EPIDEMIOLOGY METHODS

T. E. Raghunathan
Department of Biostatistics and Research Professor, Institute for Social Research, University of Michigan, Ann Arbor, USA

Keywords: Attributable risk, Case-control design, Cohort design, Disease Development, Measurement error, Missing data, Observational Studies, Odds ratio, Randomization, Relative Risk

Contents

Summary

Epidemiology aims to study patterns of disease and of the factors affecting these patterns within a well-defined conceptual or philosophical framework. Epidemiological investigations can be grouped into four broad categories: Observational epidemiology, experimental epidemiology, natural experiments, and Theoretical epidemiology. Several types study designs and measures of relationship are used in these investigations. An overview of different types of epidemiological designs is provided in Section 2. Relative Risk, Odds Ratio and Attributable Risk are the three most common measures of association often used in these investigations. These measures are described in Section 3. The emphasis is on observational studies, which forms a bulk of epidemiological investigations. The design and analysis issues concerning the prospective and retrospective studies, the two most prominent types of observational studies are discussed. Finally, the chapter concludes with a discussion of areas needing further research.

1. Introduction

Epidemiology is concerned with investigations of patterns of disease distribution and of the factors influencing these patterns. These investigations usually are conducted within a conceptual model or philosophical framework. Thus, epidemiology can be viewed as an integrative eclectic discipline deriving strength from diverse disciplines such as biology, biostatistics, economics, genetics, medicine, psychology and sociology. The field of social epidemiology, for example, is an inductive science concerned with the influence of social, economic and psychosocial factors on disease patterns. On the other

UNESCO – EOLSS SAMPLE CHAPTERS

hand, pharmaco-epidemiology is concerned with the influence of a particular set of drugs or formulations on disease occurrence.

Epidemiology is also essentially a “comparative” discipline. That is, most epidemiological investigations study diseases and the potential factors related to these diseases among different groups, different periods and different places and then make comparisons. The first step in this process is to determine the strength of statistical association between a factor and disease and then derive biological inferences from the statistical associations.

The information obtained from these investigations is used in several ways:

1. To fully understand the etiology of a specific disease or a group of diseases. For example, a study goal may be to fully understand the roles of genetics and social environment on a set of established cardiovascular disease risk factors such diabetes, hypertension, obesity and hypercholestremia.

2. To evaluate the consistency of epidemiologic data with etiological hypotheses developed using clinical or laboratory observations. For example, several clinical and laboratory observations on the relationship between oral contraceptive use and cardiovascular disease lead to several large epidemiological investigations.

3. To provide a basis for developing consistent preventive procedures, public health policies and public health practices. For example, the epidemiological investigations described in (2) led to lowering of estrogen dose in the oral contraceptives, which have demonstrably lowered the risk of cardiovascular diseases.

Given such a wide scope and uses of epidemiology, several types of epidemiological studies have been proposed. The bulk of the epidemiology is concerned with observational studies. This chapter provides an overview of the design and analytical issues in such studies. The rest of the chapter is organized into 3 sections. Section 2 briefly describes different kinds of epidemiological investigations; Section 3 describes statistical quantities used to measure the strength of association between disease and exposure (characteristic of interest). Section 4 addresses certain design and analytical issues in retrospective and prospective studies, two prominent designs used in epidemiology. Finally, Section 5 concludes the paper with a discussion of areas needing further research.

UNESCO – EOLSS SAMPLE CHAPTERS

2. Types of Investigations

Epidemiological investigations can be broadly grouped as follows:

1. **Observational Epidemiology** refers to the inference about the etiological factors that influence the disease occurrence based on the collection and analysis of data from human population groups. Much of epidemiology falls into this category. The designs used in these investigations can be classified into two categories, retrospective (cross-sectional or case-control) and prospective (longitudinal or cohort) designs. These designs are described later. There are some hybrid

versions these designs, which are also briefly described later. A bulk of this paper is confined to observational studies.

2. **Experimental Epidemiology** broadly refers to a planned experiment where the investigator has control over the population groups by deciding which groups are exposed to a factor under scrutiny. For example, much of the acceptable evaluation studies of preventive measures fall into this category. The design and analysis methods employed in these investigations are similar to those used in randomized clinical trials. A reader interested in finding more details about the design and analytical issues can find them in numerous books and a popular one is listed in the references. There are some caveats that are unique to experimental epidemiology. For instance, an evaluation of preventive measure may involve randomization of groups of people, rather than individuals.

3. **Natural Experiments** refers to a fortuitous situation where a natural course of public events closely approximate a planned control experiment. However, these are basically observational studies. For instance, the welfare reform in the United States has induced important differences in the social conditions. Thus, the effect of changes in the social conditions on the disease occurrence can studied through the patterns of disease occurrence in the two periods, before and after the welfare reforms were implemented. Such natural experiments occur through important public policy changes at both the state and national levels. A natural experiment may also occur in a setting of a cohort study. For example, the Monitoring the Future study is an on-going large school based longitudinal survey to elicit information about the drug and alcohol use and to study its impact on the well-being. During the course of this study, many new laws governing drug use were enacted, thus providing a natural experiment setting to perform “before-after” comparisons. The common methods discussed in Section 3 can be used to analyze the data from such experiments.

