

Chapter 6

SIMPSON'S PARADOX, CONFOUNDING, AND COLLAPSIBILITY

*He who confronts the paradoxical
exposes himself to reality.*

Friedrick Durrenmatt (1962)

Preface

Confounding represents one of the most fundamental impediments to the elucidation of causal inferences from empirical data. As a result, the consideration of confounding underlies much of what has been written or said in areas that critically rely on causal inferences; this includes epidemiology, econometrics, biostatistics, and the social sciences. Yet, apart from the standard analysis of randomized experiments, the topic is given little or no discussion in most statistics texts. The reason for this is simple: confounding is a causal concept and hence cannot be expressed in standard statistical models. When formal statistical analysis is attempted, it often leads to confusions or complexities that make the topic extremely hard for the nonexpert to comprehend, let alone master.

One of my main objectives in writing this book is to see these confu-

sions resolved—to see problems involving the control of confounding reduced to simple mathematical routines. The mathematical techniques introduced in Chapter 3 have indeed culminated in simple graphical routines of detecting the presence of confounding and of identifying variables that need be controlled in order to obtain unconfounded effect estimates. In this chapter, we address the difficulties encountered when we attempt to define and control confounding by using statistical criteria.

We start by analyzing the interesting history of Simpson's paradox (Section 6.1) and use it as a magnifying glass to examine the difficulties that generations of statisticians have had in their attempts to capture causal concepts in the language of statistics. In Sections 6.2 and 6.3, we examine the feasibility of replacing the causal definition of confounding with statistical criteria that are based solely on frequency data and measurable statistical associations. We will show that, although such replacement is generally not feasible (Section 6.3), a certain kind of nonconfounding conditions, called *stable*, can be given statistical or semistatistical characterization (Section 6.4). This characterization leads to operational tests, similar to collapsibility tests, that can alert investigators to the existence of either instability or bias in a given effect estimate (Section 6.4.3). Finally, Section 6.5 clarifies distinctions between collapsibility and no-confounding, confounders and confounding, and between the structural and exchangeability approaches to representing problems of confounding.