

# Contents

PREFACE.....	VII
ACKNOWLEDGMENTS.....	IX
CONTENTS.....	XI
INTRODUCTION.....	XVII
PROLOGUE.....	1
<i>Europe's Emergence from the Middle Ages.....</i>	<i>1</i>
<i>Absolutism.....</i>	<i>3</i>
<i>Refusing to be Counted.....</i>	<i>4</i>
<i>No Need for Probability.....</i>	<i>5</i>
<i>Intellectual Triumph: The Industrial Revolution.....</i>	<i>6</i>
<i>Reasoning from a Sample.....</i>	<i>7</i>
<i>Political Arithmetic.....</i>	<i>8</i>
<i>The Role of Religion in Political Arithmetic.....</i>	<i>8</i>
<i>Probability and the Return to Order.....</i>	<i>10</i>
<i>"Let Others Thrash It Out!".....</i>	<i>11</i>
<i>Early Experimental Design.....</i>	<i>11</i>
<i>Agricultural Articulations.....</i>	<i>12</i>
James Johnson.....	13
<i>Fisher, Gosset, and Modern Experimental Design.....</i>	<i>13</i>
<i>References.....</i>	<i>14</i>
1. THE BASIS OF STATISTICAL REASONING IN MEDICINE.....	17
1.1 <i>What Is Statistical Reasoning?.....</i>	<i>17</i>
1.1.2 Physicians and the Patient Perspective.....	18
1.1.2 Research and the Population Perspective.....	19
1.1.3 The Need for Integration.....	20
1.1.4 A Trap.....	20
1.1.5 Clinical Versus Research Skills.....	21
1.2 <i>Statistical Reasoning.....</i>	<i>22</i>
1.2.1 The Great Compromise.....	23
1.2.2 Example of Sampling Error.....	24
1.2.3 PRAISE I and II.....	26
1.2.4 Fleas.....	28
1.3 <i>Generalizations to Populations.....</i>	<i>29</i>
1.3.1 Advantages of the Random Sample.....	29
1.3.2 Limitations of the Random Sample.....	30
1.3.3 Example: Salary Estimation.....	31
1.3.4 Difficulty 1: Sample Size and Missing Data.....	32
1.3.5 Sample Vision.....	33
<i>References.....</i>	<i>35</i>

2. SEARCH VERSUS RESEARCH.....	37
2.1 Introduction .....	37
2.2 Catalina's Dilemma.....	37
2.2.1. Can Catalina Generalize? .....	37
2.2.2 Do Logistical Issues Block Generalization? .....	38
2.2.3 Disturbed Estimators .....	38
2.3 Exploratory Analysis and Random Research.....	40
2.4 Gender–Salary Problem Revisited.....	42
2.5 Exploratory Versus Confirmatory.....	44
2.5.1 Cloak of Confirmation.....	45
2.6 Exploration and MRFIT.....	46
2.7 Exploration in the ELITE Trials .....	47
2.8 Necessity of Exploratory Analyses.....	48
2.8.1 Product 2254RP.....	49
2.8.2 The Role of Discovery Versus Confirmation .....	49
2.8.3 Tools of Exploration.....	50
2.9 Prospective Plans and “Calling Your Shot” .....	50
2.9.1 The US Carvedilol Program .....	51
2.9.2 Let the Data Decide!.....	54
2.10 Tight Protocols .....	55
2.11 Design Manuscripts .....	56
2.12 Concordant Versus Discordant Research.....	57
2.12.1 Severe Discordance: Mortality Corruption.....	57
2.12.2 Severe Discordance: Medication Changes .....	58
2.12.3 Discordance and NSABP .....	59
2.13 Conclusions.....	59
References.....	60
3. A HYPOTHESIS-TESTING PRIMER.....	63
3.1 Introduction .....	63
3.2 The Rubric of Hypothesis Testing.....	64
3.3 The Normal Distribution and Its Estimators.....	65
3.4 Using the Normal Distribution .....	66
3.4.1 Simplifying Transformations.....	66
3.4.3 Symmetry .....	68
3.5 The Null Hypothesis: State of the Science.....	70
3.6 Type II Error and Power.....	76
3.7 Balancing Alpha and Beta .....	80
3.8 Reducing Alpha and Beta: The Sample Size .....	81
3.9 Two-Sided Testing.....	83
3.10 Sampling Error Containment.....	86
3.11 Confidence Intervals .....	87
3.12 Hypothesis Testing in Intervention Studies .....	89
3.13 Community Responsibility .....	89
4. MISTAKEN IDENTITY: P-VALUES IN EPIDEMIOLOGY .....	91
4.1 Mistaken Identity.....	91
4.2 Detective Work.....	91
4.2.1 Association versus Causation .....	92
4.3 Experimental Versus Observational Studies .....	92

