RANKING ESTIMATION METHODS IN MEDICAL IMAGING

WITHOUT THE USE OF A GOLD STANDARD

by

John W. Hoppin

———————————————

A Dissertation Submitted to the Faculty of the

GRADUATE INTERDISCIPLINARY PROGRAM IN APPLIED MATHEMATICS

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2 0 0 3

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have

read the dissertation prepared by___ John W. Hoppin

entitled ___ Ranking Estimation Methods in Medical Imaging Without a

___ Gold Standard

and recommend that it be accepted as fulfilling the dissertation

requirement for the Degree of Doctor of Philosophy

(signature) Harrison H. Barrett		June 19, 2003 Date
(signature) Eric Clarkson		June 19, 2003 Date
(signature) Matthew A. Kupinski		June 19, 2003 Date
		Date
		Date

Final approval and acceptance of this dissertation is contingent upon
the candidate's submission of the final copy of the dissertation to the
Graduate College.

I hereby certify that I have read this dissertation prepared under my
direction and recommend that it be accepted as fulfilling the dissertation
requirement.

(signature) Harrison H. Barrett   Aug 12, 2003
Dissertation Director   Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: _____

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor and friend Harry Barrett, the gold standard of dissertation advisors. He has been a constant source of knowledge and guidance and an example of how one should carry themselves in academia. I would also like to thank my other "bosses" Matt Kupinski and Eric Clarkson for all the time and effort they put into my education, not to mention sharing some good laughs. I cannot say enough good things about Harry, Matt, and Eric.

Now that I've thanked the theorists, I certainly need to give a shout-out to those that helped me with the experimental validation of the theory. Specifically, Todd Peterson and Don Wilson, who were of great help with the entire experiment, and no help whatsoever in our daily trivia game. Furthermore, I would like to thank Ben Gershman, Giorgios Kastis, and Lars Furenlid for all of their assistance with the imaging systems and phantom.

The lab has been a great place to work in terms of both good people and good scientists (even a couple combinations thereof). Therefore I would like to thank a few others, namely Lisa Gelia, Zhonglin Liu, Art Gmitro, Angel Pineda, Jane Lockwood, Jack Denny, and Dennis Patton for their help at one time or another, for one thing or another.

A big part of my time here was spent learning math across campus. That said, I certainly would not be completing this degree if it had not been for the help of Emily Lane, Dan Coombs, Joceline Lega, Tom Kennedy, and Michael Tabor. Even further away I would like to thank my undergraduate advisor, and former student of Harry's, Ed Soares along with my outside reader Ruth Pfieffer at the National Cancer Institute.

Finally, I would like to thank all of my family and friends for their support over the years. A special nod of thanks to my parents and my fiancé Janna - thanks for putting up with me..

TABLE OF CONTENTS

TABLE OF CONTENTS—*Continued*

TABLE OF CONTENTS—*Continued*

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES—*Continued*

LIST OF FIGURES—*Continued*

ABSTRACT

Imaging is often used for the purpose of estimating the value of some parameter of interest. For example, a cardiologist may measure the ejection fraction (EF) of the heart in order to know how much blood is being pumped out of the heart on each stroke. In clinical practice, however, it is difficult to evaluate an estimation method because the *gold standard* is not known, *e.g.*, a cardiologist does not know the true EF of a patient. Thus, researchers have often evaluated an estimation method by plotting its results against the results of another (more accepted) estimation method, which amounts to using one set of estimates as the pseudo-gold standard. In this dissertation, we present a maximum-likelihood approach for evaluating and comparing different estimation methods without the use of a gold standard with specific emphasis on the problem of evaluating EF estimation methods. We have named this method Regression Without Truth or RWT.

Results of numerous simulation studies will be presented and indicate that the method can precisely and accurately estimate the parameters of a regression line without a gold standard, *i.e.*, without the x-axis. We also characterize the performance of this method in comparison to conventional regression analysis using x-axis information. Also in this work we calculate the Fisher information for our method to quantify the performance of our evaluation method. Results of simulation studies are presented to show that we are very nearly efficient at estimating the parameters used in our method. In an attempt to further validate RWT we present the results of a volume estimation experiment using a physical phantom and two imaging systems (SPECT,CT).

We conclude the dissertation with a discussion of the strengths and weaknesses of RWT. In an attempt to aid users of RWT we provide multiple consistency checks for users to evaluate results of RWT. Finally, we present some areas of potential application for RWT.

Chapter 1

INTRODUCTION

A great deal of research in the field of medical imaging is dedicated to image-system design and image processing. Researchers often claim to have a system or method that produces "better" images, although they rarely support their claims with objective comparative studies. Even when objective comparisons are made, rarely do such comparisons account for the reason the images were taken and/or the observer reading the images. We advocate a statistical, task-based approach to the objective assessment of image quality [1–3]. To properly quantify the performance of an imaging system or image processing technique, we define the task the image is to be used for and measure the performance of an observer performing this task.

The most important tasks in medical imaging are classification and estimation tasks. Classification tasks consist of an observer determining some categorical (usually binary) diagnosis for a patient, e.g., tumor present or tumor absent. This task may be performed by a human observer, a computer observer, or some combination thereof. The performance of this observer may be assessed using psychophysical studies and receiver operating characteristic (ROC) analysis in the case of binary decision tasks [4]. Figures of merit such as sensitivity, specificity, and/or the area under the ROC curve (AUC) can then be used to compare different imaging modalities.

Along with classification, imaging is often performed to estimate some parameter of interest that will aid a physician in diagnosis. Examples of estimation tasks in medical imaging include estimating bone density [5] , blood-oxygen levels [6] , and cardiac ejection fraction [7]. Estimation tasks typically involve computer observers, though sometimes these computer observers require human intervention, e.g., defining regions of interest [8, 9]. The performance of estimation tasks is typically measured using the bias and variance

of the estimate, often combined to form mean-squared error (MSE) which can be used as a scalar figure of merit.

More than one imaging modality often exists to perform both classification and estimation tasks [10]. In order to evaluate the performance of an imaging modality we need to compare this performance to the truth. The gold standard is the method that returns the true estimate of the parameter. The presence of a gold standard allows for the evaluation of new imaging techniques via figures of merit such as AUC and MSE. A gold standard rarely exists for most imaging problems. Even when one does exist, the financial costs and patient risk are often too high for practical application. It is extremely rare for an inexpensive, safe, always correct classification or estimation technique to exist while other techniques are still under development.

In the case of classification tasks, biopsy is typically considered the gold standard [11]. The problems with biopsy and histological analysis is that they are not always available, and, even when available, are subject to errors. For example, biopsy will give information about those lesions detected in images. If, however, a lesion is not detected in an image, then its contribution to the false-negative fraction remains unmeasurable [12].

The problem of comparing classification tasks without a gold standard was first addressed successfully in the medical imaging literature by Henkelman, Kay, and Bronskill in their work on ROC analysis without truth in 1990 [13]. Since the publication of their work their has been a great deal of research performed comparing classification tasks without a gold standard [14–20].

