

# Contents

List of Statistical Notations: Commonly Used Names and Symbols	x
Foreword, by A. JOHN RUSH	xi
Preface	xv
Acknowledgments	xxi
1. Introduction	1
2. Disorder and Diagnosis	5
<i>The Disorder</i>	5
<i>The Diagnosis of the Disorder</i>	6
<i>Diagnostic and Prognostic Tests</i>	8
<i>The Basic Model</i>	9
<i>The Prevalence and Reproducibility of the Diagnosis</i>	14
<i>High- and Low-Risk Populations</i>	15
<i>Pre- and Posttest Risk, Prevalence, and Incidence: Confusing Terms</i>	16
<i>Summary</i>	17
3. Definition: Test Protocol, Response, Referent	18
<i>The Basic Model</i>	20
<i>Summary</i>	21
4. Families of Test Referents	23
<i>Definitions: Base and Family</i>	23
<i>A Special Family: Multiple Discriminant Rules</i>	25
<i>Summary</i>	25

5. Population and Sampling	26
<i>Outcomes: True and False Positives and Negatives</i>	26
<i>Blind Evaluation</i>	28
<i>The Basic Model</i>	29
<i>Descriptors of Outcomes: Sensitivity, Specificity, Predictive Values, and Efficiency</i>	33
<i>Naturalistic Sampling</i>	34
<i>The Two-by-Two Chi-Square Test</i>	40
<i>Retrospective Sampling</i>	44
<i>The Problem of Dropouts</i>	49
<i>Prospective Sampling</i>	53
<i>The Sampling Unit</i>	57
<i>Pseudo-Retrospective Sampling: An Invalid Approach</i>	58
<i>Summary</i>	61
6. Sensitivity and Specificity: The Signal Detection Approach	63
<i>Problems with Sensitivity and Specificity</i>	63
<i>The Basic Model</i>	65
<i>The Receiver Operating Characteristic Curve (Test ROC)</i>	71
<i>The Quality Receiver Operating Characteristic Curve (QROC)</i>	75
<i>All Q-Level Tests</i>	77
<i>What Is a Good Test? The Diagnosis ROC, QROC</i>	82
<i>What Is the Ideal Test? The Disorder ROC, QROC</i>	86
<i>The Myth of the Constancy of Sensitivity and Specificity</i>	91
<i>Summary</i>	94
7. Predictive Values: The Bayesian Approach, Risk Ratios, and Odds Ratio	96
<i>Problems with Predictive Values</i>	96
<i>The Basic Model</i>	98
<i>A Geometric Approach: The Test QROC</i>	100
<i>Risk Ratios, Odds Ratio, and the Test QROC</i>	103
<i>Summary</i>	113
8. Efficiency: Choosing Clinically Optimal Tests	114
<i>The Problem with Efficiency</i>	114
<i>The Basic Model</i>	115
<i>Consideration of Medical Consequences of Errors: The Weighted Kappa Coefficient</i>	119
<i>A Geometric Approach: The Test QROC</i>	123
<i>Summary</i>	129
9. Taking Test Costs into Account: Costworthy Tests	131
<i>Introduction</i>	131
<i>The Basic Model</i>	135
<i>Derivation of the Quality Index <math>\kappa(r,t)</math></i>	135

	<i>The Benefit Threshold of a Medical Test</i>	138
	<i>Full Evaluation of a Test Family</i>	140
	<i>A Geometric Approach: The QCROC</i>	145
	<i>The Costs and Risks of Too Much Information</i>	149
	<i>Informed Consent for Medical Tests</i>	151
	<i>Summary</i>	153
10.	<b>Basic Issues in Using Multiple Tests</b>	155
	<i>Introduction</i>	155
	<i>Tests in Parallel or Tests in Sequence</i>	156
	<i>The Quality of Combined Tests</i>	161
	<i>When Should a Test Be Repeated?</i>	163
	<i>Should a Test Be Done Blind to Previous Test Results?</i>	163
	<i>Summary</i>	164
11.	<b>Evaluating Batteries of Medical Tests: Optimal Sequences</b>	165
	<i>Introduction</i>	165
	<i>The Basic Model</i>	166
	<i>Compiling the Raw Data Base</i>	170
	<i>Compiling the Working Data Base</i>	176
	<i>Selection of the Optimal First Test and Reiteration</i>	179
	<i>Increasing the Yield</i>	180
	<i>Testing the Quality Index Using a Statistical Jackknife</i>	185
	<i>Completing the Search for the Optimal Battery Structure</i>	187
	<i>The Final Recommendation for Battery Structure</i>	188
	<i>An Alternative Strategy</i>	193
	<i>Summary</i>	194
12.	<b>Evaluating Batteries of Medical Tests: Optimal Scores</b>	199
	<i>Introduction</i>	199
	<i>The Basic Model</i>	204
	<i>Multiple Linear Discrimination</i>	209
	<i>Logistic Regression Analysis</i>	216
	<i>Some Practical Considerations in Choosing and Using Models</i>	222
	<i>Summary</i>	227
13.	<b>Evaluating Batteries of Prognostic Tests with Variable Follow-Up Times</b>	228
	<i>Introduction</i>	228
	<i>The Basic Model</i>	231
	<i>Ignoring the Problem of Variable Follow-Up: A Possibly Invalid Approach</i>	234
	<i>Counting Time and Not Patients: Another Possibly Invalid Procedure</i>	234
	<i>Tailoring the Data to Fit the Solution: An Invalid and Inefficient Approach</i>	239
	<i>Estimating Survival Curves with Variable Follow-Up Times: The Kaplan-Meier Method</i>	241

