# Statistical Models and Causal Inference

## A Dialogue with the Social Sciences

David A. Freedman

David Collier, Jasjeet Sekhon, and Philip B. Stark, eds.

Cambridge University Press, Forthcoming 2009

**Table of Contents**

## PART II
## Studies in Political Science, Public Policy, and Epidemiology

***4 Methods for Census 2000 and Statistical Adjustments***

The U.S. Census is a sophisticated, complex undertaking, carried out on a vast scale. It is remarkably accurate. Statistical adjustments are likely to introduce more error than they remove. This issue was litigated all the way to the Supreme Court, which unanimously supported the decision by the Secretary of Commerce not to adjust.

***5 On "Solutions" to the Ecological Inference Problem***

Gary King's book, *A Solution to the Ecological Inference Problem*, claims to offer "realistic estimates of the uncertainty of ecological estimates." Applying King's method and three of his main diagnostics to data sets where the truth is known shows that his diagnostics cannot distinguish between cases where estimates are accurate and those where estimates are far off the mark. King's claim to have arrived at a solution to this problem is premature.

***6 Rejoinder to King***

King's method works with some data sets but not others. As a theoretical matter, inferring the behavior of subgroups from aggregate data is generally impossible: the relevant parameters are not identifiable. King's diagnostics do not discriminate between probable successes and probable failures.

***7 Black Ravens, White Shoes, and Case Selection: Inference with Categorical Variables***

Statistical ideas can clarify issues in qualitative analysis such as case selection. In political science, an important argument about case selection evokes Hempel's Paradox of the Ravens. This paradox can be resolved by distinguishing between population and sample inferences.

***8 What is the Chance of an Earthquake?***

Making sense of earthquake forecasts is surprisingly difficult. In part, this is because the forecasts are based on a complicated mixture of geological maps, rules of thumb, expert opinion, physical models, stochastic models, and numerical simulations, as well as geodetic, seismic, and paleoseismic data. Even the concept of probability is hard to define in this context. Other models of risk for emergency preparedness, as well as models of economic risk, face similar difficulties.

***9 Salt and Blood Pressure: Conventional Wisdom Reconsidered***

Experimental evidence suggests that the effect of a large reduction in salt intake on blood pressure is modest, and health consequences remain to be determined. Funding agencies and medical journals have taken a stronger position favoring the salt hypothesis than is warranted, demonstrating how misleading scientific findings can influence public policy.

***10 The Swine Flu Vaccine and Guillain-Barré Syndrome: A Case Study in Relative Risk and Specific Causation***

Epidemiologic methods were developed to prove general causation: identifying exposures that increase the risk of particular diseases. Courts of law often are more interested in specific causation: on balance of probabilities, was the plaintiff's disease caused by exposure to the agent in question? There is a considerable gap between relative risks and proof of specific causation because individual differences affect the interpretation of relative risk for a given person. This makes specific causation especially hard to establish.

***11 Survival Analysis: An Epidemiological Hazard?***

Proportional-hazards models are frequently used to analyze data from randomized controlled trials. This is a mistake. Randomization does not justify the models, which are rarely informative. Simpler methods work better. This discussion matters because survival

analysis has introduced a new hazard: it can lead to serious mistakes in medical treatment. Survival analysis is, unfortunately, thriving in other disciplines as well.

## PART III
## New Developments: Progress or Regress?

### *12 On Regression Adjustments in Experiments with Several Treatments*

Regression adjustments are often made to experimental data to address confounders that may not be balanced by randomization. Since randomization does not justify the models, bias is likely. Neither are the usual variance calculations to be trusted. Neyman's non-parametric model serves to evaluate regression adjustments. A bias term is isolated, and conditions are given for unbiased estimation in finite samples.

### *13 Randomization Does Not Justify Logistic Regression*

The logit model is often used to analyze experimental data. Theory and simulation show that randomization does not justify the model, so the usual estimators can be inconsistent. Neyman's non-parametric setup is used as a benchmark: each subject has two potential responses, one if treated and the other if untreated; only one of the two responses can be observed. A consistent estimator is proposed.

### *14 The Grand Leap*

A number of algorithms purport to discover causal structure from empirical data with no need for specific subject-matter knowledge. Advocates have no real success stories to report. These algorithms solve problems quite removed from the challenge of causal inference from imperfect data. Nor do they resolve long-standing philosophical questions about the meaning of causation.

### *15 On Specifying Graphical Models for Causation, and the Identification Problem*

Causal relationships cannot be inferred from data by fitting graphical models without prior substantive knowledge of how the data were generated. Successful applications are rare because few causal pathways can be excluded *a priori*.

### *16 Weighting Regressions by Propensity Scores*

The use of propensity scores to reduce bias in regression analysis is increasingly common in the social sciences. Yet weighting is likely to increase random error in the estimates and to bias the estimated standard errors downward, even when selection mechanisms are well understood. If investigators have a good causal model, it seems better just to fit the model without weights. If the causal model is improperly specified, weighting is unlikely to help.

### *17 On the So-Called "Huber Sandwich Estimator" and "Robust Standard Errors"*

In applications where the statistical model is nearly correct, the Huber Sandwich Estimator makes little contribution. On the other hand, if the model is seriously in error, the parameters being estimated are likely to be meaningless – except perhaps as descriptive statistics.

### *18 Endogeneity in Probit Response Models*

The usual Heckman two-step procedure should not be used for removing endogeneity bias in probit regression. From a theoretical perspective, this procedure is unsatisfactory, and likelihood methods are superior. Unfortunately, standard software packages do a poor job of maximizing the biprobit likelihood function, even if the number of covariates is small.