Perspective: Cohort Studies

Jonathan M. Samet and Alvaro Muñoz

This issue of *Epidemiologic Reviews* on cohort studies is published five decades after the initiation of the Framingham study, one of the landmark investigations of epidemiologic research and, more generally, contemporary biomedical research. Within a few years other cohort studies—for example, of British physicians and of the atomic bomb survivors—were initiated. As follow-up of the participants in some of these studies has continued, the studies provide information on changing risks over time and the modification of risks by increasing age. The findings from many subsequent cohort studies have been widely applied as bases for developing public health policy. Evidence from cohort studies has tended to have greater credibility as a basis for decision-making than that from other epidemiologic designs, in part because cohort studies can directly describe the sequence from exposure to disease.

Thus, the place of cohort studies in epidemiologic research is assured, and cohort studies are certain to figure prominently in future epidemiologic and clinical research. In spite of its widespread application, there have been remarkably few texts and collections of papers on this central study design. Also lacking are contributions to the literature that link new developments in statistical methods for longitudinal data analysis with epidemiologic applications.

The contributions to this issue of *Epidemiologic Reviews* provide a broad review of epidemiologic aspects of the cohort study that should complement the literature on statistical techniques for longitudinal data. The papers address both the conduct of cohort studies and conceptual issues in the design of cohort studies and in the interpretation of their findings. Two contributions, those authored by Thomas and by Muñoz and Gange, focus on analytical issues. Considered together, the papers in this volume provide a starting point for learning about the use of the cohort design in epidemiologic research. They also point to many issues needing methodological research at both practical and conceptual levels.

The central role of the cohort design in epidemiologic research is highlighted in a number of the contributions, including the introduction by Samet and Muñoz. They provide an historical perspective, describing the evolution of the cohort design and offering vignettes of key studies. Several contributions address the methodological foundation of the cohort design. Tager considers the types of outcome measures that are evaluated in cohort studies, dividing them into “life-table-type” and “longitudinal.” Survival as an outcome (e.g., time to death) corresponds to the former category; this type of outcome has been central in many past cohort studies. New methods now make possible studies of Tager’s second category, which involves repeated measures. Muñoz and Gange address one particular area of methodological advancement in cohort studies, the use of intermediate markers of outcomes. They cover the conceptual basis for designs directed at intermediate markers, along with related analytical methods.

The conduct of cohort research is often challenging, as data are accrued over time and efforts are made to maintain and assure data quality. Exposure assessment may be particularly difficult in cohort studies. Exposures may change over time and measurement quality may drift. Two contributions provide guidance in this area. White and colleagues address conceptual issues in exposure assessment as well as design strategies. Whitney and colleagues offer guidance on the development of systematic strategies to assure data quality in cohort studies. Retaining participants in cohort studies is of evident importance. Hunt and White offer practical strategies for enhancing participation and follow-up.

Cohort studies may be conducted in participants drawn from the general population and selected to represent the population, or in participants in special groups, often defined by having some particular exposure. Szklı offers a perspective on population selection; he thoughtfully considers the external validity of cohort studies based in specific, defined populations.

Accepted for publication July 24, 1998.
From the Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, MD.
Reprint requests to Dr. Jonathan M. Samet, Professor and Chairman, Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205.
Willett and Colditz cover the design of more contemporary cohort studies, with some, like the Nurses' Health Study, involving large populations. Many of these more recent studies have been implemented to assess effects of specific exposures, such as oral contraceptives or diet, and the designs often have innovative features to enhance feasibility. As described by Howe, cohort studies can now be readily accomplished through approaches utilizing record linkage. Checkoway and Eisen review occupational cohort studies; the cohort design has proved essential in investigating the health effects of workplace exposures. Worker populations can be identified through employment records and exposures can be estimated; follow-up provides surveillance for adverse effects.

The cohort design is evolving at a rapid pace, as capacity for data collection and management advances and analytical approaches become increasingly flexible and informative for longitudinal data analysis. Thomas highlights these advances and indicates the enhancement of our capacity to utilize data from cohort studies informatively and fully. Muñoz and Gange address analytical issues related to repeated measures of biomarkers as either indicators of exposure or intermediate outcomes of disease progression.

The cohort design will have increasing application. We anticipate broadening application of nested designs directed at intermediate markers; with validated markers, these designs may provide findings more quickly and efficiently than full cohort studies of the ultimate disease outcome. The cohort design will also find application in characterizing the genetic basis of disease; stored specimens offer the possibility of evaluating multiple markers, using nested designs. Cohort studies conducted over long periods when different interventions and therapies are introduced and their uses documented will be beneficial for evaluating effectiveness at the population level. In addition, cohort studies of patient groups enrolled in health care plans will be carried out using record linkage approaches in order to evaluate effectiveness of care.

This collection of papers on cohort studies highlights our advancing understanding of the design and its application in epidemiologic research. The papers in this issue also point to many areas for methodological research: for example, time-dependent measurement error, and missing data problems in a longitudinal context. Vigorous collaboration between epidemiologists and biostatisticians will continue to bring solutions to these and other methodological challenges in cohort studies.

We wish to thank our colleagues for their thoughtful and provocative contributions and we hope that the Epidemiologic Reviews readership will find them as insightful as we have.