Statistical Evaluation of Diagnostic Performance
Topics in ROC Analysis

Highlights:
• Methods for statistical validations of diagnostic accuracy using ROC analysis
• Methods for estimation and comparison of diagnostic test characteristics
• Monotone transformation methods, binormality testing, and goodness-of-fit issues
• Bayesian hierarchical models for estimating diagnostic accuracy
• Multireader ROC analysis and FROC analysis
• Biomarkers, sequential designs, and bioinformatics

The book is suitable for graduate-level students and researchers in statistics, biostatistics, epidemiology, public health, biomedical engineering, radiology, medical imaging, biomedical informatics, and other closely related fields. Additionally, clinical researchers and practicing statisticians in academia, industry, and government could benefit from the presentation of such important and yet frequently overlooked topics.
Statistical Evaluation of Diagnostic Performance
Topics in ROC Analysis
Adaptive Design Theory and Implementation Using SAS and R
Mark Chang

Advanced Bayesian Methods for Medical Test Accuracy
Lyle D. Broemeling

Advances in Clinical Trial Biostatistics
Nancy L. Geller

Applied Statistical Design for the Researcher
Daryl S. Paulson

James E. De Muth

Bayesian Adaptive Methods for Clinical Trials
Scott M. Berry, Bradley P. Carlin, J. Jack Lee, and Peter Muller

Bayesian Analysis Made Simple: An Excel GUI for WinBUGS
Phil Woodward

Bayesian Methods for Measures of Agreement
Lyle D. Broemeling

Bayesian Missing Data Problems: EM, Data Augmentation and Noniterative Computation
Ming T. Tan, Guo-Liang Tian, and Kai Wang Ng

Bayesian Modeling in Bioinformatics
Dipak K. Dey, Samiran Ghosh, and Bani K. Mallick

Causal Analysis in Biomedicine and Epidemiology: Based on Minimal Sufficient Causation
Mikel Aickin

Clinical Trial Data Analysis using R
Ding-Geng (Din) Chen and Karl E. Peace

Clinical Trial Methodology
Karl E. Peace and Ding-Geng (Din) Chen

Computational Methods in Biomedical Research
Ravindra Khattree and Dayanand N. Naik

Computational Pharmacokinetics
Anders Källén

Controversial Statistical Issues in Clinical Trials
Shein-Chung Chow

Data and Safety Monitoring Committees in Clinical Trials
Jay Herson

Design and Analysis of Animal Studies in Pharmaceutical Development
Shein-Chung Chow and Jen-pei Liu

Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition
Shein-Chung Chow and Jen-pei Liu

Design and Analysis of Clinical Trials with Time-to-Event Endpoints
Karl E. Peace

Design and Analysis of Non-Inferiority Trials
Mark D. Rothmann, Brian L. Wiens, and Ivan S. F. Chan

Difference Equations with Public Health Applications
Lemuel A. Moyé and Asha Seth Kapadia

DNA Methylation Microarrays: Experimental Design and Statistical Analysis
Sun-Chong Wang and Arturas Petronis

DNA Microarrays and Related Genomics Techniques: Design, Analysis, and Interpretation of Experiments
David B. Allison, Grier P. Page, T. Mark Beasley, and Jode W. Edwards

Dose Finding by the Continual Reassessment Method
Ying Kuen Cheung

Elementary Bayesian Biostatistics
Lemuel A. Moyé
Frailty Models in Survival Analysis
Andreas Wienke

Generalized Linear Models: A Bayesian Perspective
Dipak K. Dey, Sujit K. Ghosh, and Bani K. Mallick

Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries
Daryl S. Paulson

Measures of Interobserver Agreement and Reliability, Second Edition
Mohamed M. Shoukri

Medical Biostatistics, Second Edition
A. Indrayan

Meta-Analysis in Medicine and Health Policy
Dalene Stangl and Donal A. Berry

Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies
Mark Chang

Multiple Testing Problems in Pharmaceutical Statistics
Alex Dmitrienko, Ajit C. Tamhane, and Frank Bretz

Sample Size Calculations in Clinical Research, Second Edition
Shein-Chung Chow, Jun Shao and Hansheng Wang

Statistical Design and Analysis of Stability Studies
Shein-Chung Chow

Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis
Kelly H. Zou, Aiyi Liu, Andriy I. Bandos, Lucila Ohno-Machado, and Howard E. Rockette

Statistical Methods for Clinical Trials
Mark X. Norleans

Statistics in Drug Research: Methodologies and Recent Developments
Shein-Chung Chow and Jun Shao

Statistics in the Pharmaceutical Industry, Third Edition
Ralph Buncher and Jia-Yeong Tsay

