

SELF-SELECTION—A MAJOR PROBLEM IN OBSERVATIONAL STUDIES

J. YERUSHALMY

CHILD HEALTH AND DEVELOPMENT STUDY
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

I am pleased to have been asked to participate in the Sixth Berkeley Symposium on Mathematical Statistics. It provides an opportunity for an exchange of information and ideas on some methodologic problems encountered in observational studies of etiologic factors in chronic diseases. These problems present different aspects than those ordinarily involved in experimental studies usually encountered in statistical investigations. In the latter, especially in those on inferences derived from comparisons of two samples, there is a tacit assumption that the samples have been equalized usually through a procedure of randomization.

In studies on human beings, experimentations are difficult, if not impossible. In attempts to determine the roles of environmental factors in development of disease, one must depend on observation of phenomena as they occur in nature. Often factors such as emotional stress, physical conditions, certain components in the diet, cigarette smoking, sedentary occupations, and other characteristics require prolonged periods of observation between the exposure to the environmental factor and the development of disease. In many cases the period is measured in years and sometimes decades. Consequently, a manipulative study would involve major changes in modes of life of large groups of people over prolonged periods of time. Such experimentations are impractical, if not impossible. Of necessity, therefore, the main methods are those of observations of associations between prevalence of a given disease or condition and environmental factors as they occur in nature without interference on the part of the investigator.

The most elementary and least desirable of these observational studies are the so-called "indirect studies." In these, the unit of investigation is the group rather than the individual. Often mortality or morbidity data derived from vital statistics are compared with indices of suspected environmental characteristics. Such studies are useful to provide leads for further investigations. They have,

Supported by Grant No. HD 7256 of the National Institutes of Health.