Bayesian Biostatistics and Diagnostic Medicine

Lyle D. Broemeling
Bayesian Biostatistics
and Diagnostic Medicine
Bayesian Biostatistics
and Diagnostic Medicine

Lyle D. Broemeling
The University of Texas
MD Anderson Cancer Center
Houston, Texas, U.S.A.
# Table of Contents

Preface......................................................................................................... ix  
Acknowledgments................................................................................... xi  
Author..................................................................................................... xiii  

1 Introduction............................................................................................. 1  
1.1 Introduction.......................................................................................... 1  
1.2 Statistical Methods in Diagnostic Medicine........................................ 1  
1.3 Preview of Book................................................................................... 2  
1.4 Datasets for the Book.......................................................................... 4  
1.5 Software............................................................................................... 4  
References................................................................................................. 5  

2 Diagnostic Medicine........................................................................... 7  
2.1 Introduction.......................................................................................... 7  
2.2 Imaging Modalities............................................................................. 7  
2.3 Activities in Diagnostic Imaging....................................................... 11  
2.4 Accuracy and Agreement................................................................... 12  
2.5 Developmental Trials for Imaging.................................................... 14  
2.6 Protocol Review and Clinical Trials................................................ 15  
  2.6.1 The Protocol............................................................................... 16  
  2.6.2 Phase I, II, and III Clinical Designs.......................................... 16  
2.7 The Literature..................................................................................... 18  
References................................................................................................. 18  

3 Other Diagnostic Procedures.............................................................. 21  
3.1 Introduction......................................................................................... 21  
3.2 Sentinel Lymph Node Biopsy for Melanoma...................................... 21  
3.3 Tumor Depth to Diagnose Metastatic Melanoma.............................. 22  
3.4 Biopsy for Nonsmall Cell Lung Cancer............................................ 23  
3.5 Coronary Artery Disease.................................................................... 24  
References................................................................................................. 24  

4 Bayesian Statistics.............................................................................. 27  
4.1 Introduction......................................................................................... 27  
4.2 Bayes Theorem.................................................................................. 28  
4.3 Prior Information............................................................................... 29  
4.4 Posterior Information....................................................................... 33  
4.5 Inference............................................................................................. 36  
  4.5.1 Introduction............................................................................... 36  
  4.5.2 Estimation................................................................................... 36  
  4.5.3 Testing Hypotheses................................................................. 38
8  Diagnostic Imaging and Clinical Trials ......................................... 141
   8.1 Introduction ............................................................................... 141
   8.2 Clinical Trials ........................................................................... 142
      8.2.1 Introduction ......................................................................... 142
      8.2.2 Phase I Designs ................................................................. 142
      8.2.3 Phase II Trials ................................................................. 143
      8.2.4 Phase III Trials ................................................................ 145
   8.3 Protocol .................................................................................... 145
   8.4 Guidelines for Tumor Response ................................................. 146
   8.5 Bayesian Sequential Stopping Rules ......................................... 148
   8.6 Software for Clinical Trials ...................................................... 152
      8.6.1 CRM Simulator for Phase I Trials ..................................... 153
      8.6.2 Multc Lean for Phase II Trials .......................................... 153
   8.7 Examples .................................................................................. 154
      8.7.1 Phase I Trial for Renal Cell Carcinoma ......................... 154
      8.7.2 An Ideal Phase II Trial ....................................................... 156
      8.7.3 Phase II Trial for Advanced Melanoma ......................... 158
   8.8 Exercises .................................................................................. 162
   References ..................................................................................... 163

9  Other Topics .................................................................................. 165
   9.1 Introduction ............................................................................... 165
   9.2 Imperfect Diagnostic Test Procedures .................................... 166
      9.2.1 Extreme Verification Bias ............................................... 166
      9.2.2 Verification Bias ............................................................... 170
      9.2.3 Estimating Test Accuracy with No Gold Standard ............ 173
   9.3 Test Accuracy and Survival Analysis .................................... 178
   9.4 ROC Curves with a Non-Binary Gold Standard ................. 180
   9.5 Periodic Screening in Cancer .................................................. 182
      9.5.1 Inference for Sensitivity and Transition Probability ........ 182
      9.5.2 Bayesian Inference for Lead-Time .................................. 186
   9.6 Decision Theory and Diagnostic Accuracy .......................... 189
   9.7 Exercises ............................................................................... 192
   References ..................................................................................... 193

