# Causal Inference
# in Observational Settings

**COMPASS**
**RESEARCH CENTRE**

FACULTY OF ARTS
**THE UNIVERSITY OF AUCKLAND**

Whare Wānanga o Tāmaki Makaurau

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**

New Zealand

The University of Auckland

# 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

## /III 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)

**SSAGE**

# 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

# Davis et al. (Medical Care article)





ORIGINAL ARTICLE

## 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,* Roy Lay-Yee, MA,* Alastair Scott, PhD,† and Robin Gauld, PhD‡

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

**Objective:** To assess the effect on patterns of care and patient outcomes of a substantial reduction in public hospital bed availability and multiple reorganizations in New Zealand through the 1990s.

**Research Design:** Trend analysis using both tabular and multilevel techniques.

**Subjects:** Access to discharge data, amounting to 6,639,487 records, was secured for all 34 major public hospitals in New Zealand over the period 1988–2001.

**Outcome Measures:** Number of discharges, admission rate, access levels, mean length of stay, unplanned readmission rate, and 60-day postadmission mortality rate.

**Results:** Although the number of inpatient beds in use declined by one-third over the period and the national population grew by nearly one-fifth, 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 by a quarter in readmission rates). Yet age-adjusted postadmission patient mortality decreased by a quarter over the period of study, a rate of decline that was slowed by the major workload adjustments but not by reform phase.

**Conclusions:** Other things being equal, a substantial reduction in inpatient bed availability can be effected 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

(Med Care 2007;45: 1186–1194)

Internationally there has been a considerable change in the role of the hospital through the 1990s, with higher rates of admission, shorter periods of stay, and growing rates of outpatient and day care.[1] An important strand in this change in role was a conscious restructuring of hospital workforce and redesign of work in inpatient settings across the developed world.[2] Over this same period, many of these countries also underwent bouts of broader health reform.[3] New Zealand, where the government pays for 80% of health care and public institutions dominate the health system, was no exception. The country undertook 4 sets of changes to the publicly-funded health system up to 2001 (see Fig. 1), including a succession of public hospital sector reorganizations.[4] At the same time, in a related trend, the sector experienced a substantial reduction in the availability of inpatient beds.[5]

The substantive interest in the New Zealand case is 4-fold. First, it was one of a group of countries with national health service-type systems that implemented a suite of market-oriented reforms from the late-1980s to the mid-1990s (the others being Italy, Spain, Sweden, and the United Kingdom).[6] These reforms were typically intended to create a "market" for publicly-funded health services by instituting competitive tendering between government-purchasing agencies and service providers vying among one another to win contracts to provide public services, and also by transforming public hospitals into public corporations expected to function like private cost-conscious businesses. These were features of the second and third reform phases in New Zealand (see Fig. 1). Second, this suite of reforms probably went further and faster in New Zealand than anywhere else and were part of a broader reform thrust in economic and social policy. They also drew widespread popular and political opposition.[6] Third, New Zealand simultaneously experienced both a substantial reduction in availability of public hospital beds and 4 separate structural reorganizations (Fig. 1).[4] Fourth, even though many of these reform experiments were short lived, internationally, as Or has noted, "the lack of proper evaluation . . . is striking,"[7] particularly with concerns about possible effects on access and quality.[6]

Given the strength and coherence of the reform program, and its powerfully managerial and efficiency objectives,[4] 3 key questions arise. First, how did the performance

1186

# 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?

HM Government

# What Works: evidence centres for social policy

**March 2013**

Alliance for Useful Evidence

# SQUARING THE CIRCLE
## EVIDENCE AT THE LOCAL LEVEL

**Derrick Johnstone**

May 2013

# Rationale of Handbook

1. **Traditional statistical theory**

   mainly about representation not causation (i.e. sampling)