Translational Medicine: Strategies and Statistical Methods
Dennis Cosmatos and Shein-Chung Chow
Statistical Evaluation of Diagnostic Performance
Topics in ROC Analysis

Kelly H. Zou
Aiyi Liu
Andriy I. Bandos
Lucila Ohno-Machado
Howard E. Rockette
In memory of Professor Harry Samuel Wicand of the
University of Pittsburgh who passed away at the age of 62
due to a recurrence of non-Hodgkin's lymphoma.
Contents

Authors ........................................................................................................................................ xiii
Preface ......................................................................................................................................... xvii

Section I  Introduction

1. Background and Introduction ................................................................................................. 3
  1.1 Background Information ...................................................................................................... 3
  1.2 Gold Standard, Decision Threshold, Sensitivity, and Specificity ....................................... 4
  1.3 Kappa Statistics .................................................................................................................... 5
  1.4 Receiver Operating Characteristic Curve ............................................................................. 6
  1.5 Area and Partial Area under ROC Curve .............................................................................. 10
  1.6 Confidence Intervals, Regions, and Bands .......................................................................... 11
  1.7 Point of Intersection and Youden Index ............................................................................. 12
  1.8 Comparison of Two or More ROC Curves .......................................................................... 13
  1.9 Approaches to ROC Analysis ............................................................................................ 14
References .................................................................................................................................... 15

Section II  Methods for Univariate and Multivariate Data

2. Diagnostic Rating Scales ......................................................................................................... 21
  2.1 Introduction .......................................................................................................................... 21
  2.1.1 Frequent Need for Finer Scale ....................................................................................... 22
  2.2 Interpreter-Free Diagnostic Systems .................................................................................. 23
  2.3 Human Interpreter as Integral Part of Diagnostic System ................................................ 24
  2.4 Remarks and Further Reading ........................................................................................... 25
References .................................................................................................................................... 26

3. Monotone Transformation Models ......................................................................................... 29
  3.1 Introduction .......................................................................................................................... 29
  3.2 General Assumptions .......................................................................................................... 29
  3.3 Empirical Methods .............................................................................................................. 30
  3.4 Nonparametric Kernel Smoothing ..................................................................................... 31
3.5 Parametric Models and Monotone Transformations to Binormal Distributions ................................................. 34
3.6 Confidence Intervals ................................................................. 41
  3.6.1 Sensitivity or Specificity at Specific Threshold ............ 41
  3.6.2 Confidence Intervals for Sensitivity (Specificity) at Given Specificity (Specificity) .............................................. 46
  3.6.3 Confidence Intervals for Area under Curve (AUC) .......... 47
3.7 Concordance Measures in Presence of Monotone Transformations ................................................................. 51
3.8 Intraclass Correlation Coefficient ............................................. 57
3.9 Remarks and Further Reading .................................................. 59
References ................................................................................. 59

4. Combination and Pooling of Biomarkers ........................................ 65
  4.1 Introduction ........................................................................... 65
  4.2 Combining Biomarkers to Improve Diagnostic Accuracy ...... 66
    4.2.1 Likelihood Ratio Approach ............................................... 66
    4.2.2 Uniformly Optimal Linear Combinations ....................... 67
    4.2.3 Normal Linear Combinations Maximizing AUC ............. 68
    4.2.4 Distribution-Free Approach for Linear Combinations ......... 70
  4.3 ROC Curve Analysis with Pooled Samples ......................... 72
    4.3.1 Normal Distributions ...................................................... 73
    4.3.2 Gamma Distributions ..................................................... 75
    4.3.3 Stable Distributions ....................................................... 77
    4.3.4 Distribution-Free Approach .......................................... 78
  4.4 Remarks and Further Reading ................................................ 79
    4.4.1 Combination of Biomarkers ............................................. 79
    4.4.2 Pooling of Biomarkers .................................................... 80
References .................................................................................. 80

5. Bayesian ROC Methods ............................................................. 83
  5.1 Introduction ........................................................................... 83
  5.2 Methods for Sensitivity, Specificity, and Prevalence ............. 84
  5.3 Clustered Data Structures and Hierarchical Methods ............ 89
  5.4 Assumptions and Models for ROC Analysis ......................... 92
  5.5 Normality Transformation ...................................................... 93
  5.6 Elicitation of Prior Information .............................................. 95
  5.7 Estimation of ROC Parameters and Characteristics .............. 96
  5.9 Remarks and Further Reading ................................................ 96
References .................................................................................. 98
Section III  Advanced Approaches and Applications

