Biostatistics for Oncologists is the first practical guide providing the essential biostatistical concepts, oncology-specific examples, and applicable problem sets for medical oncologists, radiation oncologists, and surgical oncologists. In addition, it serves as a review for medical oncology and radiation oncology residents or fellows preparing for in-service and board exams. All examples are relevant to oncology and demonstrate how to apply core conceptual knowledge and applicable methods related to hypothesis testing, correlation and regression, categorical data analysis and survival analysis to the field of oncology. The book also provides guidance on the fundamentals of study design and analysis.

Written for oncologists by oncologists, this practical text demystifies challenging statistical concepts and provides concise direction on how to interpret, analyze, and critique data in oncology publications, as well as how to apply statistical knowledge to understanding, designing, and analyzing clinical trials. With practical problem sets and twenty-five multiple choice practice questions with answers, the book is an indispensable review for anyone preparing for in-service exams, boards, MOC, or looking to hone a lifelong skill.

**KEY FEATURES:**

- Practically explains biostatistics concepts important for passing the hematology, medical oncology, and radiation oncology boards and MOC exams
- Provides guidance on how to read, understand, and critique data in oncology publications
- Gives relevant examples that are important for analyzing data in oncology, including the design and analysis of clinical trials
- Tests your comprehension of key biostatistical concepts with problem sets at the end of each section and a final section devoted to board-style multiple choice questions and answers
- Includes digital access to the eBook

**Recommended Shelving Category:** Oncology

ISBN 978-0-8261-6858-0
Biostatistics for Oncologists
Biostatistics for Oncologists

Kara Lynne Leonard, MD, MS
Assistant Professor of Radiation Oncology
Alpert Medical School of Brown University
Providence, Rhode Island

Adam J. Sullivan, PhD
Assistant Professor of Biostatistics
Alpert Medical School of Brown University
Providence, Rhode Island
To my family—Thank you for everything

To David Wazer—Thank you for your wisdom and encouragement in this and many other endeavors

Kara Lynne Leonard
Preface xiii

Share Biostatistics for Oncologists

SECTION I. GENERAL STATISTICAL CONCEPTS

1. Why Study Biostatistics? 3
   What Is Biostatistics? 3
   How Is Biostatistics Useful for Oncologists 3

2. Summarizing and Graphing Data 5
   Types of Data 5
      Quantitative Data 5
         Discrete Data 5
         Continuous Data 5
      Qualitative Data 5
         Nominal Data 5
         Ordinal Categorical Data 6
   Data Summaries 7
      Measures of Central Tendency 7
         Mean 8
         Median 8
         Mode 9
      Measures of Dispersion 9
         Variance 9
         Standard Deviation 9
         Interquartile Range 11
   Statistical Graphs 12
      Histograms 12
      Box Plot 13
      Scatter Plot 13
3. Sampling 15
   Populations and Sample 15
   Simple Random Sample 16
   Other Sampling Methods 16

4. Statistical Estimation 17
   Some Basic Distributions 17
      Normal Distribution 17
         Central Limit Theorem 18
         Student’s T-Distribution 19
         Standard Error of the Mean 19
      Binomial Distribution 20
      Poisson Distribution 21
   Estimation 29
      Point Estimates 29
      Interval Estimates 29

Section I Problem Set 33

Section I Problem Set Solutions 39

SECTION II. IMPORTANT STATISTICAL CONCEPTS FOR ONCOLOGISTS

5. Hypothesis Testing 45
   Type I and Type II Errors 46
   p-Values 48
   t-Tests 49
      One-Tailed Versus Two-Tailed 50
      Independent Samples 53
      Paired Data 53
   Wilcoxon Tests 54
      Wilcoxon Rank-Sum Test 54
      Wilcoxon Signed-Rank Test 54
   Analysis of Variance 54
   Testing Binomial Proportions 55
   Confidence Intervals and Hypothesis Tests: How Are They Related? 56
Sensitivity and Specificity  59
Negative Predictive Value  63
Positive Predictive Value  63
Positive Likelihood Ratio  63
Negative Likelihood Ratio  64

6. Correlation and Regression  65
Correlation  65
  Pearson Correlation Coefficient  65
  Spearman Rank Correlation  68
Regression  69
  Simple Linear Regression  70
  Multiple Linear Regression  72
  Logistic Regression  72