2. **Statistical inference=>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
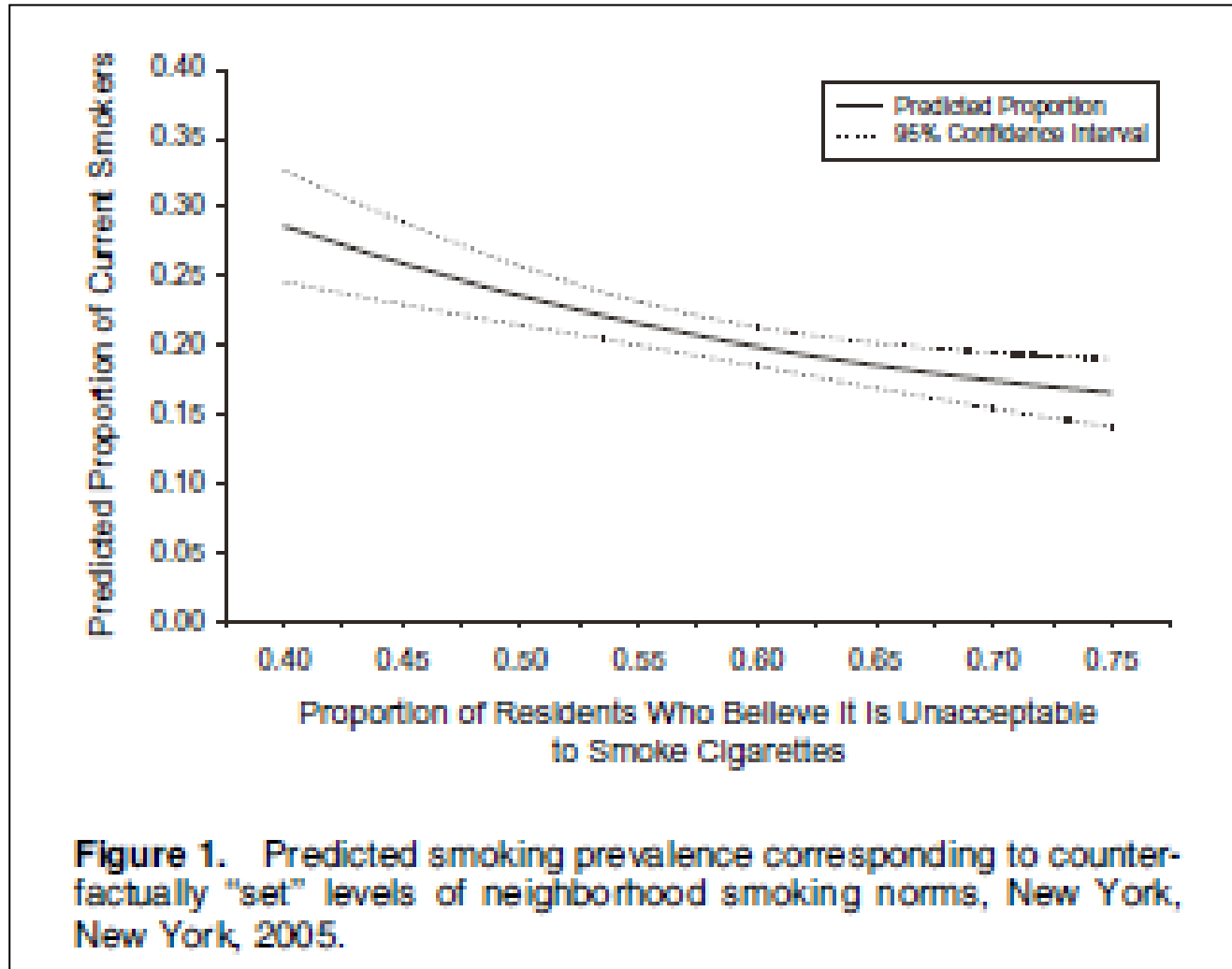
5. **Within-study comparisons**

# Ahern et al.

**Practice of Epidemiology**

**Estimating the Effects of Potential Public Health Interventions on Population Disease Burden: A Step-by-Step Illustration of Causal Inference Methods**

Jennifer Ahern, Alan Hubbard, and Sandro Galea

Causal inference methods allow estimation of the effects of potential public health interventions on the population burden of disease. Motivated by calls for epidemiologic research to be presented in ways that are more informative for intervention, the authors present a didactic discussion of the steps required to estimate the population effect of a potential intervention using an imputation-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 each individual while manipulating the exposure by "setting" it to different values, 3) averaging these probabilities across the population, and 4) bootstrapping confidence intervals. Imputed probabilities represent counterfactual estimates of the population smoking prevalence if neighborhood smoking norms could 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; OR, odds ratio.

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 prevalences) associated with variations in exposure, while holding constant a set of covariates (1–3). These models estimate differences in outcomes that are stratum specific, because they are estimated within strata of the covariates specified in the model. Although such findings constitute the backbone of modern epidemiologic research (3), they represent only 1 approach to capturing the association between an 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.