Similar problems regarding gold standard techniques exist for estimation tasks. For example, in cardiac studies, ventriculography or ultrasound might be taken as the gold standard for estimation of EF, and nuclear medicine or dynamic MRI might then be compared to the supposed standard [10]. A very common graphical device is to plot a regression line of EFs derived from the system under study to those derived from the standard and to report the slope, intercept and correlation coefficient (r) for this regression [21–25]. Another comparison approach is the use of a Bland-Altman plot, which attempts to measure

the agreement between two different modalities [21, 23–26]. Neither of these approaches allows for objective performance rankings of the imaging systems, a point we expand upon in Chap. 2. In reality, no present modality can lay claim to the status of gold standard for the estimation of quantitative cardiac parameters, and thus the aforementioned comparison techniques suffer from major inconsistencies. The goal of this work is to address the no-gold-standard problem for estimation tasks.

We have developed a method for evaluating and comparing multiple estimation methods without the benefit of a gold standard [27–29]. We have named it Regression Without Truth (RWT). The goal of this method is to quantify the relationship between estimates of a parameter of interest and the underlying gold standard without knowledge of the gold standard. Stated differently, if researchers are to estimate a quantity of interest using multiple techniques on a set of P patients, we would like to quantify the relationship between those estimates and the true value of the quantities of interest without assuming any given technique to be the gold standard.

In Chap. 3 we derive a method we have called Regression Without Truth (RWT) that assesses the relationship between the estimates and the so-called ground truth without assuming a gold standard. Chaps. 3 and 4 contain the results of numerous simulations studies indicating the success of RWT. Also in Chap. 3, we discuss how researchers can apply the results of RWT to better rank and compare different estimation methods.

In Chaps. 3 and 4 we were able to assess the performance of RWT because all of our work was done in simulation, *i.e.*, we did in fact know the gold standard. Much of this work has been received with the justified skepticism associated with computer simulations. Chaps. 5 and 6 attempt to quell such skepticism. In Chap. 5 we derive an expression for the Fisher information matrix for RWT, from which we ascertain the Cramér-Rao (CR) lower bounds for the variances of the model parameters estimated using RWT [30–32]. We also show that the variances calculated experimentally are very close to the CR bounds calculated using the Fisher information. In Chap. 6 we present the results of a volume estimation experiment performed using multiple modalities with a gold standard. We use

the results of this experiment to assess the performance of RWT when applied to real data.

We conclude the dissertation in Chap. 7 with a summary and evaluation of RWT, along with a discussion of future applications for RWT.

Chapter 2

BACKGROUND

As discussed in the introduction, there exist many medically relevant parameters physicians would like to estimate that do not have a gold standard estimator. In this dissertation we focus our attention on cardiac EF. We chose EF because it is prevalent in the literature and because it is bound between zero and one, a point we return to later in the dissertation. There have been numerous studies performed comparing cardiac EF estimation techniques [21–26]. The focus of this chapter is to discuss the different EF estimation techniques used in practice and the methods used in the literature to compare these techniques.

2.1 Cardiac Ejection Fraction

For the sake of completeness we present a brief discussion of current estimation procedures for cardiac ejection fraction (EF). Cardiac EF measures the difference in the end-diastolic and end-systolic volumes divided by the end-diastolic volume. In other words, it records the fraction of the blood in the left ventricle pumped out into the body in a given heart cycle. Physicians measure EF for patients displaying cardiac problems, for individuals about to undergo a strenuous treatment program such as chemotherapy, or even as part of an annual physical examination.

Sharir *et al.* [7] showed that the likelihood of cardiac death increases exponentially with decreasing EF. Due to this relationship, there has been increased interest in assessing the accuracy of EF measurements. Cardiac EF even played a role in the 2000 presidential elections when the election hopes of current Vice President Richard Cheney were almost squashed when he suffered a mild heart attack and was reported to have an EF around 40% [33], approximately a 2.5% chance of cardiac death according to Sharir *et al.*. With heart disease as the leading cause of death in the United States (1 out of every 2.4 deaths) [34], it

is no surprise that the advancement of cardiac EF estimation techniques has become a large field of research.

There currently exist five different approaches to estimating cardiac EF: left ventricular angiography, echocardiography, nuclear medicine, electron beam computed tomography, and magnetic resonance imaging. According to labs at the University Medical Center at the University of Arizona cardiac is estimated roughly 125 times per month using left ventricular angiography, 40 times per month using radionuclide angiography, and on the order of 350 to 400 times per month using echocardiography. The UMC does not have electron beam computed tomography imaging systems, nor do are they currently using magnetic resonance imaging to estimate cardiac EF. In this section we will discuss what these five procedures entail, along with their pros and cons. For a more extensive discourse on EF estimation methods we refer the reader to Rumberger *et al.* [10].

2.1.1 Left Ventricular Angiography

The process of left ventricular angiography (LVA) consists of placing a catheter, typically through the left ventricular cavity via the aortic valve. The catheter then disperses a contrast agent enabling the acquisition of a dynamic fluoroscopic x-ray image. Typically, three to six heart cycles are imaged. Ideally, biplane imaging is performed, though for practical purposes LVA is often performed using only a single planar projection [35]. Estimates of the EF are then made by assuming that the left ventricle is an ellipsoid. Under this assumption end-diastolic and end-systolic volumes can be estimated by determining the semimajor axis (height) and the two semiminor axes (length and width) using the biplanar projections. The semiminor axes are assumed to be equal when using monoplane imaging.

Biplane LVA is often considered the gold standard technique for estimating EF. This assumption is incorrect. While LVA, developed in the 1950's, is the oldest and most established imaging technique used to estimate EF, it does not estimate EF with the type of reproducibility required of even a pseudo-gold standard. The technique has a mean in-

terobserver difference of approximately 5% [36] and a reproducibility of approximately 10% [37]. These errors arise due to the invasive nature of the procedure as well as the incorrect assumption concerning the geometry of the left ventricle. Validation work has been done using cadaver hearts, but there are severe limitations to the conclusions one can draw from such studies [38].

2.1.2 Echocardiography

Echocardiography is the practice of cardiac imaging using ultrasound. Estimation of EF is often performed using echocardiography since the procedure is safe, quick, and inexpensive. Once imaging is performed, estimates of the EF range, *e.g.* 60 to 65%, are made via simple visual assessment. This approach tends to produce fairly high interobserver variability and is heavily dependent on the experience of the sonographer [39]. As a result of this error there exist numerous quantification and semiquantification techniques for estimating EF using echocardiography.

One such approach often used consists of estimating the left ventricular end-diastolic diameter (LVEDD) and the left ventricular end-systolic diameter (LVESD) using an electrocardiogram and then estimating cardiac EF via

$$EF = (LVEDD^2 - LVESD^2)/LVEDD^2. \tag{2.1}$$

This estimation technique is extremely limited by assuming an ellipsoidal left ventricular shape.

Another approach to estimating EF consists of imaging the apical views and using the area-length method or some sort of modified Simpson's rule [40]. This approach is again limited by shape assumptions for the left ventrical and/or segmentation of the images. Echocardiography is a relatively new technology with plenty of room for improvement, but it will never lay claim to gold standard status for EF estimation given the resolution of ultrasound.

2.1.3 Radionuclide angiography

Radionuclide angiography consists of collecting planar cardiac projections using an Anger gamma camera with a parallel-hole collimator imaging the tracer radionuclide ^{99m}Tc. The left ventrical is manually segmented at multiple stages of the heart cycle and counts within the segmented ventrical are used to calculate the EF. As a result, unlike LVA and Echocardiography, radionuclide angiography does not make any assumptions about the geometry of the left ventricular cavity. .