7. Categorical Data Analysis  77
Contingency Tables  77
  2 × 2 Tables  77
  R × C Tables  78
  Fisher’s Exact Test  79
  Chi-Square Test  80
  Chi-Square Test Versus Logistic Regression  82
Effect Size Estimators  83
  Relative Risk  83
  Odds Ratio  85
  Relative Risk Versus Odds Ratio  87
McNemar’s Test  88
Mantel–Haenszel Method  91
  Homogeneity Test  92
  Summary Odds Ratio  93

8. Survival Analysis Methods  95
Time-to-Event Data  95
Kaplan–Meier Curves  95
  Median Survival  100
Log-Rank Test  101
Wilcoxon Rank-Sum Test  102
Cox Proportional Hazards Model  103
Hazard Ratio  104

9. Guide to Choosing the Appropriate Statistical Test  105

10. Noninferiority Analysis  107

Section II Problem Set  109

Section II Problem Set Solutions  125

SECTION III. BASICS OF EPIDEMIOLOGY

11. Study Designs  141
   Experimental Studies  141
   Clinical Trials  141
      Common Outcomes for Clinical Trials in Oncology  141
      Phase I Clinical Trials  143
      Phase II Clinical Trials  147
      Phase III Clinical Trials  150
      Phase IV Clinical Trials  152
      Meta-Analysis  153
      Primary Outcomes  154
   Field Trials  154
      Community Intervention Trials  154
   Nonexperimental Studies  155
      Cohort Studies  155
      Retrospective Cohort Study  155
      Prospective Cohort Study  156
      Case-Control Studies  156
      Cohort Studies Versus Case-Control Studies  157
      Cross-Sectional Studies  158
      Matched Studies  158
   Analysis of Studies  159
      Crude Analysis  159
      Bias  159
      Confounding  160
Stratified Analysis  161
Effect Modification  161
Connections to Regressions  162
Sample Size  163
Hypothesis Testing  163
t-Test  164
  Testing Binomial Proportions  164
Survival Analysis  164
Regression Analysis  164

Section III Problem Set  167
Section III Problem Set Solutions  169

SECTION IV. SELF-ASSESSMENT QUESTIONS AND ANSWERS

12. Self-Study Multiple Choice Questions  173
13. Self-Study Answers and Rationales  179

Index  183
As a radiation oncology resident, I was often reminded of the importance of biostatistics in the study of oncology. An understanding of biostatistics is necessary for reading and comprehending published literature, for performing retrospective research, and for designing and analyzing prospective clinical trials. Biostatistical concepts are also tested on oncology board exams.

I noticed that, unlike radiation biology and medical physics education, statistics education was much less standardized. Biostatistics courses for oncology trainees are often taught by biostatisticians without a background in oncology or by MDs without formal statistical training. As such, many practitioners, both trainees and attending physicians, have a limited understanding of biostatistics.

I have had some formal statistical training as part of a MS program in experimental psychology. The MS program statistics courses focused heavily on experimental design concepts directly applicable to the biostatistics used in oncology. With this statistics background, I came to recognize that there are a few core statistical concepts critical to the practice of oncology. These concepts are repeated over and over again. To develop a comprehension of these concepts, I would pay careful attention when I read about them in newly published papers, and I would apply them to my own retrospective research.

The key, as I saw it, to understanding biostatistics in oncology was to focus on the core concepts and to apply those concepts to oncology-specific examples. I began to develop tools and mnemonics to share with my co-residents. As a junior attending, I designed a biostatistics course for radiation oncology residents in our program. I have been teaching the course annually for the past 4 years. The course has been well received, and residents have noted an improved understanding of sensitivity and specificity, categorical data analysis, and time-to-event data analysis. Residents have also mentioned to me that their scores on the statistics portion of the in-service exam have improved after taking the course.

In talking with my colleagues in medical oncology and surgical oncology, I learned that the need for basic, oncology-specific education in biostatistics was not unique to radiation oncology. Each specialty shared
enthusiasm for a concise resource that simplified biostatistics concepts relevant to the field.