One alternate approach, which could be more informative, would assess how a particular potential intervention on the exposure being studied might reduce disease burden across the population (2, 4). Several methods can estimate population parameters under hypothetical interventions. In simple situations, standardization can estimate a population-level causal effect (5, 6). Certain causal inference methods generalize standardization to situations with covariates that are continuous as well as categorical, covariates that are time dependent, models that include multiplicative interactions, and nonlinear model forms (5–10). Although many causal inference methods were developed to control time-dependent confounding, the machinery allows the estimation of population parameters under hypothetical interventions for cross-sectional studies. Causal inference analyses of epidemiologic data start with the specification of a causal effect that is of interest. The population average causal effect is specified as the difference in the outcome (e.g., the

Correspondence to Dr. Jennifer Ahern, School of Public Health, University of California, Berkeley, 101 Haviland Hall, Berkeley, CA 94720-7358 (e-mail: jahern@berkeley.edu).

1140      Am J Epidemiol 2009;169:1140–1147

# 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

# Ahern et al.



**Figure 1.** Predicted smoking prevalence corresponding to counterfactually "set" levels of neighborhood smoking norms, New York, New York, 2005.

# Cook et al.

Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons

**Thomas D. Cook**
**William R. Shadish**
**Vivian C. Wong**

**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 test 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 experiments 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 pretest. 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 nonrandomized comparison groups manifestly differ at pretest but where the selection process into treatment is completely or very plausibly known. Here too, experimental results are recreated. Two of the remaining studies result in correspondent experimental and nonexperimental results under some circumstances but not others, while two others produce different experimental and nonexperimental estimates, though 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 measured *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

JOURNAL OF POLICY ANALYSIS AND MANAGEMENT

# 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.** **<u>Counterfactual thinking</u>**

2. <u>Statistical reasoning</u>
   – Fisher never related his work on likelihoods and models to his work on experimental design

3. <u>Causal diagrams</u>
   – Using diagrams to clarify causal relationships

4. <u>The econometric paradigm</u>
   – Relying on research design rather than questionable statistical assumptions

**5.** **<u>Within-study comparisons</u>**

# 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**

# Data Inference in Observational Settings

## /III 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

**⑤SAGE**

# 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**

# Avendano

# 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

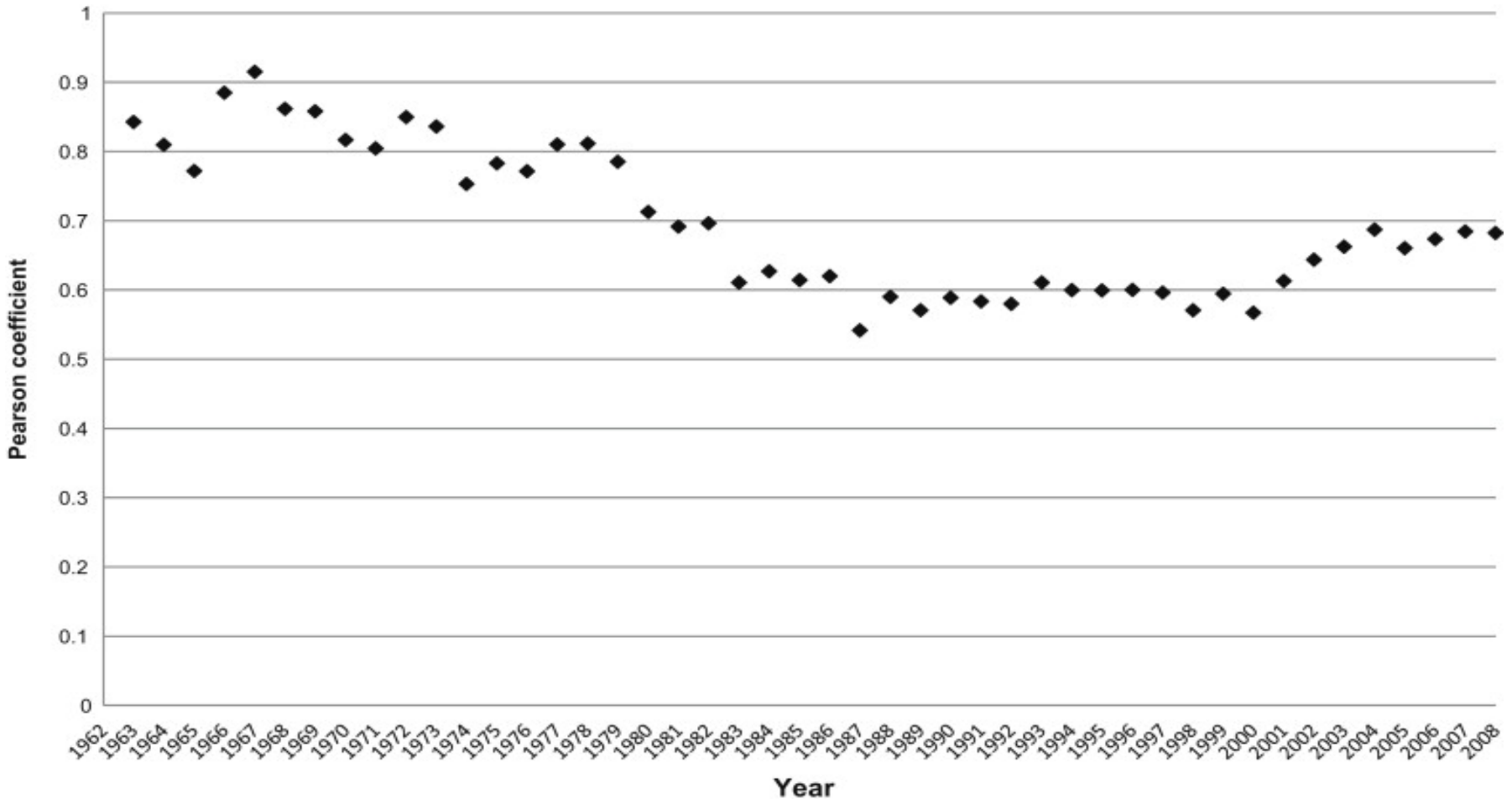Fig. 2. Pearson correlation between household income inequality (Gini) and infant mortality rate for each year in the period 1963–2008 in 34 OECD countries.