Index ............................................................................................... 195
Bayesian methods are being used more often than ever before in biology and medicine. For example, at the University of Texas MD Anderson Cancer Center, Bayesian sequential stopping rules routinely are used for the design of clinical trials. This book is based on the author’s experience working with a variety of researchers, including radiologists, pathologists, and medical oncologists. The majority of that experience has been with the Division of Diagnostic Imaging, where radiologists determine the extent of disease among patients undergoing treatment. Diagnosis, via medical imaging, is essential in order to assess the effect of the various therapies provided to the patient. Another source of information for the author has been the ability to work with medical oncologists in their design of Phase I, II, and III clinical trials. The author has found Bayesian methods for the design and analysis of clinical trials to be quite useful because prior information, in the form of previous related studies, is always available and easily incorporated into the design of future studies.

Based on this experience and the wealth of information available to the author, this book should give the biostatistics student a good idea of what to expect and how to work with healthcare researchers. It is an introductory book with a Bayesian flavor and is directed toward diagnostic medicine. Students with a good background in the basic methods courses of regression and the analysis of variance and in the introductory courses in probability and mathematical statistics should benefit greatly from the book. With this type of background, the student will be able to learn Bayesian statistics and how to apply it to important problems in medicine and biology. In addition, it should serve as a useful reference for those providing statistical assistance to medical scientists.

In the book, the reader is introduced to various diagnostic medical procedures, then presented with the fundamentals of Bayesian statistics and associated computing methods. Next, the foundation for the analysis of diagnostic test accuracy is outlined and the Bayesian way to analyze such data is explained, using many author-assisted studies. Of special interest is the estimation of the area under the receiver operating characteristic (ROC) curve for determining diagnostic accuracy. Also described in the book is a novel way to estimate the area when the image data are clustered.

Some of the material in this book is similar to that found in Statistical Methods in Diagnostic Medicine by Zhou, Obuchowski, and McClish and The Statistical Evaluation of Medical Tests for Classification and Prediction by Pepe. Several examples from these sources are analyzed from a Bayesian perspective. However, this book is entirely from a Bayesian perspective and presents
a great deal of material not stressed in the above-mentioned references. This material includes Bayesian methods of agreement between readers and the role diagnostic medicine plays in the design of clinical trials, and should complement as well as expand on the books by Pepe and by Zhou et al. A unique feature of this book is that the Minitab® and WinBUGS® packages are employed to provide Bayesian inferences. After reading the book, the student will be able to provide a Bayesian analysis for a large variety of interesting and practical problems.
Acknowledgments

The author would like to acknowledge the many departments and people at the University of Texas MD Anderson Cancer Center in Houston who assisted me during the writing of this book. First and foremost, I would like to thank the faculty in the Division of Diagnostic Imaging, who shared their valuable research expertise with me. Many of the examples of this book are a result of consulting with the radiologists in that division. In particular, Drs. Isis Gayed, Reggie Munden, Vikas Kundra, Edith Marom, and Sanjay Gupta are gratefully acknowledged.

Also, I would like to thank many people in the Department of Biostatistics and Applied Mathematics for the opportunity to be part of a developing biostatistics program. The assistance of the department in the areas of protocol development, the editorial services of Margaret Newell, and the programming assistance for Bayesian statistical methods were invaluable.

The National Cancer Institute awarded (1 P50 CA09345901 A1(PC-C)) Melanoma Spore to the UT MD Anderson Cancer Center (Dr. Elizabeth Grimm, principal investigator), and this provided support during the past 3 years.