6. Sequential Designs of ROC Experiments ................................................................. 103
   6.1 Introduction ........................................................................................................ 103
   6.2 Group Sequential Tests Using Large Sample Theory ..................................... 104
       6.2.1 Approximating Test Statistics via Brownian Motion ......................... 104
       6.2.2 Group Sequential Testing Procedures .................................................. 106
       6.2.3 Choosing Stopping Boundaries ............................................................. 106
   6.3 Sequential Evaluation of Single ROC Curve .................................................... 110
       6.3.1 Binormal Model ....................................................................................... 110
       6.3.2 Nonparametric Model ......................................................................... 112
   6.4 Sequential Comparison of Two ROC Curves ................................................... 114
       6.4.1 Binormal Model ....................................................................................... 115
       6.4.2 Nonparametric Model ......................................................................... 116
   6.5 Sequential Evaluation of Binary Outcomes ...................................................... 118
       6.5.1 Evaluating Accuracy of Diagnostic Tests .............................................. 119
       6.5.2 Comparison of Two Diagnostic Tests ................................................... 121
   6.6 Sample Size Estimation ..................................................................................... 123
   6.7 Remarks and Further Reading ......................................................................... 124
       6.7.1 Inference upon Termination ................................................................. 124
       6.7.2 Sample Size Re-Estimation ................................................................... 125
       6.7.3 Computing Software ............................................................................. 125

References .................................................................................................................. 125

7. Multireader ROC Analysis .................................................................................... 129
   7.1 Introduction ........................................................................................................ 129
   7.2 Overall ROC Curve and Its AUC .................................................................... 130
   7.3 Statistical Analysis of Cross-Correlated Multireader Data ............................. 132
       7.3.1 Covariate-Free Approaches ................................................................... 133
           7.3.1.1 Methods for Fixed Reader Inferences ........................................... 133
           7.3.1.2 Methods for Random Reader or Fixed Reader Inferences .......... 136
           7.3.1.3 Variance Components Estimation .............................................. 141
       7.3.2 Covariate Adjustment Approaches ............................................................ 144
           7.3.2.1 Models for ROC Curves .............................................................. 145
           7.3.2.2 Models for Summary Indices ....................................................... 147
           7.3.2.3 Bayesian Approaches ................................................................. 148
   7.4 Remarks and Further Reading ......................................................................... 150

References .................................................................................................................. 152

Appendix 7.A: Closed Form Formulation of DBM Approach for Comparing
   Two Modalities Using Empirical AUC ................................................................. 155
Appendix 7.B: Variance Estimators of Empirical AUCs ......................................... 158
8. Free-Response ROC Analysis .............................................................. 163
  8.1 Introduction .................................................................................. 163
  8.2 FROC Approach .......................................................................... 165
    8.2.1 Rating-Free FROC Analysis .................................................... 165
    8.2.2 FROC Space and Use of Guessing Process for Interpretation and Comparison .............................................................................................................. 168
    8.2.3 FROC Curve .............................................................................. 171
    8.2.4 Rating-Based FROC Analysis .................................................... 172
    8.2.5 Estimation of FROC Curve ....................................................... 175
    8.2.6 Summary Indices ....................................................................... 179
    8.2.7 Comparison of FROC Performance Levels ................................. 184
  8.3 Other Approaches of Detection–Localization Performance Assessment .................................................................................................................. 187
    8.3.1 Alternative FROC (AFROC) ..................................................... 187
    8.3.2 Localization ROC (LROC) ........................................................ 188
    8.3.3 Regions of Interest (ROI) Approach .......................................... 191
  8.4 Remarks and Further Reading ....................................................... 196
References ............................................................................................. 197

9. Machine Learning and Predictive Modeling ........................................ 201
  9.1 Introduction .................................................................................. 201
  9.2 Predictive Modeling ...................................................................... 202
  9.3 Cross-Validation .......................................................................... 203
  9.4 Bootstrap Resampling Methods ..................................................... 204
    9.4.1 Bootstrap Method 1 ................................................................... 205
    9.4.2 Bootstrap Method 2 ................................................................... 205
  9.5 Overfitting and False Discovery Rate ............................................. 206
  9.6 Remarks and Further Reading ....................................................... 207
References ............................................................................................. 208

Section IV Discussions and Extensions

10. Summary and Challenges ............................................................... 213
  10.1 Summary and Discussion ............................................................ 213
  10.2 Future Directions in ROC Analysis .............................................. 213
  10.3 Future Directions in Reliability Analysis ....................................... 215
  10.4 Final Remarks .............................................................................. 217
Appendix: Notation List ........................................................................... 219
Authors