	<i>Some General Observations on the Choice of Follow-Up Time and Its Impact on Test Evaluation</i>	247
	<i>Evaluating a Family of Prognostic Tests with a Variable Follow-Up Time</i>	250
	<i>An Alternative? Randomized Clinical Trials for the Evaluation of Screening Programs (Test + Intervention) for Prevention</i>	256
	<i>Evaluating a Battery of Prognostic Tests with Variable Follow-Up Periods: Sequential Structuring</i>	259
	<i>Evaluating a Battery of Prognostic Tests Using Variable Follow-Up: Scoring a Battery with the Proportional Hazards Model</i>	259
	<i>Summary</i>	263
14.	<b>Evaluation of Medical Tests: The Past, Present, and Future</b>	266
	<i>Introduction: The Past and the Present</i>	266
	<i>The Future: Unsolved Problems</i>	272
	<i>Problem: Evaluation of Monitoring Tests</i>	272
	<i>Problem: Multicategory Responses</i>	273
	<i>Problem: Stopping Rules</i>	274
	<i>Other Problems</i>	275
	<i>The Present: An Overall Summary</i>	276
	<i>What Is the Disorder? What Diagnosis Is to Be Used? Is the Diagnosis Clinically Valid and Reliable?</i>	276
	<i>In What Clinical Population or Populations Is the Test Proposed for Use? How Can the Sampling Be Done so as to Obtain a Representative Sample from That Population or Populations?</i>	277
	<i>If the Test Is to Be Evaluated as a Prognostic Test, Will the Follow-Up Be Fixed or Variable? If the Follow-Up Is to Be Fixed, Fixed at What Time?</i>	277
	<i>Will Sampling Be Naturalistic, Retrospective, or Prospective?</i>	277
	<i>How Large a Sample Is Needed?</i>	278
	<i>How Are Dropouts and Missing Responses Avoided?</i>	278
	<i>What Are the Clinical Benefits in this Situation?</i>	279
	<i>What Are the Tests Under Evaluation—Their Protocols, Responses, and Referents? What Are the Test Costs? If This is a Single Test (a Single Fixed Cost for All the Responses) or Is This to Be Considered a Battery of Tests (Separable Costs)?</i>	279
	<i>How Are the Blinding of the Diagnosis and Test Results Assured? If There Are Multiple Tests in a Battery Under Evaluation, Are These Blinded to Each Other?</i>	280
	<i>Is There any Internal or External Standard Provided for the Performance of an Excellent Test?</i>	280
	<i>Was the Data Base Properly Compiled and Thoroughly Checked for Errors?</i>	281
	<i>For Each Single Test Under Evaluation, for Each Test in the Battery Under Evaluation, and for Each Population Under Evaluation, Have the Descriptive Statistics Been Properly Computed?</i>	282

*What Is the Quality of Each Single Test Under Evaluation, and for Each Population Under Evaluation (and in an Evaluation of a Prognostic Test with Variable Follow-Up Time, for Each Possible Value of Follow-Up Time)?* 282

*Has the Optimal Referent for a Test, or the Optimal First Test in a Battery of Tests Been Appropriately Selected?* 284

*If There Is a Battery of Tests Under Consideration, Has the Battery Been Appropriately Evaluated?* 285

References 286

Index 291

About the Author 295