There are numerous ways of measuring cardiac EF using radionuclide angiography. The most common approach is a multiple gated acquisition (MUGA) study, also called a gated blood-pool scan. MUGA consists of injecting a very small amount of radiotracer into the blood stream, often mixed with a patients blood prior to injection for better uniformity. Gated cardiac imaging is then performed using an electrocardiogram, and 20 to 100 frames are observed per heart cycle. This process is carried for the duration of around 1000 heart beats. Abnormal heart beats are removed, and the left ventricle is manually segmented in the planar images [41]. The maximum number of counts in a given cycle is considered the end-diastole, and likewise the minimum is considered the end-systole. The number of counts is assumed proportional to the volume and EF is estimated by the following,

$$EF = (diastolic\ counts - systolic\ counts)/(diastolic\ counts). \tag{2.2}$$

Cardiac EF is more difficult to study in patients with larger hearts due to increased gamma-ray absorption, but corrections can be made in such situations. The largest problems in MUGA studies are assuming a uniform distribution of the radiotracer in the bloodstream and the resolution of gamma-ray imaging.

2.1.4 Electron Beam Computed Tomography

Electron beam computed tomography (EBCT) or ultrafast-CT was designed by Boyd [42] in the early 1980's. Gated imaging is performed using a iodinated contrast median and

end-diastolic and end-systolic volume estimation is straightforward given the quality of the tomographic images. In principle this is an ideal technique for estimating EF, but from a clinical standpoint, EBCT is at this point impractical. EBCT involves a large exposure (~5 rad for a cardiac study) and the use of iodinated contrast media, making serial studies impossible.

2.1.5 Magnetic Resonance Imaging

Estimating EF using cardiac MRI has become possible due to general innovations in the imaging modality. EF can be estimated with MRI using echo-planar imaging sequences allowing for numerous images to be collected during a heart cycle. Biplane MRI can be performed and EF estimation can then be accomplished as it is in LVA and echocardiography. Motion artifacts, claustrophobia, and time of acquisition are the major problems present in cardiac MRI. Breath-hold MRI techniques allow for better cardiac imaging, though in the case of EF the process of holding one's breath impacts the *true* cardiac EF. The confines of an MRI machine also impact cardiac imaging in about 10-20% of patients due to cardiac motion artifacts.

Acquisition time is a problem in cardiac imaging due to the beating of the heart. It is also a problem in general since imaging time on MRI systems are in high demand. While newer MRI systems have software for calculating EF, older systems must perform estimation off-line. As a result, many hospitals are simply incapable of performing functional cardiac imaging.

2.2 Current Methods of Comparison

Comparing estimation techniques when a gold standard is available is rather straightforward. For example, let us envision an experiment in which we would like to compare two newly developed EF estimation techniques, namely Estimator I and Estimator II, and let us assume that a gold standard technique GS exists. We can simply measure the EFs for

a group of patients using Estimator I, Estimator II, and GS and then compare the two new techniques using linear regression analysis. In Fig. 2.1 we show plots of the Estimator I and Estimator II estimates versus the gold standard. In each plot we have estimated the slope and intercept of the regression line along with a parameter σ that quantifies the spread of the data away from the regression line. Using these parameters we can then compare the two techniques quantitatively by calculating their respective reproducibility, a point we expand upon in the next chapter. Furthermore, we can calibrate either of our new estimators using the respective slope and intercept determined by the linear regression analysis so as to increase the accuracy of our estimator in the future.

As stated in the introduction, the current techniques used to compare different estimation tasks are regression and/or Bland-Altman analysis. The first attempts to use values of the slope, intercept, and correlation coefficient (r) to compare the performance of a new, less accepted, modality to a more accepted pseudo-gold standard. An example is given in Fig. 2.2. This approach has many shortcomings. Unlike the example discussed above in which a gold standard existed, the importance of the calculated slope, intercept, and r values are unclear. If a researcher were to believe in the estimates given by the pseudo-gold standard, then the slope and intercept of the regression line could, as discussed, be used to adjust the new estimates to calibrate the new system. This calibration is not often done, however, because rarely does such confidence exist in any of the available estimation techniques.

Calculating the correlation coefficient r for the regression plot is not particularly informative when comparing two estimation tasks [43–45]. A non-zero value of r implies correlation, which is of very little help considering the two estimators are attempting to measure the same quantity. Rather, researchers would like to state that a large r value implies strong agreement. This is not necessarily true. The value of r depends on the magnitude of the spread of the data points around the regression line *and* the variance of the true parameter across the subjects. As a result, the interpretation of r can be very misleading. For example, if for a given comparative study we were to measure the EFs for 100 patients
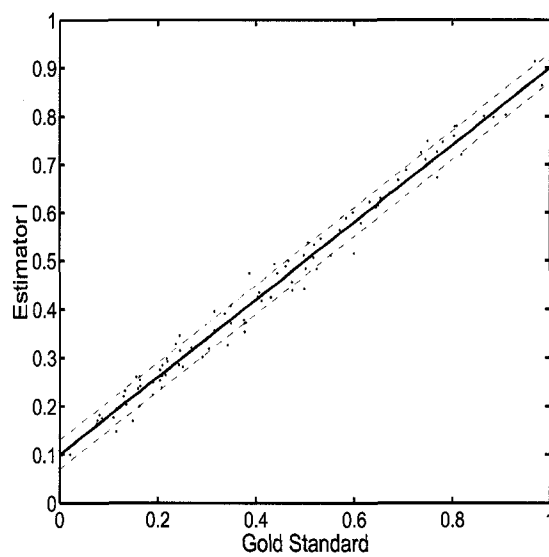
(a)
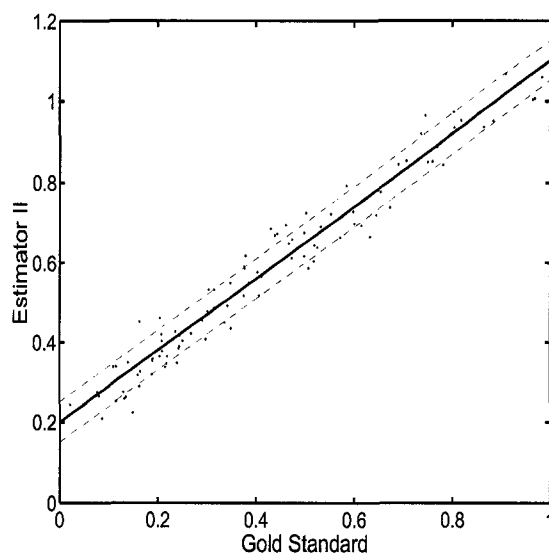
(b)