Lastly, I want to give special thanks to my wife, Ana L. Broemeling, who always encouraged me during the many months of writing this book.
Author

Lyle D. Broemeling, Ph.D., is a research professor in the Department of Biostatistics and Applied Mathematics at the University of Texas MD Anderson Cancer Center in Houston. He received his B.A., M.S., and Ph.D. degrees from Texas A&M University at College Station. Since 1987, he has been active in consulting with medical researchers at the University of Texas Medical Branch in Galveston, the University of Texas School of Public Health, and the University of Texas MD Anderson Cancer Center in Houston. Dr. Broemeling’s main interest is in developing Bayesian statistical methods and his previous books include Bayesian Analysis of Linear Models and Econometrics and Structural Change, written with Dr. Hiroki Tsurumi, Professor of Economics, Rutgers University.
Chapter 1

Introduction

1.1 Introduction

This book is about how to use Bayesian statistical methods to design and analyze studies involving diagnostic medicine. It grew out of the author’s experience in consulting with many investigators of the Division of Diagnostic Imaging at The University of Texas MD Anderson Cancer Center (MDACC) in Houston. In a modern medical center, diagnostic imaging is ubiquitous and crucial for patient management, from the initial diagnosis to assessing the extent of disease as the patient is being treated.

1.2 Statistical Methods in Diagnostic Medicine

Biostatistics plays a pivotal role in the imaging literature, as can be discerned by reading articles in the mainline journals, such as Academic Radiology, The American Journal of Roentgenology, and Radiology, and the more specialized, such as The Journal of Computed Assisted Tomography, The Journal of Magnetic Resonance Imaging, The Journal of Nuclear Medicine, and Ultrasound in Medicine. As we will see, the usual methods ranging from the t-test and chi-squared test to others, such as the analysis of variance and various regression techniques, are standard fare for medical diagnostic studies. There are also some methods that are somewhat unique to the field, including ways to estimate diagnostic test accuracy and methods to measure the agreement between imaging modalities and/or readers. Also important to imaging research are the elements of good statistical design, including replication, randomization, and blocking in the planning of clinical investigations.

Diagnostic imaging is employed in clinical trials, such as Phase II trials, where the main objective is to determine the response to a new therapy and where the response is based on an image measurement. Patient sample size, based on Bayesian sequential stopping rules, is another application that has proven to be quite beneficial in the development of new medical therapies.
The Bayesian approach will be used throughout this book for all aspects of the design and analysis of diagnostic studies. Also, standard non-Bayesian methods will be employed on certain occasions when deemed appropriate.

1.3 Preview of Book

Chapter 2 is an introduction to the fundamental areas of interest in diagnostic medicine, including a brief description of the main imaging modalities, namely X-ray, computed tomography, magnetic resonance imaging, and nuclear medicine procedures. A brief explanation of how diagnostic imaging is involved in large population screening and in the day-to-day patient management is provided. The estimation of diagnostic accuracy is not only essential to the diagnosis of a patient, but also for the assessment of patient progress during therapy. Estimating agreement between various readers and/or modalities is crucial for the training of radiologists, for comparing imaging techniques, and in assessing the success of therapy. The development of an imaging modality involves three phases (Phase I, Phase II, and Phase III developmental trials) and these will be explained along with numerous examples. A description of the role diagnostic imaging plays in the various phases of a clinical trail is provided. In addition, the literature of the field, including the various books and journals, is reviewed.

The purpose of Chapter 3 is to introduce the reader to other areas of diagnostic medicine. In addition to diagnostic imaging, there are many other services that provide diagnostic information, including pathology and surgery. For example, in the treatment of breast cancer, it is imperative to know the extent of metastasis to the axillary lymph nodes. Sentinel lymph node (SLN) biopsies are performed to determine if there is lymph node involvement, and diagnostic imaging (including nuclear medicine and interventional radiology), pathology, and surgery are key components for the procedure. This chapter will examine the role of SLN biopsy in lung cancer and in melanoma. Of course, the biostatistical methods for determining accuracy and agreement are the same as those for diagnostic imaging, and their application to nonimaging tests will be shown.

