1

INTRODUCTION

1.1 WHAT IS A BIOMARKER?

According to the *Dictionary of Epidemiology*, a *biomarker* (or *biological marker*) is "a cellular, biochemical, or molecular indicator of exposure; of biological, subclinical, or clinical effects; or of possible susceptibility" (Porta 2008, p. 21). As Porta points out, the term "biomarker" is often ambiguous; this is perhaps an indication that there is insufficient understanding of the pathophysiological or mechanistic role of the "marker."

The ambiguity may also be due to the fact that biomarkers are involved in one way or another with so many different disciplines (clinical trialists, statisticians, regulators, etc.) and clinical research applications. In fact, there is so much potential ambiguity associated with the term "biomarker" that several efforts have been made to provide a formal definition of exactly what a biomarker is.

For example, in a 1987 US National Research Council report, biomarkers were defined to be "indicators signaling events in biological systems or samples." In 2001, a Biomarkers Definitions Working Group (BDWG), convened by the US National Institutes of Health, proposed the following definition of biological marker (biomarker): "A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention."

As interest in the development, validation, and application of new biomarkers has increased, numerous classification systems for biomarkers have been proposed. These include Type 0–Type 6 biomarkers; Type I and II biomarkers (Mildvan et al.

Analysis of Biomarker Data: A Practical Guide, First Edition. Stephen W. Looney and Joseph L. Hagan.

1997); prognostic and predictive biomarkers; genomic, proteomic, and combinatorial biomarkers; screening and stratification biomarkers, and so on. (See Table 1 of DeCaprio (2006) for details.) Most of these classification systems reflect the intended use of the biomarker data in a particular discipline; however, all biomarkers are related in the sense that each of them is designed to be an "indicator" of something, as noted in the *Dictionary of Epidemiology* definition cited above. Our primary focus in this book is on markers of exposure, although the statistical techniques we describe can be applied to almost any type of biomarker. By using "real" data taken from published biomarker studies to exemplify the proper application of these techniques, we have tried to illustrate the broad applicability of statistical methods in the analysis of biomarker data, regardless of the particular type of biomarker that is being considered.

1.2 BIOMARKERS VERSUS SURROGATE ENDPOINTS

In their report describing preferred definitions for biomarkers and surrogate endpoints, the BDWG defined a *clinical endpoint* as: "A characteristic or variable that reflects how a patient feels, functions, or survives." They then defined a *surrogate endpoint* as: "A biomarker that is intended to substitute for a clinical endpoint." A surrogate endpoint is thus expected to "predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence." As they pointed out, all surrogate endpoints are biomarkers, but not all biomarkers are surrogate endpoints. In fact, "it is likely that only a few biomarkers will achieve surrogate endpoint status." (Note that they discouraged the use of the term *surrogate marker*, and advocated the exclusive use of *surrogate endpoint* instead (BDWG 2001, p. 91).)

Because of the requirement that one must be able to substitute a surrogate endpoint in place of the corresponding clinical endpoint, the process of validating a surrogate endpoint goes far beyond what is usually required when validating a biomarker (see Chapter 5). In fact, the BDWG claimed that the term *validation* is unsuitable for describing the process of linking biomarkers to clinical endpoints; they proposed that the process of determining surrogate endpoint status be referred to as *evaluation.* They reserved use of *validation* to describe the process of addressing what they referred to as the "performance characteristics" (e.g., sensitivity, specificity, and reproducibility) of a measurement process or assay technique. This is consistent with our use of the term *biomarker validation* in Chapter 5.

Because of the complexity involved in evaluating a surrogate endpoint, various approaches have been proposed, almost all of which involve examining the effect of a treatment for the clinical endpoint (typically referred to as the "disease") on the surrogate for the endpoint. In a landmark paper, Prentice (1989) formulated a definition of surrogate endpoints and defined a set of operational criteria for their evaluation. In their subsequent work, Freedman et al. (1992) proposed that one should focus attention on the proportion of the treatment effect explained by the surrogate for the disease endpoint, whereas Buyse and Molenberghs (1998) proposed that the primary focus should be on the *relative effect* of the treatment on the surrogate. Various

authors also advocated the use of meta-analytic data in the evaluation of a surrogate endpoint (Freedman et al. 1992; Lin et al. 1997; Daniels and Hughes 1997). The application of meta-analytic techniques to surrogate endpoint evaluation was further developed by Buyse et al. (2000); Gail et al. (2000); Molenberghs et al. (2002); and others. The very comprehensive textbook edited by Burzykowski et al. (2005) thoroughly discuss all of these statistical approaches and subsequent developments. The Institute of Medicine report (Micheel and Ball 2010) approaches the evaluation of surrogate endpoints from a more clinical perspective.