Kelly H. Zou, PhD is a director of statistics in the Specialty Care Business Unit at Pfizer, Inc. A native of Shanghai, China, she earned a BA (summa cum laude) in mathematics from Chaminade University of Honolulu, a master’s and PhD in statistics from the University of Rochester, and completed a postdoctoral fellowship in biostatistics and radiology from Harvard Medical School and its affiliated Brigham and Women’s Hospital. She has been a lecturer on health care policy at Harvard Medical School; an instructor, assistant professor, and associate professor in radiology at Harvard Medical School; the principal statistician in radiology at Brigham and Women’s Hospital; and the director of biostatistics at Children’s Hospital Boston. In addition to her career in academia, Dr. Zou served as an associate director of rates at Barclays Capital and director of statistics in outcomes research and evidence-based strategies at Pfizer, Inc.

Dr. Zou served as the principal investigator on a number of research grants funded by the National Library of Medicine, National Institute of General Medical Sciences, and the Agency for Healthcare Research and Quality of the U.S. National Institutes of Health. Her work on ROC analysis and statistical classification led her to be the recipient of the Stauffer Award for the best article published in Academic Radiology, the first place winner of the American Statistical Association and Biopharmaceutical Statistics Section poster competition at the 2009 Joint Statistical Meetings, and second and third place winner at the 2010 meeting. She was the recipient of the Travel Stipend Award from the Society of Health Services Research in Radiology and received the Reviewer with Special Distinction Award from Radiology.

Dr. Zou has published more than 100 peer-reviewed articles listed via PubMed. She served as an associate editor of Statistics in Medicine, Radiology, and Medical Physics, and as a referee for more than 10 professional statistical and medical journals. She served as vice chair of the Committee on Applied Statistics American Statistical Association; member of the Corporate Sponsorship Committee, Biopharmaceutical Section, American Statistical Association; chair of Judiciary Committee, Radiology Alliance to Health Services Research, Association of University Radiologists; and member of the Faculty Taskforce, Joint Committee on the Status of Women, Harvard Medical School and Harvard School of Dental Medicine. She was the theme editor on Mathematical and Statistical Methods for Diagnoses and Therapies. Her research interests include diagnostic tests, medical imaging, health policy, outcomes research, and clinical trials.
Aiyi Liu, PhD is a senior investigator in the Biostatistics and Bioinformatics Branch, Division of Epidemiology, Statistics and Prevention Research, of the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health (NIH). He earned a BA in mathematics and a master’s degree in statistics from the University of Science and Technology of China, and subsequently taught for 5 years in the Department of Mathematics at the same university. He earned a PhD in statistics from the University of Rochester, and completed his postdoctoral training at St. Jude Children’s Research Hospital in Memphis, Tennessee. Prior to his appointment at NIH, Dr. Liu was an assistant professor of biostatistics in the Department of Biomathematics and Biostatistics, Georgetown University Lombardi Cancer Center, Washington, DC, and taught in the department’s master’s program and the center’s tumor biology program.

Dr. Liu’s research interests include general statistical theory and methods, sequential methodology, multivariate data analysis, diagnostic biomarkers, and ROC curve analysis. He has published more than 80 articles in peer-reviewed journals including Biometrics, Biometrika, and the Journal of the American Statistical Association. He served as a referee for more than 15 journals and organized a number of invited sessions for statistical meetings. Dr. Liu is an associate editor for the Journal of Statistical Planning and Inference, and acted as a guest co-editor for Philosophical Transaction of the Royal Society A. He is an elected member of the International Statistical Institute, and a member of the American Statistical Association, International Chinese Statistical Association, and Institute of Mathematical Statistics. His tenure as a member of the Committee on Award of an Outstanding Statistical Application of the American Statistical Association started in 2011.

Andriy I. Bandos, PhD is a research assistant professor in the Department of Biostatistics at the University of Pittsburgh. He earned a master’s in mathematics from Karazin Kharkiv National University (Ukraine) and a PhD in biostatistics from the University of Pittsburgh. Over the past several years, he has worked extensively on methodology development and analysis of observer performance studies in diagnostic imaging. His work has been published in a number of biostatistical and radiology-related journals.

Dr. Bandos developed and teaches a graduate-level course on ROC analysis at the University of Pittsburgh. He is a reviewer for a number of academic journals and governmental agencies. His current research interests include statistical evaluation of diagnostic performance, ROC analysis, free-response ROC (FROC) methodology, nonparametric methods, and resampling approaches in statistics.