FIGURE 2.1. An example of regression analysis when a gold standard is available.
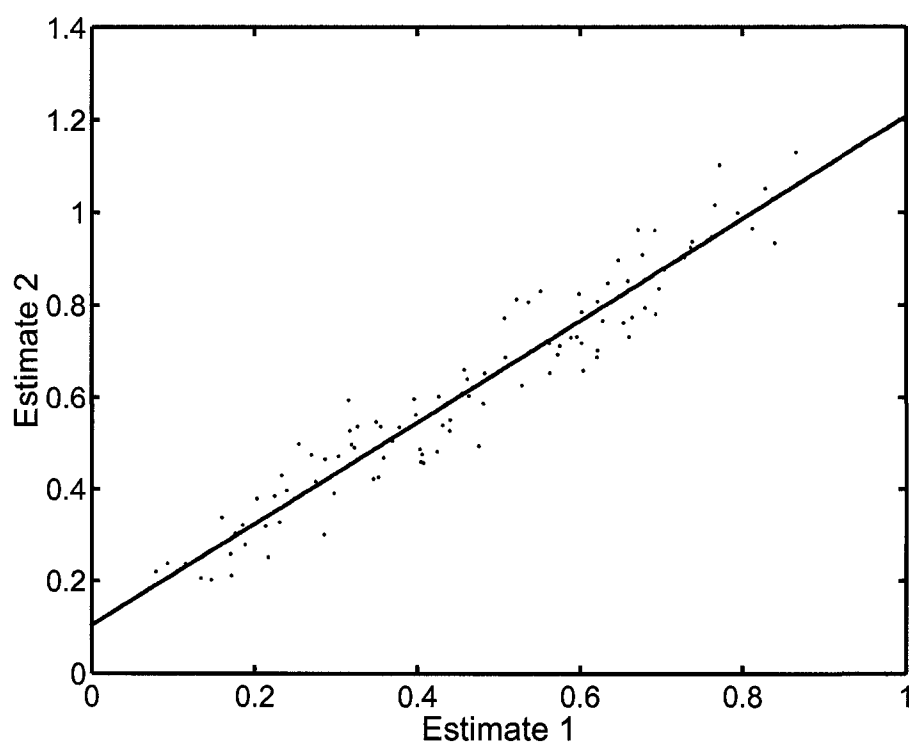
FIGURE 2.2. An example of a regression plot comparing two sets of estimates.

with true EFs between 0.6 and 0.7 using two different modalities we would very likely have a lower r value than if we were to run the same study, using the same modalities to measure the EFs for 100 patients with EFs between 0.4 and 0.9.

Bland and Altman presented a simple approach to comparing estimation techniques in 1983 which attempts to quantify the level of agreement between two methods for calculating the same quantity [43]. Given two sets of estimates for the same parameter the Bland-Altman plot depicts the difference between the estimates vs. the mean of the estimates. An example is shown in Fig. 2.3. If 95% of the estimates fall within two standard deviations of the mean of the differences, then the the two methods of estimation are said to "agree" and thus one method could, in theory, replace another. A shortcoming of this approach lies in the definition of agreement which appears to be rather arbitrary. Their definition implies that if the differences of the estimates follow a Gaussian distribution then "agreement" is achieved independent of how big or small those differences are.

Both regression and Bland-Altman analysis attempt to draw conclusions about the correlation or agreement between two methods. Neither approach determines which method is doing a better job of estimating the true parameter of interest. In order to achieve such a ranking one must incorporate the underlying true parameter into the method of comparison even though it is unknown. In this dissertation we present and evaluate a method of comparison that does just that.
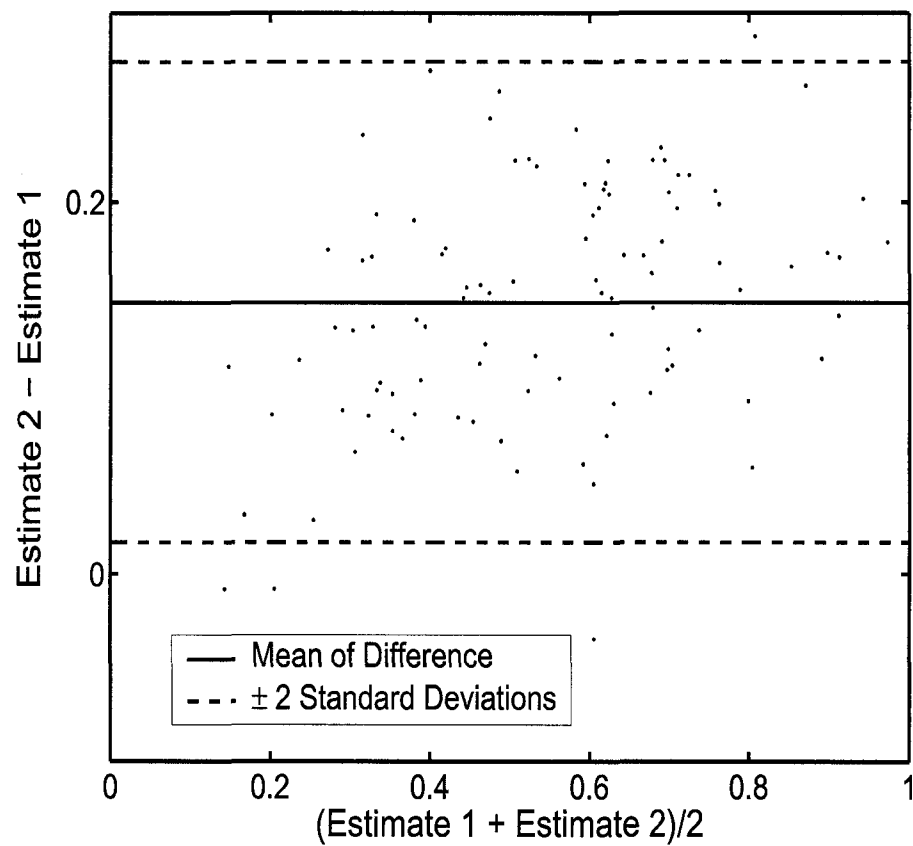
FIGURE 2.3. An example of a Bland-Altman plot used to quantify the "agreement" between two different estimation techniques. Note, this plot is just a rotation and rescaling of Fig. 2.2.

Chapter 3

REGRESSION WITHOUT TRUTH

Ideally we would like to be able to compare estimation techniques by performing regression analysis with a gold standard technique plotted along the x axis. In this chapter, we develop a statistical model that allows us to perform regression analysis without a gold standard. In Sec. 3.2 we present the results of numerous simulation studies used to analyze the performance of our method. In Sec. 3.3, we discuss our method in the context of the current statistics literature for the sake of completeness. We conclude the chapter with a discussion of the assumptions made in our model, the results of our simulations, and how the results of our technique can be used to compare estimation techniques.

3.1 Approach

We begin with the assumption that there exists a linear relationship between the true EF and its estimated value. We will describe this relationship for a given modality m and a patient p using a regression line with a slope a_m, intercept b_m, and noise term ϵ_{pm}. We represent the true EF for a given patient with Θ_p and an estimate of the EF made using modality m with θ_{pm}. The linear model is thus represented by

$$\theta_{pm} = a_m \Theta_p + b_m + \epsilon_{pm}. \tag{3.1}$$

We make the following assumptions:

1. Θ_p does not vary for a given patient across modalities and is statistically independent from patient to patient.