4. **Theoretical Epidemiology** refers to development mathematical and statistical models to explain the patterns of occurrence of diseases. Several models to explain the outbreak and spread of infectious diseases have been developed. Modern computing power enables numerical simulation to check and refine these models. It is also possible to use the results from the observational and planned or natural experiments to develop mathematical models. These mathematical models can then used to simulate “what-if" scenarios of disease patterns when the associated factors are altered in some meaningful ways.

-
-
-

UNESCO – EOLSS SAMPLE CHAPTERS

TO ACCESS ALL THE **13 PAGES** OF THIS CHAPTER,
Visit: http://www.eolss.net/Eolss-sampleAllChapter.aspx

Bibliography

Breslow, N. E., and Day, N. E. (1988). *Statistical Methods in Cancer Research, Volume 1, The Design and Analysis of Case-control Studies*, Oxford University Press [This is an excellent book describing design and analysis of observational studies. This volume is concerned with retrospective or case-control designs].

Breslow, N. E., and Day, N. E. (1988). *Statistical Methods in Cancer Research, Volume 2, The Design and Analysis of Cohort Studies*, Oxford University Press [This is a companion volume to the previous reference and discusses design and analysis issues in cohort studies].

Cochran, W. G. (1977). *Sampling Techniques*. 3rd Ed. Wiley: New York [This is a basic text book describing probability sampling techniques].

Cornfield, J. (1956). A statistical problem arising from retrospective studies. *Proc. Third Berkeley Sympo. Math. Stat. Prob.*,4, 135-148.[This is one of the fundamental papers demonstrating that relative risk can be approximately estimated from retrospective studies. A classic paper which must be read by any student of epidemiology]

Hosmer, D. W., Jr and Lemeshow, S. (1989).*Applied Logistic Regression*. Wiley: New York.[One of the most readable books for applied researchers describing common statistical techniques used in epidemiology]

Kish, L. (1995), *Survey Sampling*, Wiley: New York.[Another classic text book on probability sampling techniques with a emphasis on practical aspects]

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Incomplete Data*, Wiley: New York. [This is a basic textbook describing the impact of missing data and the methods for analyzing incomplete data using weighting, imputation or maximum likelihood methods].

Pocock, S. J. (1984). *Clinical Trials: A Practical Approach*. Wiley: New York. [This text describes the issues in randomized trials often used in experimental epidemiology].

Prentice, R. L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies, *Biometrika*, 66, 403-412.[This is a companion to the paper by Cornfield. This generalizes Cornfield's results when one needs to adjust for several covariates while assessing the disease-exposure relationship.]

Rosenbaum, P. R. (1995). *Observational Studies*, Springer-Verlag.[This book discusses issues concerning making causal inferences from observational studies.]

Biographical Sketch

Trivellore Eachambadi Raghunathan was born in Nagpur, India. He obtained his Bachelor Science with Physics, Mathematics and Statistics as concentrations (1977) and Master of Science in Statistics in 1979, at the Institute of Science, Nagpur University. He received several gold medals and cash prizes for being ranked among the top 10 of all undergraduates and as the top student among all post graduate students. He taught at the Institute of Science for 3 years. He received another Masters degree in Statistics from Miami University, Oxford, Ohio in 1983 and a Ph.D. degree in Statistics in 1987 from Harvard University. In 1987, he joined the Department of Biostatistics, University of Washington, Seattle, Washington as an Assistant Professor. He moved to the University of Michigan in 1994 with a joint appointment at the Institute for Social Research and the Department of Biostatistics in the School of Public Health. . He was the director of the Survey Methodology Program at the Institute for Social Research from 1997-2001. He became a Professor of Biostatistics and a Research Professor at the Institute for Social Research in 2002. He is an Associate Director of the Center for Research on Ethnicity, Culture and Health (CRECH). He is a faculty member at the Center of Social Epidemiology and Population Health (CSEPH). He is also affiliated with the University of Michigan Transportation Research Institute (UMTRI). He also teaches in the Joint Program in Survey Methodology at the University of Maryland, a consortium of University of Michigan, University of Maryland and Westat. He continues to be involved in several projects at the Cardiovascular Health Research Unit (CHRU) at the

UNESCO – EOLSS SAMPLE CHAPTERS

University of Washington. His research interests are in the analysis of incomplete data, multiple imputation, Bayesian methods, design and analysis of sample surveys, small area estimation, confidentiality and disclosure limitation, longitudinal data analysis and statistical methods for epidemiology. He has developed a SAS based software for imputing the missing values for a complex data set and can be downloaded from www.isr.umich.edu/src/smp/ive. He is author on more than 80 refereed publications and more than 25 non-refereed publications. He regularly participates on several national committees, referees articles and grants and provides other editorial services. He was elected a fellow of American Statistical Association in 2003.

UNESCO – EOLSS
SAMPLE CHAPTERS