and derive smoking levels

• Prevalence estimates if norms “manipulated”
  • 17% (versus 29%) if all neighbourhoods prohibitive
Figure 1. Predicted smoking prevalence corresponding to counterfactually “set” levels of neighborhood smoking norms, New York, New York, 2005.
Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons

Abstract

This paper analyzes 12 recent within-study comparisons contrasting causal estimates from a randomized experiment with those from an observational study sharing the same treatment group. The aim is to see whether different causal estimates result when a counterfactual group is formed, either with or without random assignment, and when statistical adjustments for selection are made in the group from which random assignment is absent. We identify three studies comparing experiment and regression-discontinuity (RD) studies. They produce quite comparable causal estimates at points around the RD cutoff. We identify three other studies where the quasi-experiment involves careful intact group matching on the pretreatment. Despite the logical possibility of hidden bias in this instance, all three cases also reproduce their experimental estimates, especially if the match is geographically local. We then identify two studies where the treatment and nonidentified comparison groups manifestly differ at pretreatment but where the selection process two treatment is completed or very plausibly known. Here too, experimental results are reproduced. Two of the remaining studies result in corresponding experimental and nonexperimental results under some circumstances but not others, while two others produce different experimental and nonexperimental estimates, through in each case the observational study was poorly designed and analyzed. Such evidence is more promising than what was achieved in past within-study comparisons, most involving job training. Reasons for this difference are discussed. © 2008 by the Association for Public Policy Analysis and Management.

INTRODUCTION

Comprehensive program evaluation depends on validly determining a program's causal impacts. Debate has been vigorous about the role experiments and observational studies should play in identifying such impacts. The main reason for preferring experiments is that, when perfectly implemented, they create intervention and control groups that do not initially differ in expectation and so do not differ on any confounded or unmeasured variables. However, the regression-discontinuity design (RD) and instrumental variables (IV) also provide unbiased causal inference in theory. So additional technical justification for preferring experiments is required. It comes from experimental estimates being more precise than RD and IV estimates (Goldberger, 1972) and also from the experiment's assumptions being more transparent in research practice. IV's main assumption is that the instrument is only correlated with outcome through treatment. This assumption is well warranted when...
Within-Study Comparison of Causal Effect

• Experiment and Regression Discontinuity
  • Comparable causal estimates around RD cut-off

• Matched intact comparison groups
  • Comparable effect estimates where intact comparison groups with overlap on pre-test means and even slopes.

• Different populations, but known selection
  • Modelling of selection process can reduce bias
Two Background Papers

1. **Counterfactual thinking**

2. **Statistical reasoning**
   - Fisher never related his work on likelihoods and models to his work on experimental design

3. **Causal diagrams**
   - Using diagrams to clarify causal relationships

4. **The econometric paradigm**
   - Relying on research design rather than questionable statistical assumptions

5. **Within-study comparisons**
Outline

Rationale, motivation

Five background papers

- Handbook Outline
  - Volume I – Background
  - Volume II – Analytical techniques
  - Volume III – Temporal relations
  - Volume IV – Experimental analogues

- Two exemplar papers

- Implications for theory and practice
SPECIAL INTRODUCTORY OFFER!

Data Inference in Observational Settings

Four-Volume Set

SAGE Benchmarks in Social Research Methods

Edited by Peter Davis University of Auckland

Most social research is carried out in observational settings: that is, most social researchers collect information in the "real world" trying to do as little possible to alter the circumstances of study. However, there is a fundamental problem with this kind of research, in that it is very hard to draw "causal" conclusions, because of the complexity and obduracy of social reality. This is not just a problem for social scientists interested in policy or social action. It applies across the board, more generally, because it becomes difficult to know, without the conditions for credible inference, what conclusions can be drawn from any piece of empirical research that aspires to be anything more than descriptive of social phenomena.

Drawing from a variety of sources - from logicians and philosophers, to applied statisticians, computer scientists, econometricians, epidemiologists and social researchers - this collection provides an invaluable resource for scholars in the field.

Volume One: Background
Volume Two: Analytical Techniques
Volume Three: Temporal Relations
Volume Four: Experimental Analogues

December 2013 • 1628 pages
Cloth (978-1-4462-6650-2) Price £600.00

Special Introductory Offer • £500.00
(on print orders received before the end of month of publication)
Sage Handbook Series

• Sage Benchmarks in Social Research Methods
• Four-volume readers
• 60 “readings”
• Previous examples
  – Social Statistics
  – Causality
  – Computational Social Science
  – Cluster Analysis
• Title: “Data Inference in Observational Settings”
Structure of Book

• Volume I – Background
  • Causal inference
  • Potential outcomes
  • “Evaluation research”

• Volume II – Analytical techniques
  • Matching methods
  • Propensity scoring
  • Causal diagrams

• Volume III – Temporal relations
  • Panel studies
  • Family studies
  • Instrumental variables

• Volume IV – Experimental analogues
  • Experimental paradigm
  • Regression discontinuity
  • Quasi-experiments, natural experiments
Outline

Rationale, motivation

Five background papers

Handbook Outline

• Volume I – Background
• Volume II – Analytical techniques
• Volume III – Temporal relations
• Volume IV – Experimental analogues

Two exemplar papers

Implications for theory and practice
Two Exemplar Papers

1. Matching/Propensity scores

2. Using panel data

3. Fixed effects

4. Instrumental variables

5. A natural experiment
Correlation or causation? Income inequality and infant mortality in fixed effects models in the period 1960–2008 in 34 OECD countries

Mauricio Avendano

Abstract

Income inequality is strongly associated with birth outcomes across countries, but whether this association is causal has not been established. To help answering these questions, we used data on income inequality in 34 OECD countries over the period 1960–2008. We estimated fixed effects models of infant mortality on the changes in income inequality, controlling for a variety of country-specific factors and using growth rate changes. We find that an increase in income inequality is associated with a 1% increase in the infant mortality rate (95% CI = 0.78, 1.38). Inverse mills ratio estimates suggest that this effect is robust to the inclusion of other confounders. These results are consistent with the hypothesis that income inequality increases infant mortality.