2. The parameters a_m and b_m are characteristic of the modality and independent of the patient.

3. The error terms, ϵ_{pm}, are statistically independent and normally distributed with zero mean and variance σ_m^2.

Note that the zero-mean assumption for ϵ_{pm} entails no loss of generality since the mean is lumped into b_m in equ. 3.1. Using assumption number 3 we write the probability density function for the noise ϵ_{pm} for a given patient p and M modalities as

$$pr(\{\epsilon_{pm}\}) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2\sigma_m^2}\epsilon_{pm}^2\right), \tag{3.2}$$

where the term $\{\epsilon_{pm}\}$ signifies the set of M noise terms. In other words, we assume a multivariate noise model with a diagonal covariance matrix. We could relax this assumption by adding non-zero terms in the off-diagonal components of the covariance matrix. One could also assume a different noise model, even one that is dependent on Θ_p. Solving for ϵ_{pm} in equ. 3.1, we rewrite equ. 3.2 as the probability of the estimated EFs for multiple modalities and a specific patient given the linear model parameters (a_m's,b_m's,σ_m's) and the true EF as

$$pr(\{\theta_{pm}\}|\{a_m, b_m, \sigma_m^2\}, \Theta_p) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(-\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2\right). \tag{3.3}$$

The notation $\{\theta_{pm}\}$ represents the estimated EFs for a given patient p over M modalities. Using the following property of conditional probability

$$pr(x_1, x_2) = pr(x_1|x_2)pr(x_2), \tag{3.4}$$

as well as the marginal probability law,

$$pr(x_1) = \int dx_2 \, pr(x_1, x_2), \tag{3.5}$$

we write the probability of the estimated EF for a specific patient across all modalities given the linear model parameters as

$$pr(\{\theta_{pm}\}|\{a_m, b_m, \sigma_m^2\}) =$$
$$\int d\Theta_p \, pr(\Theta_p)S \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2\right), \tag{3.6}$$

where

$$S = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}}. \tag{3.7}$$

From assumption number 1 above, the likelihood of the linear model parameters can be expressed as

$$L = \prod_{p=1}^{P} \left(S \int d\Theta_p \, pr(\Theta_p) \exp \left(\sum_{m=1}^{M} \left(-\frac{1}{2\sigma_m^2} (\theta_{pm} - a_m\Theta_p - b_m)^2 \right) \right) \right), \tag{3.8}$$

where P is the total number of patients. Upon taking the log, we rewrite products as sums we obtain,

$$\lambda = \ln(L) = P\ln(S) +$$
$$\sum_{p=1}^{P} \ln \left(\int d\Theta_p \, pr(\Theta_p) \exp \left(\sum_{m=1}^{M} \left(-\frac{1}{2\sigma_m^2} (\theta_{pm} - a_m\Theta_p - b_m)^2 \right) \right) \right). \tag{3.9}$$

It is this scalar λ, the log-likelihood, that we seek to maximize to obtain our estimates of a_m, b_m, and σ_m^2. These estimates will be maximum-likelihood estimates for our parameters when the data match the model. Although $pr(\Theta_p)$ may appear to be a prior term, we are *not* using a maximum-*a posteriori* approach; we are simply marginalizing over the unknown parameter Θ_p which we are treating as a nuisance parameter. We are *not* estimating Θ_p, rather we are estimating the linear model parameters in an attempt to compare the different modalities. Thus we have derived an expression for the log-likelihood of the model parameters which does not require knowledge of the true EF Θ_p, *i.e.*, without the use of a gold standard. This procedure is analogous to fitting lines without the use of the x axis.

3.1.1   True ($pr_t(\Theta_p)$) versus Assumed ($pr_a(\Theta_p)$) Distributions

Although the expression for the log-likelihood in equ. 3.9 does not require the true EF Θ_p, it does require some knowledge of their distribution $pr(\Theta_p)$. We will refer to this distribution, as it appears in equ. 3.9, as the assumed distribution ($pr_a(\Theta_p)$) of the EFs. In this chapter we will investigate the effect different choices of the assumed distributions have on estimating

the linear model parameters. We first sample parameters from a true distribution $(pr_t(\Theta_p))$ and generate different estimated EFs for the different modalities by linearly mapping these values using known a_m's and b_m's, then add normal noise to these values with known σ_m's. These EF estimates form the values θ_{pm}, that will be used in the process of determining the estimates of the linear model parameters by optimizing equ. 3.9. We will look at cases in which the assumed and true distributions match (data match model), as well as cases in which they do not match (data do not match model).

For our experiments we will investigate beta distributions and truncated normal distributions as our choices for both the assumed and true distributions. These distributions have been chosen because EF is bounded between 0 and 1 and has been shown to follow a unimodal distribution [7, 46]. Furthermore, Kastis showed EF data to be well fit by beta and truncated normal distributions [46]. The beta distribution has probability density function given by

$$pr(\theta) = \frac{\theta^{\nu-1}(1-\theta)^{\omega-1}}{B(\nu,\omega)}, \tag{3.10}$$

where $\theta \in [0,1]$ and the beta function $B(\nu,\omega)$ is a normalizing constant. The truncated normal distribution is given by

$$pr(\theta) = A(\mu,\sigma)\exp\left(-\frac{1}{2\sigma^2}(\theta-\mu)^2\right)\Pi(\theta), \tag{3.11}$$

where $A(\mu,\sigma)$ is a normalizing constant involving error functions and $\Pi(\theta)$ is a rect function that truncates the normal from 0 to 1. It should be noted that μ and σ are the mean and standard deviation for the normal distribution, not necessarily the mean and standard deviation of the truncated normal. While ν, ω, μ, and σ appear to be hyperparameters they are not; they are simply parameters characterizing the density, $pr(\Theta_p)$, that we used to marginalize Θ_p in equ. 3.3. The beta and truncated normal distributions are discussed in more detail in App. A.

Using a truncated normal for the assumed distribution in equ. 3.9, we find the following

closed-form solution for the log-likelihood:

$$\lambda = P\ln(S) + \sum_{p=1}^{P}\left(\frac{\beta^2 - 4\alpha\gamma}{4\alpha}\right)\ln\left(\frac{A(\mu,\sigma)}{2}\sqrt{\frac{\pi}{\alpha}}\left[\text{erf}\left(\frac{2\alpha+\beta}{2\sqrt{\alpha}}\right) - \text{erf}\left(\frac{\beta}{2\sqrt{\alpha}}\right)\right]\right)$$

(3.12)

where

$$\alpha = \frac{1}{2\sigma^2} + \sum_{m=1}^{M}\frac{a_m^2}{2\sigma_m^2},$$

$$\beta = -\frac{\mu}{\sigma^2} - \sum_{m=1}^{M}\frac{a_m(\theta_{pm} - b_m)}{\sigma_m^2},$$

$$\gamma = \frac{\mu^2}{2\sigma^2} + \sum_{m=1}^{M}\frac{(\theta_{pm} - b_m)^2}{2\sigma_m^2}.$$

The expression for the log-likelihood with a beta assumed distribution does not easily simplify to a closed-form solution, and thus we used numerical integration techniques to evaluate the one-dimensional integral in equ. 3.9.

We used a quasi-Newton optimization method in Matlab on a Dell Precision 620 running Linux to maximize the log-likelihood as a function of our parameters [47, 48]. For each experiment we generated EF data for 100 patients using one of the aforementioned distributions. We then ran the optimization routine to estimate the parameters and repeated this entire process 100 times in order to compute sample means and variances for the parameter estimates. The tables that follow consist of the true parameters used to create the patient data as well as the sample means and standard deviations obttained through the simulations.