# Avendano

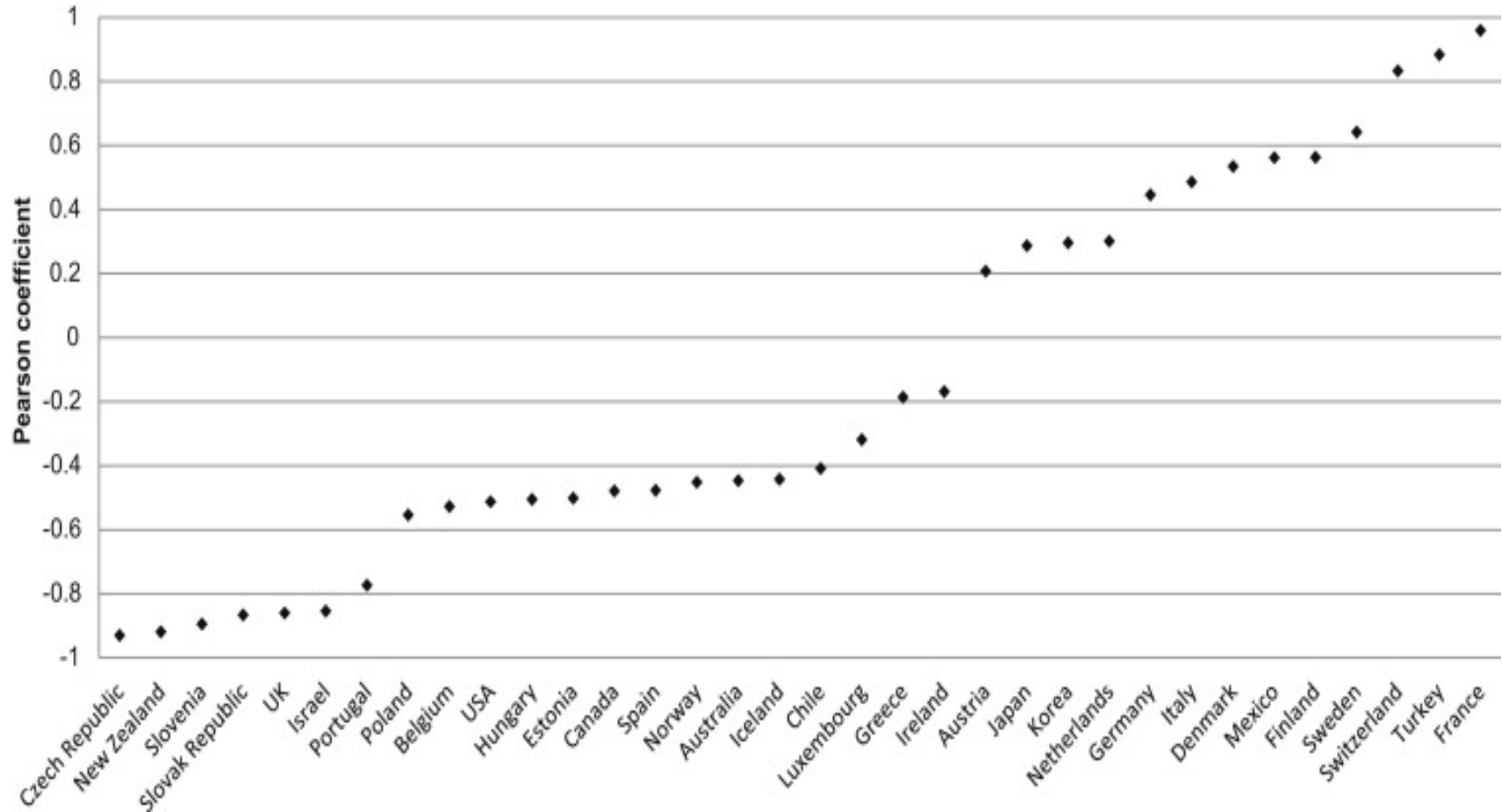**Year-to-year correlation: Income inequality and infant mortality, 1960-2009**