4.3.1 The Role of Randomization.....	92
4.3.2 Observational Studies.....	94
4.4 <i>Determining Causation</i> .....	95
4.4.1 Causality Tenets.....	96
4.5 <i>Clinical Significance Without P-Values</i> .....	98
4.5.1 Thalidomide.....	98
4.5.2 The Radium Girls.....	99
4.6 <i>Tools of the Epidemiologist</i> .....	100
4.6.1 Case Reports and Case Series.....	100
4.6.2 <i>Categories of Observational Studies</i> .....	104
4.6.2.1 Directionality: Forward or Backward?.....	104
4.6.2.2 Retrospective Versus Prospective.....	105
4.6.3 Variable Adjustments.....	107
4.7 <i>Fenfluramines</i> .....	108
4.8 <i>Design Considerations</i> .....	109
4.9 <i>Solid Structures from Imperfect Bricks</i> .....	110
4.10 <i>Drawing Inferences in Epidemiology</i> .....	111
4.11 <i>Study counting: The ceteris paribus fallacy</i> .....	112
4.12 <i>Critiquing Experimental Designs</i> .....	113
4.13 <i>Conclusions</i> .....	113
<i>References</i> .....	114
5. SHRINE WORSHIP.....	117
5.1 <i>Introduction</i> .....	117
5.2 <i>The Nightmare</i> .....	117
5.3 <i>P-value Bashing</i> .....	118
5.4 <i>Epidemiology and Biostatistics</i> .....	118
5.4.1 The Link Between Epidemiology and Statistics.....	119
5.4.2. Exposure–Disease Relationships.....	119
5.5 <i>The Initial Schism</i> .....	120
5.6 <i>Appearance of Statistical Significance</i> .....	122
5.6.1 Introducing the 0.05 Level.....	122
5.6.2 “Dangerous Nonsense”.....	124
5.7 <i>The P-value Love Affair in Healthcare</i> .....	126
5.8 <i>Use and Abuse of P-values</i> .....	127
5.8.1 Confidence Intervals as False Substitution.....	129
5.8.2 War Too Important to be Left to the Generals?.....	130
5.9 <i>Proper Research Interpretation</i> .....	131
<i>References</i> .....	133
6. P-VALUES, POWER, AND EFFICACY.....	137
6.1 <i>Introduction</i> .....	137
6.2 <i>P-values and Strength of Evidence</i> .....	137
6.2.1 Example.....	138
6.3 <i>Power</i> .....	141
6.4 <i>No Way Out?</i> .....	143
6.5 <i>Sample Size Computations</i> .....	144
6.5.1 The Anvil.....	144
6.6. <i>Non-statistical Considerations</i> .....	145
6.6.1 The LRC Sample Size.....	146