3.2 Results

3.2.1 Estimating the Linear Model Parameters for a Given Assumed Distribution

We first investigated the results of choosing the assumed distribution to be the same as the true distribution. The asymptotic properties of maximum-likelihood estimates would

TABLE 3.1. Values of the estimated linear model parameters using matching assumed and true distributions.

	a_1	a_2	a_3	b_1	b_2	b_3
True Values	0.6	0.7	0.8	-0.1	0.0	0.1
$pr(\Theta)$=Beta	0.59±.03	0.69±.03	0.79±.05	-0.10±.02	0.00±.02	0.11±.03
$pr(\Theta)$=Normal	0.58±.04	0.68±.04	0.78±.06	-0.09±.02	0.01±.02	0.11±.03
		σ_1	σ_2	σ_3		
True Values		0.05	0.03	0.08		
$pr(\Theta)$=Beta		0.048±.005	0.029±.009	0.079±.007		
$pr(\Theta)$=Normal		0.048±.006	0.028±.010	0.080±.007		

predict that in the limit of large patient populations the estimated linear model parameters would converge to the true values [49]. The results, shown in Table 3.1, are consistent with this prediction. For the experiment below we have chosen $\nu = 1.5$ and $\omega = 2$ for the beta distribution and $\mu = 0.5$ and $\sigma = 0.2$ for the truncated normal distribution. Fig. 3.1 illustrates the results of an individual experiment using the truncated normal distribution.

In an attempt to understand the impact of the assumed distribution on the method we next used a flat assumed distribution, which is in fact a special case of the beta distribution ($\nu = 1$, $\omega = 1$). We used the same beta and truncated normal distributions for the true distribution as was chosen in the previous experiment, namely $\nu = 1.5, \omega = 2, \mu = 0.5$ and $\sigma = 0.2$. As shown in Table 3.2, the parameters estimated using a flat assumed distribution are not as accurate as those in the experiment with matching assumed and true distributions. However, the systematic underestimation on the a_m's and the systematic overestimation on the b_m's has not affected the ordering of these parameters. In fact, the estimated parameters have been shifted roughly the same amount. It should also be noted that the estimates of the σ_m's are still accurate. We will return to both of these points later in this chapter.

(a)

(b)

(c)
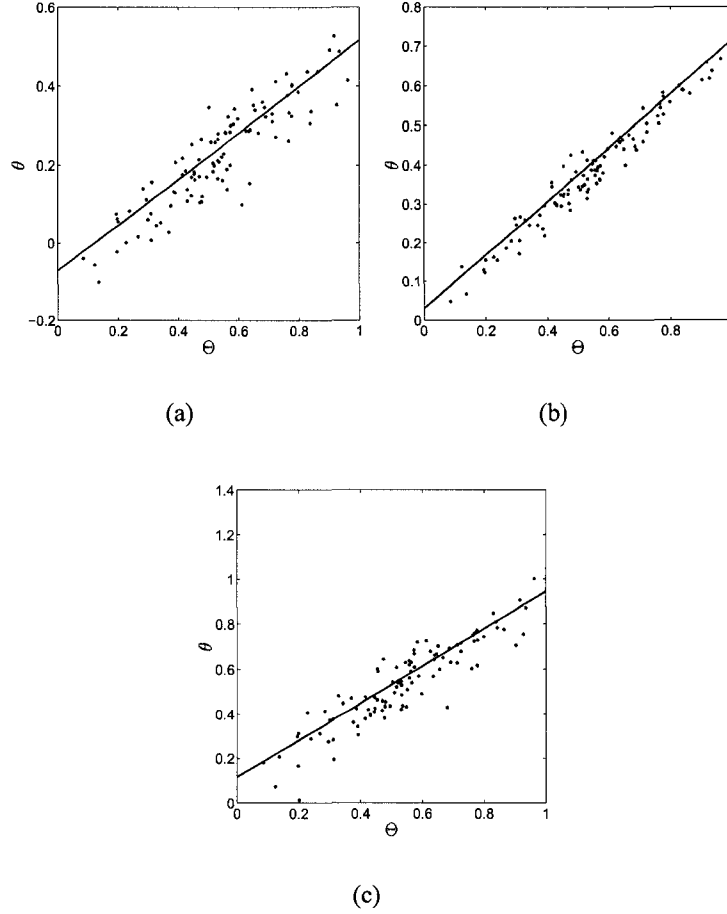
FIGURE 3.1. The results of an experiment using 100 patients, 3 modalities, and the same true parameters as shown in Table 3.1. In each graph we have plotted the true ejection fraction against the estimates of the EF for three different modalities ((a), (b) and (c)). The solid line was generated using the estimated linear model parameters for each modality. The dashed lines denote the estimated standard deviations for each modality. The estimated a_m, b_m and σ_m for each graph are (a)0.59, -0.07, 0.06, (b)0.69, 0.03, 0.025 and (c)0.83, 0.12, 0.082. Note that although we have plotted the true EF on the x-axis of each graph, this information was not used in computing the linear model parameters.

TABLE 3.2. Values of estimated linear model parameters using a flat assumed distribution $(pr_a(\Theta) = 1)$.

	a_1	a_2	a_3
True Values	0.6	0.7	0.8
$pr_t(\Theta)$=Beta	0.53±.03	0.61±.03	0.70±0.05
$pr_t(\Theta)$=Normal	0.50±.01	0.56±.03	0.64±.08
	b_1	b_2	b_3
True Values	-0.1	0.0	0.1
$pr_t(\Theta)$=Beta	-0.09±.02	0.02±.02	0.13±.03
$pr_t(\Theta)$=Normal	-0.05±.02	0.07±.03	0.18±.04
	σ_1	σ_2	σ_3
True Values	0.05	0.03	0.08
$pr_t(\Theta)$=Beta	0.049±0.005	0.031±0.009	0.079±0.007
$pr_t(\Theta)$=Normal	0.048±0.005	0.033±0.008	0.080±0.007

3.2.2 Estimating the Linear Model Parameters and the Parameters of the Assumed Distribution

After noting the impact of the choice of the assumed distribution on the estimated parameters, we investigated the effect of varying this distribution. In the case of the beta distribution this was simply a case of adding ν and ω to the list of parameters over which we were attempting to maximize the likelihood. In similar fashion, we added μ and σ to the list of parameters for the truncated normal distribution. In the case of the beta distributions, we limited the search in the region $1 \le \nu, \omega \le 5$, since values of ν and ω between 0 and 1 create singularites at the boundaries, an impossibility considering the nature of EF. In the case of the truncated normal distributions we limited the search in the region $0 \le \mu \le 1$ and $0.1 \le \sigma \le 10$. We began by choosing the form of the assumed distribution and the true distribution to be the same, i.e., we estimated the parameters of the beta distribution while using beta distributed data. We found that the method successfully approximated the values of all parameters, including those on the assumed distribution, as displayed in Table 3.3. The results of an individual experiment are displayed graphically in Figs. 3.2 and 3.3.

