

# Biostatistics

## I. INTRODUCTION

Statisticians contribute an essential intellectual component to most clinical trials (1,2). The following introduces statistical concepts and formulas that enable an understanding of information occurring at later points in this textbook. These concepts and formulas include the Kaplan–Meier plot, the Z statistic, the t statistic, P values, the hazard ratio, the concept of *sample group* versus *population group*, and the concept of *superiority* analysis versus *non-inferiority* analysis. An earlier chapter in this book will explain the concept of *intent-to-treat* analysis versus *per protocol* analysis.

### a. Kaplan–Meier plot

Kaplan–Meier plots are used to represent deaths occurring during the course of clinical trials in oncology, and hence are often called survival plots (3). But Kaplan–Meier plots are also used to represent other types of events in clinical trials, such as time to metastasis, time to a non-fatal heart attack, time to disappearance of pain in studies of arthritis drugs, and time to recovery from an infection after antibiotic treatment (4,5).

A Kaplan–Meier plot is a curve or, more accurately, a step–function. In this curve, the X axis is time and the Y axis is the *cumulative proportion* of study subjects experiencing the event of interest at any given time. In clinical trials in oncology, the event of interest is often death. Where a study subject dies, this death is shown by a downward (vertical) step in the curve. Intervals of time, during the clinical trial where there are no deaths, are shown by horizontal lines (no downward steps). Typically, horizontal components of the plot are shorter near the beginning of the clinical trial, because many subjects are participating in the trial (they have not yet died) and thus many subjects are at risk for triggering the event of interest, while horizontal components of the plot are longer near the end of the clinical trial, because relatively few subjects are still participating in the trial (6).

<sup>1</sup> The author thanks Dr. Jenna Elder, Ph.D., of PharPoint Research, Inc., Wilmington, NC, for reviewing the draft chapter and for providing insightful suggestions.

<sup>2</sup> The author thanks Dr. Harvey Motulsky, M.D., of GraphPad Software, Inc., La Jolla, CA, for reviewing the draft chapter and for responding with perceptive suggestions.

<sup>3</sup> Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53:457–481.

<sup>4</sup> Duerden M. *What are Hazard Ratios? What is...? Series.* Hayward Medical Communications. Hayward Group, Ltd.; 2009;8.

<sup>5</sup> Machin D, Gardner MJ. Calculating confidence intervals for survival time analyses. *Brit Med J.* 1988;296:1369–1371.

<sup>6</sup> Kirkwood BR, Sterne JA. *Essential Medical Statistics.* 2nd ed. Malden, MA: Blackwell Science Ltd.; 2003;278.