Chapter 4 provides an introduction to Bayesian statistics, beginning with Bayes theorem and how prior information is found and used. Inferential techniques of estimation and testing hypotheses based on the posterior and predictive distributions are introduced. Also, how they are to be applied in diagnostic imaging is revealed. The computing algorithms along with the corresponding software for direct and indirect sampling from posterior distributions are briefly outlined. This book is unique in that the sample sizes are to be determined by fully Bayesian techniques, and the software that is used to estimate the sample size is described. Some of the packages are off-the-shelf, while others have been developed at MDACC.
Chapter 5 introduces the estimation of accuracy by sensitivity, specificity, and positive and negative predictive values for ordinal and continuous diagnostic measurements. Numerous examples taken from the literature illustrate the various concepts involved in test accuracy. The area under the receiver operating characteristic (ROC) curve gives the overall intrinsic accuracy of an imaging modality, and Bayesian techniques that estimate this area are explained for ordinal and continuous data. Special software for Bayesian ROC analysis is introduced and illustrated with several examples analyzed by the author at MDACC. Also discussed are some specialized topics in test accuracy, such as problems of localization and detection, where multiple images are taken per patient. This induces a correlation between images within patients and special methods that take into account that correlation. The last topic of the chapter is on Bayesian sample size estimation for diagnostic accuracy.

In Chapter 6, the topics on diagnostic accuracy explained in the previous chapter are generalized to include patient covariate information. We know that test accuracy depends on many factors including patient characteristics, such as age, gender, therapy received, other biomarkers, stage of disease, etc. The importance of the risk score when taking into account patient covariates is emphasized and illustrated with many examples. Again, examples are based on the author’s consultation with investigators in the imaging division of MDACC.

Often, the clinical investigator wants to know the extent of agreement between several readers or observers who are making diagnostic decisions. For example, one study at MDACC consisted of five readers estimating the size of a lung cancer lesion at various times in response to therapy. It is important to know the inter- and intra-observer agreement because it has a bearing on the decision to declare a therapy a success or a failure. Or suppose several readers are using a confidence level ordinal score to classify the status of lesions seen on a mammogram of a screening trial for breast cancer. How well do the readers agree? Chapter 7 discusses the statistical methods for estimating the agreement between and within observers, including a Bayesian version of the Kappa statistic to estimate agreement with ordinal data; for continuous data, regression techniques (including a Bayesian version of Bland-Altman) for calibration will estimate the agreement. In another example, three readers, reading the same image, measure the length and width of the major axis of spicules observed on mammograms. To measure inter-observer variability, analysis of variance methods, using random effects for readers and patients, calculate the agreement via the intra-class correlation coefficient.

Imaging techniques are utilized to measure the extent of response to therapy. For example, in many Phase II clinical trials for disease with solid tumors, the efficacy of the therapy is measured by the change in the size of the lesion from start of treatment. Imaging the tumor size is crucial. Chapter 8 provides the necessary detail in explaining the protocol review process for cancer clinical trials, how the tumor response is categorized, using World
Health Organization (WHO) and response evaluation criteria in solid tumors (RECIST) criteria, and lastly how Bayesian sequential methods are employed to monitor the trial and to estimate the sample size. Also discussed is the software development of Bayesian methods for the design and analysis of clinical trials at MDACC. Examples taken from the protocol review at MDACC illustrate how to apply Bayesian methods to this important application of diagnostic medicine.

Chapter 9 introduces other topics in diagnostic medicine that are not considered in the previous eight chapters. For example, how is the accuracy of a test estimated when there is no reliable gold standard, or how is accuracy estimated when only those that test positive are subject to the gold standard? Or suppose the gold standard is not binary, but is possibly continuous, then how is accuracy to be determined? Thus, this chapter emphasizes topics that do not fit the standard mold, but are variations of the basic themes introduced in the previous chapters. Other areas of medicine in addition to diagnostic imaging employ diagnostic tests for the management of the patient. For example, the whole idea of biomarkers, including the expanding area of genetic microarrays, is to use such information for medical diagnoses or as prognostic factors for patient morbidity and survival.

