Contents

Preface *xiii*

Acronyms *xvii*

1 The Basics *1*

1.1 Distinguish Randomized and Observational Studies 2

1.2 Beware of Linear Models 3

1.3 Understand Omnibus Quantities 6

1.4 Independence, Equal Variance, and Normality 7

1.5 Models As Simple As Possible, But Not More Simple 11

1.6 Do Not Multiply Probabilities More Than Necessary 12

1.7 Know the Sample Space for Statements of Risk 13

1.8 Use Two-sided p-Values 14

1.9 p-Values for Sample Size, Confidence Intervals for Results 16

1.10 Use at Least Twelve Observations in Constructing a Confidence Interval 18

1.11 Know the Unit of the Variable 19

1.12 Know Properties Preserved When Transforming Units 20

1.13 Be Flexible About Scale of Measurement Determining Analysis 23
1.14 Be Eclectic and Ecumenical in Inference 24
1.15 Consider Bootstrapping for Complex Relationships 25
1.16 Standard Error from Sample Range/Sample Size 26

2 Sample Size *29*
2.1 Begin with a Basic Formula for Sample Size 31
2.2 No Finite Population Correction for Survey Sample Size 33
2.3 Calculating Sample Size Using the Coefficient of Variation 35
2.4 Do Not Formulate a Study Solely in Terms of Effect Size 38
2.5 Overlapping Confidence Intervals Do Not Imply Nonsignificance 39
2.6 Sample Size Calculation for the Poisson Distribution 40
2.7 Sample Size for Poisson With Background Rate 41
2.8 Sample Size Calculation for the Binomial Distribution 43
2.9 When Unequal Sample Sizes Matter; When They Don't 45
2.10 Sample Size With Different Costs for the Two Samples 47
2.11 The Rule of Threes for 95% Upper Bounds When There Are No Events 49
2.12 Sample Size Calculations Are Determined by the Analysis 50

3 Covariation *53*
3.1 Assessing and Describing Covariation 55
3.2 Don't Summarize Regression Sampling Schemes with Correlation 56
3.3 Do Not Correlate Rates or Ratios Indiscriminately 58
3.4 Determining Sample Size to Estimate a Correlation 59
3.5 Pairing Data is not Always Good 61
3.6 Go Beyond Correlation in Drawing Conclusions 63

3.7 Agreement As Accuracy, Scale Differential, and Precision 65
3.8 Assess Test Reliability by Means of Agreement 68
3.9 Range of the Predictor Variable and Regression 70
3.10 Measuring Change: Width More Important than Numbers 72

4 Epidemiology *75*
4.1 Start with the Poisson to Model Incidence or Prevalence 76
4.2 The Odds Ratio Approximates the Relative Risk Assuming the Disease is Rare 77
4.3 The Number of Events is Crucial in Estimating Sample Sizes 82
4.4 Using a Logarithmic Formulation to Calculate Sample Size 84
4.5 Take No More than Four or Five Controls per Case 86
4.6 Obtain at Least Ten Subjects for Every Variable Investigated 87
4.7 Begin with the Exponential Distribution to Model Time to Event 89
4.8 Begin with Two Exponentials for Comparing Survival Times 91
4.9 Be Wary of Surrogates 92
4.10 Prevalence Dominates in Screening Rare Diseases 95
4.11 Do Not Dichotomize Unless Absolutely Necessary 99
4.12 Select an Additive or Multiplicative Model on the Basis of Mechanism of Action 100

5 Environmental Studies *103*
5.1 Think Lognormal 103
5.2 Begin with the Lognormal Distribution in Environmental Studies 104
5.3 Differences Are More Symmetrical 106
5.4 Beware of Pseudoreplication 108

5.5 Think Beyond Simple Random Sampling 109
5.6 Consider the Size of the Population Affected by Small Effects 111
5.7 Statistical Models of Small Effects Are Very Sensitive to Assumptions 112
5.8 Distinguish Between Variability and Uncertainty 113
5.9 Description of the Database is As Important as Its Data 115
5.10 Always Assess the Statistical Basis for an Environmental Standard 116
5.11 Measurement of a Standard and Policy 117
5.12 Parametric Analyses Make Maximum Use of the Data 119
5.13 Distinguish Between Confidence, Prediction, and Tolerance Intervals 120
5.14 Statistics Plays a Key Role in Risk Assessment, Less in Risk Management 122
5.15 Exposure Assessment is the Weak Link in Assessing Health Effects of Pollutants 124
5.16 Assess the Errors in Calibration Due to Inverse Regression 125

6 Design, Conduct, and Analysis 129

6.1 Randomization Puts Systematic Effects into the Error Term 129
6.2 Blocking is the Key to Reducing Variability 131
6.3 Factorial Designs Should be Used to Assess Joint Effects of Variables 132
6.4 High–Order Interactions Occur Rarely 134
6.5 Balanced Designs Allow Easy Assessment of Joint Effects 136
6.6 Analysis Follows Design 137
6.7 Plan to Graph the Results of an Analysis 139
6.8 Distinguish Between Design Structure and Treatment Structure 142
6.9 Make Hierarchical Analyses the Default Analysis 143

6.10 Distinguish Between Nested and Crossed Designs—Not Always Easy 145
6.11 Plan for Missing Data 146
6.12 Address Multiple Comparisons Before Starting the Study 149

7 Words, Tables, and Graphs *153*
7.1 Use Text for a Few Numbers, Tables for Many Numbers, Graphs for Complex Relationships 153
7.2 Arrange Information in a Table to Drive Home the Message 155
7.3 Always Graph the Data 158
7.4 Never Use a Pie Chart 160
7.5 Bargraphs Waste Ink; They Don't Illuminate Complex Relationships 162
7.6 Stacked Bargraphs Are Worse Than Bargraphs 163
7.7 Three-Dimensional Bargraphs Constitute Misdirected Artistry 166
7.8 Identify Cross-sectional and Longitudinal Patterns in Longitudinal Data 167
7.9 Use Rendering, Manipulation, and Linking in High Dimensional Data 170

8 Consulting *175*
8.1 Structure a Consultation Session to Have a Beginning, a Middle, and an End 176
8.2 Ask Questions 177
8.3 Make Distinctions 178
8.4 Know Yourself, Know the Investigator 180
8.5 Tailor Advice to the Level of the Investigator 181
8.6 Use Units the Investigator is Comfortable With 182
8.7 Agree on Assignment of Responsibilities 184
8.8 Any Basic Statistical Computing Package Will Do 185
8.9 Ethics Precedes, Guides, and Follows Consultation 186
8.10 Be Proactive in Statistical Consulting 187

8.11 Use the Web for Reference, Resource, and Education 189
8.12 Listen to, and Heed the Advice of Experts in the Field 190

Epilogue *193*

References *195*

Author Index *207*

Topic Index *211*