6.7	<i>The “Good Enough for Them” Approach</i> .....	147
6.8	<i>Efficacy Seduction</i> .....	147
6.8.1	Efficacy and Sample Size.....	148
6.8.2	Large P-values and Small Effect Sizes.....	149
6.9	<i>Number Needed To Treat</i> .....	150
6.10	<i>Absolute versus Relative Risk</i> .....	151
6.11	<i>Matching Statistical with Clinical Significance</i> .....	153
6.12	<i>Power for Smaller Efficacy Levels</i> .....	155
6.13	<i>Conclusions</i> .....	156
	<i>References</i> .....	156
7.	SCIENTIFIC REASONING, P-VALUES, AND THE COURT .....	157
7.1	<i>Introduction</i> .....	157
7.2	<i>Blood Pressure and Deception: The Frye Test</i> .....	158
7.3	<i>Rule 402</i> .....	159
7.4	<i>The Daubert Rulings</i> .....	160
7.5	<i>The Havner Ruling</i> .....	161
7.6	<i>Relative Risk and the Supreme Court</i> .....	163
7.7	<i>P-values, Confidence Intervals, and the Courts</i> .....	164
7.8	<i>Conclusions</i> .....	165
	<i>References</i> .....	165
8.	ONE-SIDED VERSUS TWO-SIDED TESTING .....	167
8.1	<i>Introduction</i> .....	167
8.2	<i>Attraction of One-Sided Testing</i> .....	167
8.3	<i>Belief Versus Knowledge in Healthcare</i> .....	167
8.4	<i>Belief Systems and Research Design</i> .....	168
8.5	<i>Statistical Versus Ethical Optimization</i> .....	169
8.6	<i>“Blinded by the Light”: CAST</i> .....	170
8.7	<i>LRC Results</i> .....	173
8.8	<i>Sample Size Issues</i> .....	174
8.9	<i>Hoping for the Best, Preparing for the Worst</i> .....	175
8.10	<i>Symmetrics versus Ethics</i> .....	176
8.11	<i>Conclusions</i> .....	179
	<i>References</i> .....	180
9.	MULTIPLE TESTING AND COMBINED ENDPOINTS.....	181
9.1	<i>Introduction</i> .....	181
9.2	<i>Definition of Multiple Analyses</i> .....	182
9.3	<i>Efficiency Versus Parsimony</i> .....	182
9.3.1	Efficiency.....	183
9.3.2	Epidemiologic Strength.....	183
9.3.3	The Need To Explore.....	184
9.4	<i>Hypothesis Testing in Multiple Analyses</i> .....	184
9.4.1	Don’t Ask, Don’t Tell.....	184
9.5	<i>Familywise Error</i> .....	186
9.7	<i>The Bonferroni Inequality</i> .....	186
9.8	<i>Is Tight Control of the FWER Necessary?</i> .....	188
9.9	<i>Alternative Approaches</i> .....	190

9.9.1	Sequentially Rejective Procedures .....	190
9.9.2	Resampling P-values .....	190
9.10	<i>Analysis Triage</i> .....	191
9.10.1	Primary Versus Secondary Analyses .....	192
9.10.2	Secondary Analyses .....	192
9.10.3	Example of Endpoint Triaging .....	192
9.11	<i>Combined Endpoints</i> .....	194
9.12	<i>Why Use Combined Endpoints</i> .....	194
9.12.1	Epidemiologic Considerations .....	195
9.12.2	Sample Size Concerns .....	195
9.12.3	Improved Resolving Power .....	195
9.13	<i>Combined Endpoint Construction</i> .....	196
9.13.1	Property 1: Coherence .....	196
9.13.2	Property 2: Equivalence .....	197
9.13.3	Therapy Homogeneity .....	198
9.14	<i>Measuring Combined Endpoints</i> .....	199
9.14.1	Prospective Identification .....	199
9.14.2	Ascertaining Endpoints .....	200
9.15	<i>Conclusions</i> .....	201
	<i>References</i> .....	202
10.	SUBGROUP ANALYSES .....	205
10.1	<i>Bona Fide Gems or Fool's Gold</i> .....	205
10.2	<i>What Are Subgroups?</i> .....	205
10.3	<i>The Amlopidine Controversy</i> .....	206
10.4	<i>Definitions</i> .....	208
10.5	<i>Interpretation Difficulties</i> .....	208
10.5.1	Re-evaluating the Same Patients .....	208
10.5.2	Random Findings .....	209
10.5.3	Clinical Trial-Mediated Subgroup "Effects" .....	211
10.5.4	The Importance of Replication .....	215
10.6	<i>Stratified Randomization</i> .....	216
10.7	<i>Proper Versus Improper Subgroups</i> .....	217
10.8	<i>"Intention-to-Treat" Versus "As Treated"</i> .....	218
10.9	<i>Effect Domination Principle</i> .....	219
10.10	<i>Confirmatory Subgroup Analyses</i> .....	220
10.11	<i>Assessment of Subgroup Effects</i> .....	221
10.11.1	Effect Modification and Interactions .....	221
10.11.2	Within-Stratum Effects .....	222
10.12	<i>Data Dredging — Caveat Emptor</i> .....	224
10.13	<i>Conclusions</i> .....	224
	<i>References</i> .....	225
11.	P-VALUES AND REGRESSION ANALYSES .....	229
11.1	<i>The New Meaning of Regression</i> .....	229
11.2	<i>Assumptions in Regression Analysis</i> .....	230
11.3	<i>Model Estimation</i> .....	231
11.3.1	Cross-Sectional Versus Longitudinal .....	232
11.4	<i>Variance Partitioning</i> .....	233
11.5	<i>Enter Dichotomous Explainer Variables</i> .....	236
11.6	<i>The Meaning of "Adjustment"</i> .....	237