Introduction

During the last decades, a wide array of studies has examined the relationships between income inequality and health in high-income countries. The rationale behind these studies is that income inequality, independently of individual income, is associated with population health, so that more equal societies have better health and lower mortality (Vespa, Hamilton-Chistin, and Woodward, 1997; Kaplan, Strick, Synn, Collett, and Fuller, 1998; Lynch et al., 2001; Navarro et al., 2001; Wilkinson and Pickett, 2006;). Although income inequality is indeed correlated with overall mortality across countries, whether this association is causal has been brought into question by a series of studies showing that in many instances, the association does not persist after controlling for potential confounders (Vespa & Miler, 2003). There is disagreement, however, on the right choice of confounders and methodologies. These studies often question the direction and strength of the association (Glyson et al., 2003; Kesselheim et al., 2006; Miler & Miler, 2007; Diderichsen, 2001).

A notable exception is the association between income inequality and infant mortality, as suggested by Veit-Clock et al. (2001). In this study, we use fixed effects models to analyze data on changes in income inequality and infant mortality in 34 OECD countries over the period 1960–2008. We find that an increase in income inequality is associated with a 1% increase in the infant mortality rate (95% CI = 0.78, 1.38). These results are consistent with the hypothesis that income inequality increases infant mortality.
**Fixed Effects – Inequality and Mortality**

- Income inequality related to infant mortality
  - Strong ecological association income inequality with infant mortality across countries - but is it causal?

- Fixed effects controls variation across countries
  - Approach relies on changes in inequality within countries over time – 34 OECD countries over 38 years, Gini and IMR.

- Gini changes not associated with IMR changes
  - Possible that social policies reducing IMR cluster in relatively egalitarian countries, but their effects are not via income.
Avendano

Correlation: Inequality and infant mortality, 1963-2008

Fig. 2. Pearson correlation between household income inequality (Gini) and infant mortality rate for each year in the period 1963–2008 in 34 OECD countries.
Fig. 3. Within-country year-to-year Pearson correlation between household income inequality (Gini) and infant mortality rates in 34 OECD countries for the period 1960–2008.
Effects of Prenatal Poverty on Infant Health: State Earned Income Tax Credits and Birth Weight

Kate W. Strully, David H. Rehkopf, and Ziming Xuan

Abstract
This study estimates the effects of prenatal poverty on birth weight using changes in state Earned Income Tax Credits (EITC) as a natural experiment. We seek to answer two questions about poverty and child wellbeing: First, are there associations between prenatal poverty and lower birth weights even after factoring out unmeasured potential confounders? Because birth weight predicts a range of outcomes across the life course, lower birth weights that result from poverty may have lasting consequences for children’s life chances. Second, how have recent expansions of a work-based welfare program (i.e., the EITC) affected maternal and infant health? In recent decades, U.S. poverty rates have begun increasing again due to earnings and labor markets, but the consequences for children’s wellbeing remain controversial. We find that state EITCs increase birth weights and reduce maternal smoking. However, results related to AFDC/TANF and varying EITC effects across maternal age rates cautionary messages.

Keywords
Infant health, poverty, Earned Income Tax Credit

In life course models of stratification, early-life environment is crucially important. Exposure to poverty and negative environments during critical stages of early life can negatively affect children’s future developmental trajectories (e.g., cognitive and physical development), which may have lasting negative effects on educational attainment and adult earnings (Duncan and Brooks-Gunn 1997; Waggerl et al. 2006). According to recent research, prenatal poverty and birth weight are important variables in life course processes of stratification (Conley, Strully, and Bennett 2003; Cramer 1995). As a measure of health at the start of life, birth weight is a general indicator of a baby’s in utero environment and development, and maternal poverty during the prenatal period is a robust predictor of lower birth weights (Bennett 1997). Low birth weight can in turn predict a range of negative
Natural Experiment – Welfare and Health

• Do work/income incentives affect infant health?
  • It is hypothesised that work/income schemes will raise
    incomes and employment for unmarried mothers with high
    school or less, and in turn improve infant health.

• Using a “natural experiment” design
  • Variation between US states in introduction of income/work
    incentives to estimate effects prenatal poverty/infant health.

• Labour market, incomes, birth weight, smoking
  • Schemes increased employment 19%, incomes 32%,
    increased infant birth weight, slightly reduced smoking
Model of Pathways

Figure 2. State EITCs as a Natural Experiment
Two Exemplar Papers

1. Matching/Propensity scores

2. Using panel data

3. Fixed effects

4. Instrumental variables

5. A natural experiment
Outline

- Rationale, motivation

- Two background papers

Handbook Outline

- Volume I – Background
- Volume II – Analytical techniques
- Volume III – Temporal relations
- Volume IV – Experimental analogues

- Two exemplar papers

- Implications for theory and practice
Concluding Thoughts

• Can insistence on causal purity go too far?
  – UK Cabinet Office advice on RCTs?
  – Smoking and lung cancer; climate change
  – Status of predictive and descriptive work?

• Evidence-based movement in policy
  – This requires credible, usable evidence

• Enhance role of the social sciences
  – Need conceptual and methodological credibility
  – Social sciences limited (see CoRES, MBIE, NSC)
National Science Foundation, 2012

Bringing People Into Focus
How Social, Behavioral and Economic Research Addresses National Challenges