In the previous experiment the estimated parameters associated with both the beta and

(a)

(b)

(c)
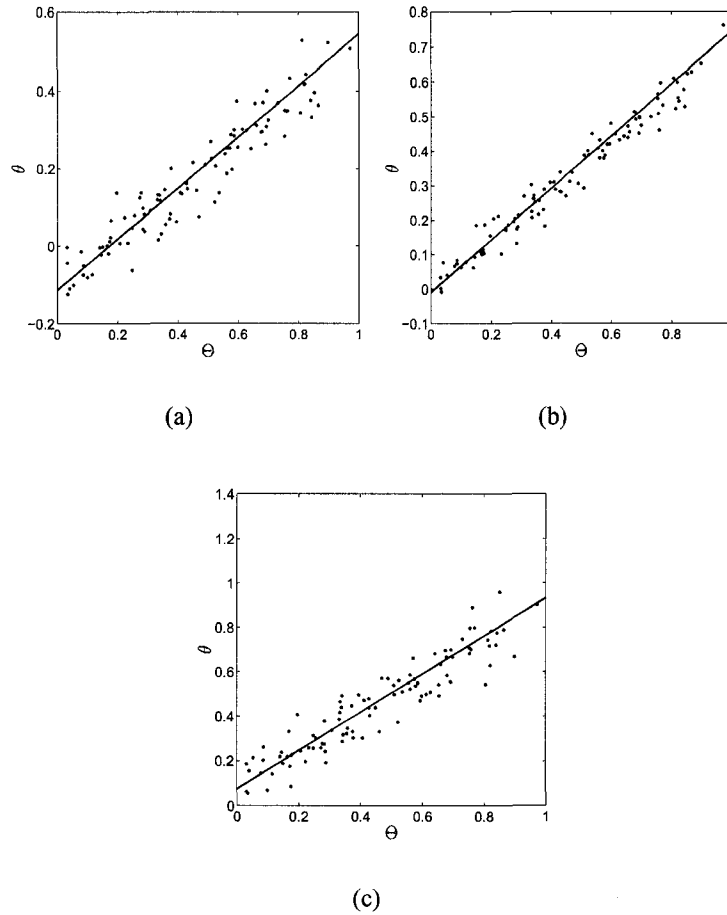
FIGURE 3.2. The results of an experiment using 100 patients, 3 modalities, and the same true parameters as shown in Table 3.3. In each graph we have plotted the true ejection fraction against the estimates of the EF for three different modalities ((a), (b) and (c)). The solid line was generated using the estimated linear model parameters for each modality. The dashed lines denote the estimated standard deviations for each modality. The estimated a_m, b_m and σ_m for each graph are (a)0.66,-0.11,0.050, (b)0.75,0.01,0.035 and (c)0.86,0.07,0.073. Note in this study the parameters of the beta distribution were estimated along with the linear model parameters.
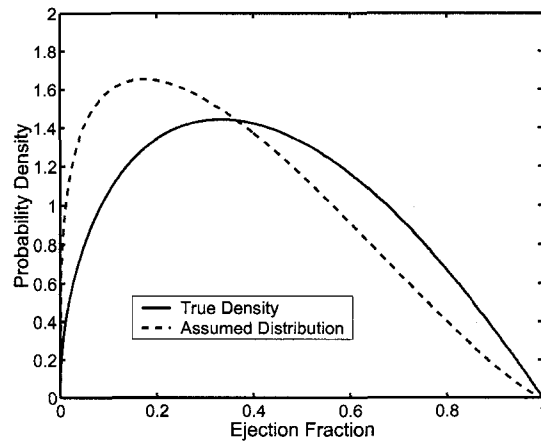
(a)
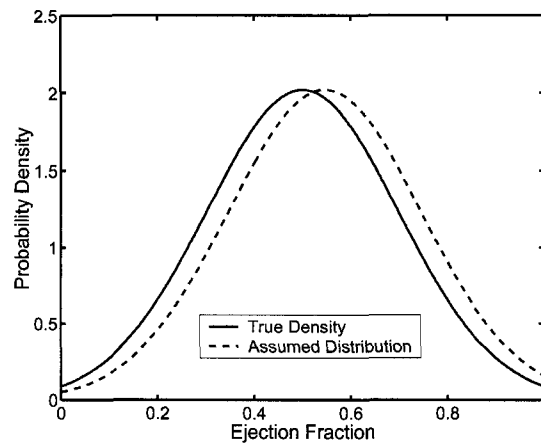
(b)

FIGURE 3.3. When the form of the assumed distribution matches the true distribution we see that RWT results in estimates of the assumed distribution which closely resemble the true distribution. In (a), the true distribution is a beta with parameters $\nu = 1.5$ and $\omega = 2.0$ that have been estimated to be $\nu = 1.27$ and $\omega = 2.29$. In (b), the true distribution is a truncated normal with parameters $\mu = 0.5$ and $\sigma = 0.25$ that have been estimated to be $\mu = 0.55$ and $\sigma = 0.20$.

TABLE 3.3. Values of estimated linear model and distribution parameters with the assumed distribution and the fixed true distribution having the same form.

	a_1	a_2	a_3
True Values	0.6	0.7	0.8
$pr(\Theta)$=Normal	0.59±.03	0.69±.04	0.79±.04
$pr(\Theta)$=Beta	0.60±.09	0.70±.09	0.79±.11
	b_1	b_2	b_3
True Values	-0.1	0.0	0.1
$pr(\Theta)$=Normal	-0.09±.03	0.01±.03	0.11±.04
$pr(\Theta)$=Beta	-0.10±.03	0.01±.03	0.11±.04
	σ_1	σ_2	σ_3
True Values	0.05	0.03	0.08
$pr(\Theta)$=Normal	0.050±.002	0.029±.004	0.080±.003
$pr(\Theta)$=Beta	0.048±.006	0.030±.011	0.080±.006
	Distribution	Parameters	
True Values	$\mu = 0.5, \nu = 1.5$	$\sigma = 0.2, \omega = 2.0$	
$pr(\Theta)$=Normal	$\mu = 0.50±.03$	$\sigma = 0.20±.02$	
$pr(\Theta)$=Beta	$\nu = 1.50±.53$	$\omega = 2.08±.99$	

truncated normal distributions were very close to their true values. We now show the results when the assumed distribution differs from the true distribution in Table 3.4. We know from our previous experiment that when the form of the assumed and true distributions match, the correct distribution parameters are estimated (on average). However, it remains to be seen what distribution parameters will be estimated when the forms of the two distributions differ. Thus in Fig. 3.4 we display the true distribution as well as the assumed distribution with the mean estimates of the distribution parameters. Note that the assumed distribution cannot equal the true distribution because they are from two different distribution families, *i.e.* beta and truncated normal. The assumed distribution does, however, take on a form which approximates the true distribution in an attempt to maximize the likelihood.
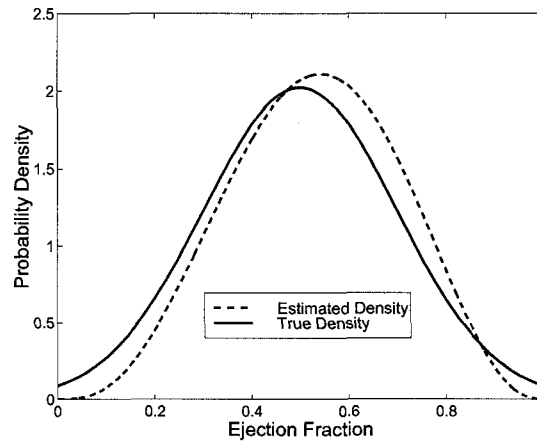
(a)
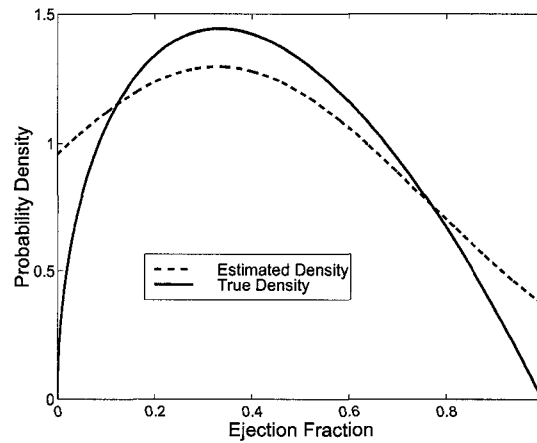
(b)

FIGURE 3.4. When the form of the assumed distribution does not match that of the true distribution, we see that the optimal distribution parameters are such that the form of the assumed distribution approximates the true distribution. In (a), the true distribution is a truncated normal which is approximated automatically by the method using a beta distribution ($\nu = 3.93$, $\omega = 3.47$). In (b), the roles are reversed, as a truncated normal automatically approximates a beta distribution ($\mu = 0.33$, $\sigma = 0.42$).

