PREFACE

Long-term follow-up (cohort) studies of human populations, particularly of industrial workers, of patients treated with radiation and cytotoxic chemotherapy, and of victims of nuclear and other disasters, have provided the most convincing evidence of the link between exposure to specific environmental agents and cancer occurrence. Of the chemicals and industrial processes for which working groups convened by the IARC have decided that there is 'sufficient evidence' of human carcinogenicity, cohort studies provided the definitive evidence in the great majority of cases. In the studies carried out in the 1950s and 1960s, high risks were associated with specific exposures. Relatively simple statistical methods were sufficient to demonstrate the effect, and the finer quantitative features of the relationship were not emphasized. It was not uncommon for reports of occupational hazards to be based primarily on the computation of standardized death rates or mortality ratios (SMRs) for a few causes of death, with virtually no attention paid to internal comparisons among differentially exposed workers. Since then, the picture has changed. More attention is now paid to the quantification of risk and the use of more refined dose-response models. Interest has also turned to a wider range of exposures and the interplay between physiological measures of nutritional status, dietary factors and other variables of modes of life. Multivariate methods are then necessary, often making use of serial measurements on the same individuals.

Increasingly, modern concepts of statistical inference and modelling are being used to maximize the information obtainable from these major endeavours and to provide the most precise estimates possible of quantitative risk. Indeed, some cohort studies have stimulated the development of new statistical methods of particular relevance to this field.

The primary purpose of this monograph is to bring together in one place the statistical developments that have taken place during the past few years that are of relevance to the design and analysis of cohort studies, and to illustrate their application to several sets of data of importance in the field of cancer epidemiology. We hope to present these new statistical methods in such a way that epidemiologists and other research workers without extensive statistical training can appreciate the possibilities they offer and, in many cases, can apply them to their own work. In addition, by providing a thorough introduction to the design and execution of cohort studies, including a detailed description of six landmark investigations of this type, we hope to interest students of statistical science in this field so that they may turn their attention

both to the proper application of current methods and to the further development of those methods.

In the preface to the first volume in this series we stressed the essential similarity of statistical methods applicable to the case-control and cohort approaches to epidemiological research, the flexibility of new methods for handling a variety of data configurations and the wide range of problems that could be approached from a common conceptual foundation. This pursuit of unity and flexibility continues to be our goal. We show how elementary methods that have long been used for analysis of cohort data relate to explicit statistical models, and how they may be extended so as to achieve greater understanding of the collected data. The SMR, for example, has been used virtually without change for over 200 years to make age-adjusted comparisons of regional and occupational mortality. We show how this statistic may be derived as a maximum likelihood estimate in a well-defined statistical model, and how an extension of that model leads to a regression analysis of the SMR as a function of one or more risk factors. This approach shows us that the well-known 'lack of comparability' of SMRs is due to the problem of statistical confounding and may be alleviated by a proper analysis. Further extensions of the basic model permit variations in the SMR to be estimated as a nonparametric function of time for purposes of exploratory analyses of data.

Experience with the first volume taught us that one of its most important features, made possible through the generosity of our collaborators, was the provision of appendices containing several condensed, but nonetheless bona-fide, sets of data. These were used in worked examples that readers could follow to test their understanding of the material (and, occasionally, to find our mistakes). The present volume contains appendices that give grouped data from a study of respiratory cancer among smelter workers in Montana, USA, and both grouped and individual data records on 679 Welsh nickel refiners who had high rates of lung and nasal sinus cancer. Summary data from several other studies that appear in tables scattered throughout the monograph may also be useful for this purpose.

A major source of dissatisfaction with the first volume was its lack of a subject index. We have attempted to remedy the situation by including a combined index to both volumes.

N.E. Breslow and N.E. Day