1.4 Datasets for the Book

The datasets used for this book come from the following sources: (1) the protocol review process of clinical trials at MDACC, where the author was either a reviewer or a collaborator on the protocol; (2) the author’s consultations with the scientific and clinical faculty of the Division of Diagnostic Imaging at MDACC with some 32 datasets; (3) the six datasets accompanying the excellent book by M. S. Pepe (see: http://www.fhcrc.org/labs/pepe/Book) *The Statistical Evaluation of Medical Tests for Classification and Prediction*; (4) the information contained in the examples of the WinBUGS® package; and (5) other miscellaneous sources, including the examples and problems in *Statistical Methods in Diagnostic Medicine* by Zhou, McClish, and Obuchowski.2 Also, various articles by N. A. Obuchowski appearing in *The American Journal of Roentgenology* and *Academic Radiology* provided the author with useful information for this book, and several of her examples are included.

1.5 Software

WinBUGS will be used for the Bayesian analysis when sampling from the posterior distribution is appropriate. On the other hand, when direct sampling from the posterior distribution is called for, Minitab® is often employed
Introduction

for the posterior analysis. Many specialized Bayesian programs for the
design and analysis of clinical trials have been developed at the Department
of Biostatistics and Applied Mathematics at MDACC, some of which will be
used for the design of clinical trials.

Why is the Bayesian approach taken here? The author has been a Bayesian
theorist since 1974, when he was on a sabbatical leave to study at University
College London. A colleague persuaded him of the advantages of such an
approach. The main advantage, of course, is that it is a practical way to
utilize prior information, which, in a medical setting, is all around and
should be used to one’s advantage. It would be a pity not to use it.

References

1. Pepe, M.S., The Statistical Evaluation of Medical Tests for Classification and
Chapter 2

Diagnostic Medicine

2.1 Introduction

In this chapter is a brief description of diagnostic imaging and other diagnostic techniques routinely used at a major healthcare institution. At MD Anderson Cancer Center (MDACC), the Division of Diagnostic Imaging is made up of the following departments: Diagnostic Radiology, Experimental Diagnostic Imaging, Imaging Physics, Nuclear Medicine, and Interventional Radiology. There were approximately 100 faculty members during the 2003–2004 academic year. Diagnostic imaging provides an extremely important role in the overall care of the cancer patient, including diagnosis, staging, and monitoring of patients during their stay in the hospital. Most of the examples in this book are taken from diagnostic imaging studies; however, there are many other ways to perform diagnoses, and some of these are explained in Chapter 3.

2.2 Imaging Modalities

The primary modalities for diagnostic imaging are X-ray, fluoroscopy, mammography, computer tomography (CT), ultrasonography (US), magnetic resonance imaging (MRI), and nuclear medicine. Each one has advantages and disadvantages with regard to image quality, depending on the particular clinical situation. Broadly speaking, image quality consists of three components. The first is contrast. Contrast is good when important physical differences in anatomy and tissue are displayed with corresponding different shades of gray levels. The ability to display fine detail is another important aspect of image quality and is called resolution. Anything that interferes with image quality is referred to as noise, which is the third component. Obviously, noise needs to be minimized in order to improve image quality.

Medical images are best thought of as being produced by tracking certain probes as they pass through the body. A stream of X-rays is passed through
the patient and captured on film as the stream exits. An X-ray is a stream of photons, which are discrete packets of energy. As they pass through the body, various tissues interact with the photons and these collisions remove and scatter some of the photons. The various tissues reduce the amount of energy in various parts of the stream by different amounts. A shadow is produced that appears on special photographic film producing an image. If the density of the object that is the target is much higher than that of the surrounding environment (as bone), the X-ray does a good job of locating it. Some lesions have densities that are quite similar to the surrounding medium and are difficult to detect. Generally speaking, the X-ray has very good resolution and the noise is easy to control, but has low contrast in certain cases. The X-ray is routine in all medical settings and is the most utilized of all imaging devices.

A close relative of the X-ray is fluoroscopy. In this modality, the exiting beam is processed further by projecting it onto an image intensifier, which is a vacuum tube that transforms the X-ray shadow onto an optical image. This mode has about the same image quality as the X-ray, but allows the radiologist to manage images in real time. For example, it allows the operator to visualize the movement of a contrast agent past certain landmark locations in the gastrointestinal (GI) tract or vascular system.