Lucila Ohno-Machado, MD, PhD is a professor of medicine and the founding chief of the Division of Biomedical Informatics at the University of California San Diego, (UCSD). She earned her medical degree from the
University of Sao Paulo, Brazil and her doctorate in medical information sciences from Stanford University. Before joining UCSD, she was the director of the training program of the Harvard–MIT–Tufts–Boston University consortium in Boston, and the director of the Decision Systems Group at Brigham and Women’s Hospital, Harvard Medical School.

Dr. Ohno-Machado’s research focuses on the development of new evaluation methods for predictive models of disease, with special emphasis on the analysis of model calibration and implications in healthcare. She is an elected member of the American College of Medical Informatics, the American Institute for Medical and Biological Engineering, and the American Society for Clinical Investigation. She is the editor-in-chief of the Journal of the American Medical Informatics Association (JAMIA).

Howard E. Rockette, PhD, is a former chair and currently a professor of biostatistics at the University of Pittsburgh Graduate School of Public Health. He is a fellow of the American Statistical Association and a member of the Society for Epidemiological Research and the Society of Clinical Trials. He has written 150 peer-reviewed articles, more than a third of which relate to the development of statistical methodology or its application to the evaluation of diagnostic imaging systems. He served on numerous advisory committees for government and industry, and on the editorial board of Annals of Epidemiology and Academic Radiology.
Preface

We now extend the body of literature via the journey of “ROC Trek” into where classifiers have never gone before.

Matt Gönen

ROC on

Statistical evaluation of diagnostic performance in general, and receiver operating characteristic (ROC) analysis in particular, are important for assessing the performance of medical tests and statistical classifiers, and also for evaluating predictive models or algorithms. In diagnostic and prognostic tasks, the receiver operating characteristic (ROC) curve plays an essential role by providing a graphic display that illustrates the discrimination performance of a medical diagnostic test. ROC analysis originated during World War II as a method of assessing classification accuracy for differentiating signals from noise in radar detection. Recently, the ROC methodology has been adapted to many disciplines, including several clinical areas that depend heavily on screening and diagnostic tests.

The purpose of this book is to present innovative approaches in ROC analysis that are relevant to a wide variety of clinical applications including medical imaging, cancer research, epidemiology, and bioinformatics. We begin by reviewing the conventional ROC methodology. Monotone transformation models taking data to parametric forms are chosen to improve goodness of fit of modeling approaches. Likelihood-based algorithms for estimating an ROC curve are discussed, along with the associated characteristics of several models for univariate and multivariate data. The covered topics include monotone transformation techniques in parametric ROC analysis, ROC methods for combined and pooled biomarkers, Bayesian hierarchical transformation models, sequential designs and inferences in the ROC setting, predictive modeling, multireader ROC analysis, and free-response ROC (FROC) methodology. The topics highlighted include:

- Methods for statistical validations of diagnostic accuracy using ROC analysis
- Methods for estimation and comparison of diagnostic test characteristics
- Monotone transformation methods, binormality testing, and goodness-of-fit issues
- Bayesian hierarchical models for estimating diagnostic accuracy
• Multireader and multimodality ROC analysis and FROC analysis
• Biomarkers, sequential designs, and bioinformatics

The intended audience of this book includes graduate-level students and researchers in the areas of statistics, biostatistics, epidemiology, public health, biomedical engineering, radiology, medical imaging, biomedical informatics, and other closely-related fields. Additionally, clinical researchers and practicing statisticians in academia, industry, and government may benefit from the presentation of such important and yet frequently overlooked topics.

The CA19-9 and CA125 cancer marker data were provided by Professor H. Samuel Wieand, of the University of Pittsburgh, who passed away on June 10, 2006 from a recurrence of non-Hodgkin’s lymphoma. This book is in memory of Professor Wieand for his important contribution to the field of ROC analysis following his classical work examining the nonparametric statistics for comparing diagnostic markers with paired or unpaired data.

The authors wish to express our deepest gratitude to Rob Calver, the senior statistics acquisitions editor at Taylor & Francis Group, who tirelessly provided tremendous assistance in making this project a reality. We are most grateful to two anonymous experts who reviewed earlier chapters. We especially thank the many authors who have contributed extensively to the body of ROC literature including widely used software programs.

The authors were partially funded by Grants R03-HS13234, R01-LM007861, R01-GM074068, R01-EB002106, R01-EB006388, R01-LM009520, and U54-HL108460 from the U.S. National Institutes of Health (NIH). Dr. Aiyi Liu’s research was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the NIH.

By gaining an in-depth understanding and accurate validation of complex biomarkers and high-dimensional modality data, we hope that appropriate therapies and improved outcomes may ultimately be achieved using methods described in this book.

ROC ROCKS!

Section I

Introduction