11.7 <i>Super Fits</i> .....	239
11.8 <i>Pearls of Great Price</i> .....	241
11.9 <i>Effect Modifiers and Alpha Allocation</i> .....	242
11.9.1 <i>Effect Modification Models</i> .....	242
11.9.2 <i>The Difficulty of Effect Modification</i> .....	244
11.10 <i>Conclusions</i> .....	247
12. BAYESIAN ANALYSIS: POSTERIOR <i>P</i> -VALUES .....	249
12.1 <i>An Arbitrary Process</i> .....	249
12.2 <i>The Frequentists</i> .....	250
12.2.1 <i>The Frequentist and Long-Term Accuracy</i> .....	250
12.2.2 <i>Reverse Perspective</i> .....	250
12.2.3 <i>The Likelihood Principle</i> .....	251
12.3 <i>The Bayesian Philosophy</i> .....	255
12.3.1 <i>Bayes and Heart Failure</i> .....	255
12.4 <i>Feasibility of Prior Distributions</i> .....	263
12.5 <i>The Loss Function</i> .....	265
12.6 <i>Example of Bayesian Estimation</i> .....	265
12.7 <i>Bayesians and P-values</i> .....	266
12.8 <i>Bayes Testing: Asthma Prevalence</i> .....	267
12.9 <i>Conclusions</i> .....	271
<i>References</i> .....	272
CONCLUSIONS: .....	273
APPENDIX A. STANDARD NORMAL PROBABILITIES .....	279
APPENDIX B: SAMPLE SIZE PRIMER.....	283
<i>B.1 General Discussion of Sample Size</i> .....	283
<i>B.2 Derivation of Sample Size</i> .....	285
B.2.1 <i>Phase 1: Under the Null Hypothesis</i> .....	286
B.2.2 <i>Phase 2: Under the Alternative Hypothesis</i> .....	286
B.2.3 <i>Phase 3: Consolidation</i> .....	287
<i>B.3 Example</i> .....	288
<i>B.4 Continuous Outcomes</i> .....	289
B.4.1 <i>Phase I: The Null Hypothesis</i> .....	290
B.4.2 <i>Phase II: The Alternative Hypothesis</i> .....	290
B.4.3 <i>Phase III: Consolidation</i> .....	291
B.4.4 <i>Example</i> .....	292
<i>References</i> .....	293
APPENDIX C: DAUBERT AND RULE 702 FACTORS .....	295
<i>C.1 The Daubert Factors</i> .....	295
<i>C.2 The 702 Factors</i> .....	295
INDEX.....	297