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.

# Strully et al.

## Effects of Prenatal Poverty on Infant Health: State Earned Income Tax Credits and Birth Weight

Kate W. Strully,[a] David H. Rehkopf,[b] and Ziming Xuan[c]

### 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 relief has become increasingly tied 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 ages raise 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; Wagmiller 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

[a]University at Albany, SUNY
[b]University of California, San Francisco
[c]Harvard School of Public Health

**Corresponding Author:**
Kate W. Strully, University at Albany, SUNY, 1400 Washington Avenue, AS-308, Albany, NY 12222
E-mail: kstrully@albany.edu

# 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
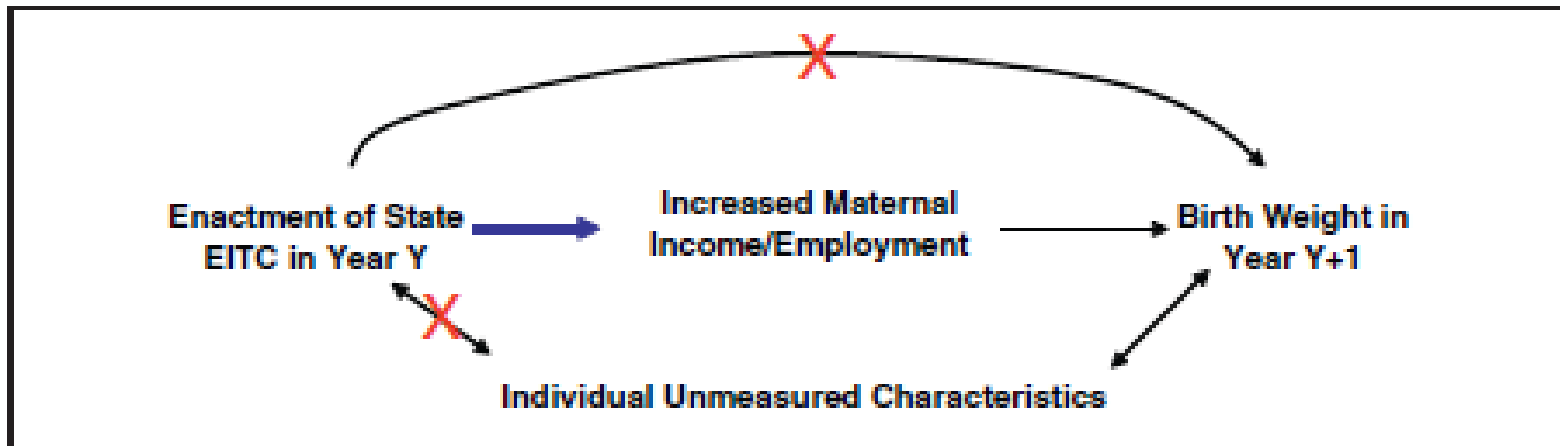
# Model of Pathways



**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

- <u>Can insistence on causal purity go too far?</u>
  - UK Cabinet Office advice on RCTs?
  - Smoking and lung cancer; climate change
  - Status of predictive and descriptive work?
- <u>Evidence-based movement in policy</u>
  - This requires credible, usable evidence
- <u>Enhance role of the social sciences</u>
  - Need conceptual and methodological credibility
  - Social sciences limited (see CoRES, MBIE, NSC)

# National Science Foundation, 2012