Still another variation of the X-ray is computer tomography or CT, which overcomes some of the limitations of X-rays. The superimposition of shadows of overlapping tissues and other anatomical structures often obscure detail in the image. CT does produce images quite differently than X-rays; however it does use X-rays, but the detection and processing of the shadows is quite sophisticated and is the distinctive feature of the modality that vastly improves the image over that of the X-ray. CT has good contrast among the soft tissues (e.g., lung and brain tissue) and good resolution. The X-ray takes information from a three-dimensional structure and projects it onto a two-dimensional image, which causes the loss of detail due to overlapping tissues. To overcome this problem, the patient is placed in a circle and inside the circle is an X-ray source and embedded in the circle is an array of detectors that capture the shadow of the X-ray beam. The X-ray source irradiates a thin slice of tissue across the patient and the detector captures the shadow. The X-ray source moves to a close adjacent location and the process is repeated, say, 700 times. The X-ray source circumscribes the patient through 360 degrees. The source then repeats the above process with another thin slice. For a given slice, there are 700 projections of that slice and these 700 projections are processed via computer and back projection algorithms to produce the two-dimensional representation. The computer works backward from the projections to reconstruct the spatial distribution of the structure of the thin slice. In other words, CT answers the following question: What does the original structure need to resemble in order to produce the 700 generated projections?

A good example of CT (using the GE Imatron C-100 Ultrafast) is screening for coronary heart disease, where the coronary artery calcium score indicates the degree of disease severity. See Mielke et al. and Dasgupta where the accuracy of coronary artery calcium to diagnose heart disease is estimated
by the area under the receiving operating characteristic (ROC) curve. These examples will be examined from a Bayesian perspective in later chapters.

Mammography is still another variation of the X-ray. While some small masses can be detected by a physician or by self-examination, mammography has the ability to detect very small lesions. However, the smaller they are the more difficult they are to detect. Mammography consists of a specialized X-ray tube and generator, a breast compression device, an antiscatter grid, and film. The procedure must be able to reveal small differences in breast density, possibly indicative of suspicious mass, and it must also be able to detect small calcifications that may have importance for diagnosis. All the attributes of good image quality are required, namely high contrast, good resolution, and low noise. Later in this book, the role of mammography in screening for breast cancer will be described.

A completely different form of imaging is magnetic resonance imaging (MRI). A beam of photons is not passed through the body, but instead the body is placed in a large magnet and the hydrogen atoms (in the water molecules) line up in the same direction as the magnetic field. When the magnetic field is disrupted by directing radio energy into the field, the magnetic orientation of the hydrogen atoms is disrupted. The radio source is switched off and the magnetic orientation of the hydrogen atoms returns to the original state. The manner (referred to as T1 and T2 relaxation times) and way in which they return to the original state produces the image. Essentially, what is being measured is the proton density per unit volume of imaged material. The actual image looks like an X-ray; however the principal foundations of MRI are completely different. The same image processing technology as used in CT can be used in MRI to process the images. For example, thin slices and backward projection methods are often made to improve the MRI image quality. MRI has excellent resolution and contrast among the soft tissues and displays good anatomical detail.

Nuclear medicine is the joining of nuclear physics, nuclear chemistry, and radiation detection. A radioactive chemical substance called a radiopharmaceutical is injected, usually intravenously (IV), where it concentrates in a particular tissue or organ of interest. The substance emits gamma rays that are detected by gamma cameras and then the captured gamma particles are counted by the camera. There are two principal gamma cameras: PET (positron emission tomography) and SPECT (single photon emission computed tomography). Nuclear imaging is often used to view physiological processes. For example, FDG-PET (florodeoxyglucose-PET) is frequently used to measure glucose metabolism, where the radiopharmaceutical (18) F-florodeoxyglucose is absorbed by every cell in the body. The higher the observed radioactivity as measured by PET, the higher the glucose metabolism. In some cancer studies, the malignant lesion has an increased glucose metabolism compared to the adjacent nonmalignant tissue and, thus, is useful in the diagnosis and staging of disease.

Another area where nuclear medicine is useful is in cardiac perfusion studies. For example, radiation therapy of esophageal cancer often induces