Although surrogate endpoints are certainly a very important special case of biomarkers, we feel that the specialized techniques developed for evaluating them, especially as these techniques relate to treatment of the clinical endpoint, are beyond the scope of this text. Hence, we do not discuss surrogate endpoints as a separate topic elsewhere in this book. However, the methods that we describe for analyzing biomarker data and validating a biomarker (as defined by BDWG), certainly apply to surrogate endpoints as well.

1.3 ORGANIZATION OF THIS BOOK

In Chapter 1, we define what we mean by a biomarker and then describe our understanding of the differences and similarities between biomarkers and surrogate endpoints.

In Chapter 2, we cover basic principles of effective design of a study that will make use of biomarker data, including selecting the most appropriate type of study design (cross-sectional, case–control, etc.), choosing the appropriate measure of association once the type of design has been selected, designing the statistical analysis that will be applied to the study data once they have been obtained, and choosing the appropriate sample size for the study that is being planned. We also describe several features of what we consider to be the effective presentation of statistical results once the study data have been analyzed.

In Chapter 3, we provide a survey of elementary statistical methods that are widely used when analyzing biomarker data. To be specific, the methods that we cover include: graphical and tabular summaries; descriptive statistics; basic concepts of statistical inference, including point estimation, confidence interval estimation, and hypothesis testing; comparisons of means between two groups and among more than two groups; statistical inference for correlation coefficients; simple and multiple linear regression; and analysis of cross-classified data, including the chi-square test of independence and methods for comparing proportions across two or more groups. Our intention in this chapter is not to provide comprehensive coverage of all of elementary statistical methods, but rather to describe selected methods in sufficient detail so that someone who is relatively inexperienced in the application of statistics will be able to carry out these analyses appropriately and with a minimum of effort.

In Chapter 4, we describe various "challenges" that one is likely to encounter in the analysis of biomarker data and offer our recommendations on preferred methods for dealing with them. These challenges include: (1) violations of underlying assumptions

(normality, homogeneity of variance), (2) lack of independence between the groups being compared, (3) proper analysis of correlated data, (4) clustered data, (5) contaminated data, (6) non-detectable observations, (7) choosing the appropriate measure of association between predictor and outcome, and (8) choosing the appropriate method of analysis for cross-classified data (i.e., contingency tables). Each of these challenges is illustrated using data from a "real" biomarker study, most of which were taken from the scientific literature.

In Chapter 5, we provide a detailed discussion of the methods we recommend for evaluating the quality of a newly proposed (or existing) biomarker (also called *biomarker validation*). Our focus is on establishing that the biomarker has adequate reliability and validity.

Throughout Chapters 3–5, we provide what we hope is sufficient mathematical detail for those who are interested, but our primary emphasis is on the proper application of the statistical methods. Sections marked with an asterisk (*) contain a more theoretical treatment of the topic at hand and can safely be omitted without loss of continuity with the remainder of the text.

To the greatest extent possible, we provide software code for performing the statistical methods that we describe. Our software of choice is SAS because of its flexibility and widespread use in industry, government, and academia; but, in some instances, we also indicate how to perform an analysis using R (R Core Development Team 2014) or other statistical software (SPSS, STATA, etc.). The data sets for the fully worked examples in the book are provided along with the code used to analyze them. Shorter segments of code are included in the body of the text and are available on the companion website; longer segments are not provided in the text, but are available on the website.

We do not anticipate that the primary audience for this textbook will be students, so we have not provided extensive problem sets. However, we do recognize that exercises (with solutions) are an effective tool for anyone who is trying to learn how to perform a particular type of statistical analysis for the first time, or for someone who is trying to refresh their memory of statistical methods that they may have studied years ago. From personal experience, we know that exercises with solutions can be extremely helpful for experienced statisticians who are trying to learn about a statistical method they have never used before, or about applications of certain statistical methods in a scientific field that is new to them. Exercises with solutions are also useful as "test cases," when someone is trying to write their own software to carry out a statistical method for which easily accessible software is not available. With this in mind, we have provided small problem sets at the end of Chapters 3, 4, and 5. These contain exercises that are similar to the worked examples included in the text. To the greatest extent possible, these exercises are based on "real" data taken from published biomarker studies. Solutions to these exercises, including the SAS or R code needed to carry out the analyses, are provided at the end of the text.