As I became even more aware of the need for an oncology-specific biostatistics text, I reached out to Adam, a biostatistician who focuses on education, and together, we created *Biostatistics for Oncologists*.

The book begins with the basic foundations of biostatistics that are tested on board exams (Section 1). In Section 2, these basics are then expanded on to include the concepts used in retrospective study design, analysis, and interpretation. The final section (Section 3) focuses on prospective clinical trials, guiding readers in their understanding of published clinical trials and in the design and analysis of novel clinical trials.

Most importantly, *Biostatistics for Oncologists* presents each concept with an example relevant to oncology. All examples focus on oncology-specific outcomes such as tumor response rate and progression-free survival. The section on Phase I clinical trials includes an example of radiotherapy dose escalation trials. The text also includes oncology-specific problem sets and oncology-specific multiple-choice questions for self-assessment and board study.

All of us in the field use our knowledge of biostatistics to read and evaluate published oncology literature. My hope is that oncologists reading a newly published clinical trial will reference this text to understand and evaluate the statistics section and, will use this text when developing an abstract for an annual meeting, or will flip through the book when developing concepts for new clinical trials. Hopefully, after reading this book, you will find biostatistics less overwhelming, more manageable and applicable to your practice, and even enjoyable.

*Kara Lynne Leonard, MD, MS*
Share
Biostatistics for Oncologists
The purpose of research and data analysis is to study and make conclusions about a population. Examples of populations in oncology include patients with stage III lung cancer, patients with metastatic breast cancer, or patients who receive a new chemotherapeutic agent or radiotherapy with a novel technique.

**POPULATIONS AND SAMPLE**

Because it is often inconvenient, impractical, or impossible to study an entire population, a sample typically has to be chosen to represent the population. This representative sample is the group that will be studied to make determinations about an entire population. If the sample appropriately typifies a population, conclusions drawn about the sample may be directly applied to the population at large.

**Example**

Let us use an example from the Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy (STAMPEDE) trial looking at the use of systemic therapy in advancing or metastatic prostate cancer (1). This study concerns a population of men with advanced or metastatic prostate cancer. This is an enormous population, and all men fitting this description cannot be practically studied. To represent this population, a sample of patients is enrolled. In this case, the sample consisted of 2,962 men enrolled at multiple participating sites. The men in the sample were randomly assigned to receive hormone therapy alone, hormone therapy plus zoledronic acid, hormone therapy plus docetaxel, or hormone therapy plus docetaxel and zoledronic acid. The primary end point was overall survival. Docetaxel chemotherapy, given at the time of long-term hormone therapy initiation, showed evidence of improved survival accompanied by an increase in adverse events. This conclusion is then extrapolated to the population at large and the authors recommend that docetaxel chemotherapy should become part of the standard of care for men in this population (those with advanced or metastatic prostate cancer) commencing long-term hormone therapy.
SIMPLE RANDOM SAMPLE

The most scientifically appropriate sample is a simple random sample. In the case of a simple random sample, each member of the sample is chosen at random and, as such, each person in the population has the exact same chance of being chosen for a sample. Such simple random samples are not practical and most of the time not even possible for trials in oncology.

Example

A feasible example of a simple random sample in oncology is as follows: An attending physician working with a second-year medical student wants to develop a manageable project for her. The attending physician is interested in looking at outcomes for all patients with stage IV colon cancer treated at his institution over the past 15 years. This population includes about 500 patients, and he would like to have the student review the records of a sample of 20%, or 100, of these patients. The attending physician suggests that the medical student choose a simple random sample to represent the population. He suggests that she enter all 500 patients into Microsoft Excel and then use the random number generator (RAND) to select the simple random sample, taking care that the same number is not chosen more than once. The data could then be collected on this more manageable sample to represent and draw conclusions about the entire population.

OTHER SAMPLING METHODS

After defining the population of interest, a sample must be selected. This can be done in a variety of ways. Other sampling methods include probability sampling such as systematic sampling, stratified sampling, probability-proportional-to-size sampling, cluster sampling, quota sampling, minimax sampling, accidental sampling, line-intercept sampling, panel sampling, snowball sampling, or theoretic sampling.

The most commonly used probability sampling methods in medicine are systematic sampling, stratified sampling, and cluster sampling.

REFERENCE