TABLE 3.4. Values of estimated linear model parameters using different forms of the varying assumed distribution and the fixed true distribution.

	a_1	a_2	a_3
True Values	0.6	0.7	0.8
$pr_a(\Theta)$=Normal/$pr_t(\Theta)$=Beta	0.56±.04	0.65±.05	0.74±.06
$pr_a(\Theta)$=Beta/$pr_t(\Theta)$=Normal	0.66±.10	0.78±.09	0.89±.12
	b_1	b_2	b_3
True Values	-0.1	0.0	0.1
$pr_a(\Theta)$=Normal/$pr_t(\Theta)$=Beta	-0.09±.02	0.01±.02	0.12±.03
$pr_a(\Theta)$=Beta/$pr_t(\Theta)$=Normal	-0.14±.06	-0.06±.06	0.03±.07
	σ_1	σ_2	σ_3
True Values	0.05	0.03	0.08
$pr_a(\Theta)$=Normal/$pr_t(\Theta)$=Beta	0.050±.005	0.029±.004	0.080±.007
$pr_a(\Theta)$=Beta/$pr_t(\Theta)$=Normal	0.050±.007	0.025±.011	0.079±.009

3.3 Linear Mixed-Effects Models

From a statistical model viewpoint, RWT is a type of latent-variable analysis which falls into the category of general linear mixed-effects model [50–53]. A mixed-effects model is one which consists of both random and fixed effects. Random effects are stochastic terms in a model while fixed effects are terms whose value remains constant. Let us revisit our model from equ. 3.1,

$$\theta_{pm} = a_m \Theta_p + b_m + \epsilon_{pm}.$$

The a_m's and b_m's are considered fixed effects, Θ_p is a random effect, and the ϵ_{pm} are the noise terms often referred to as the residual components.

Linear mixed-effects models have been used in the past to address the no-gold-standard problem for classification tasks. [15–20]. RWT is the first attempt at solving the no-gold-standard problem for estimation tasks.

3.4 Discussion

We have developed a method for characterizing an observer's performance in estimation tasks without the use of a gold standard. Although a gold standard is not required for this method, it is necessary to make some assumptions on the distribution of the parameter of interest (*i.e.*, EF). We have found that when the assumed distribution matches the true distribution, the estimates of the linear model parameters are both accurate and precise. Conversely, when the assumed and true distributions do not match, we find that our linear model parameters are no longer as accurate. This led us to investigate the role of the assumed distribution in the accuracy of the linear model parameters. By optimizing both the distribution parameters and the model parameters we found that one can effectively find both the model parameters and the form of the assumed distribution.

3.4.1 Assumptions

The key advantage of RWT over conventional regression analysis is that it does not require a gold standard. The performance of RWT, however, is hindered by this lack of information. Furthermore, like conventional regression analysis, RWT assumes a known functional form for the relationship between the gold standard and the data. However, unlike conventional regression analysis, this relationship cannot be visually assessed without the gold standard. We must also assume a functional form of the gold standard density $pr(\Theta)$, but there are parameters characterizing the shape of this density which are free to vary in RWT. We have assumed a Gaussian noise model, which is also implicit in conventional regression analysis, but other noise models are easy to implement in the likelihood expression.

A major underlying assumption of the method proposed here is that the true parameter of interest does not vary according to modality. This assumption may not be accurate in the context of estimating EF, which may vary moment to moment with a patient's mood and/or breathing pattern. This assumption may be valid, however, for other estimation tasks. Another assumption we have made is the linear relationship between the true and

$$C_{a \times a} = \begin{bmatrix} 1 & 0.96 & 0.92 \\ 0.96 & 1 & 0.94 \\ 0.92 & 0.94 & 1 \end{bmatrix} \quad C_{b \times b} = \begin{bmatrix} 1 & 0.90 & 0.85 \\ 0.90 & 1 & 0.89 \\ 0.85 & 0.89 & 1 \end{bmatrix}$$

$$\text{(a)} \qquad\qquad\qquad\qquad \text{(b)}$$

$$C_{\sigma \times \sigma} = \begin{bmatrix} 1 & -0.58 & 0.16 \\ -0.58 & 1 & -0.37 \\ 0.16 & -0.37 & 1 \end{bmatrix}$$

$$\text{(c)}$$

FIGURE 3.5. Correlation matrices for the bias of the estimates of the linear model parameters estimated by RWT using 100 runs with three modalities and 100 patients. Note that the bias is highly correlated for the slopes (a) and intercepts (b) and only mildly correlated for the σ terms (c).

estimated parameters of interest. This linear relationship was chosen in large part due to mathematical simplicity, but is, nonetheless, a good first step. More complicated, nonlinear models can easily be accommodated by this method and are discussed briefly in the next chapter. Ideally, we would like to choose a model based on some sort of physical knowledge of the estimation technique.

3.4.2 Using RWT

An estimator of a medically relevant parameter should be both accurate and precise. For the linear models discussed above, accuracy can be approximately achieved by adjusting the measurements using the estimated model parameters $\hat{a}_m$ and $\hat{b}_m$. Namely, we can rewrite equ. 3.1 as

$$\Theta_p = \frac{\theta_{pm} - \hat{b}_m}{\hat{a}_m} - \frac{\epsilon_{pm}}{\hat{a}_m}. \tag{3.13}$$

After this correction the variance in the adjusted measurements, *i.e.*, the precision, is $\sigma_m^2/\hat{a}_m^2$. An estimate of this quantity $\hat{\sigma}_m^2/\hat{a}_m^2$ can be used as a figure of merit for cross-modality comparisons.

When comparing different imaging modalities one would typically prefer the modality with the smallest error, *i.e.*, the smallest σ_m/a_m. Estimating σ_m/a_m facilitates modality comparisons without knowledge of a gold standard. As discussed earlier, the estimates of the slopes a_m's retained the proper ordering amongst modalities even when a bias is introduced by mismatching true and assumed distributions. In Fig. 3.5 we present correlation matrices for the bias of the linear model parameters estimated using RWT. We generated these matrices using estimates of the linear model parameters for 100 runs of RWT with three modalities and 100 patients. The estimates of σ_m were very accurate regardless of the choice of the assumed and true distributions. Combining these observations we feel confident that σ_m/a_m will serve as a good figure of merit to compare imaging systems even when the data do not match the model.

The estimates of the slope and intercept values describe the systematic error (or bias) of the modality. If one were confident in these estimates they could be employed to adjust and correct systematic error for each modality. Another interesting result of the experiments is the successful estimation of the distribution parameters to fit the form of the true distribution. This could serve as an insight into the distribution of the true parameter for the population studied, *i.e.*, the patient distribution of EFs. We will further discuss applications of RWT in Chap. 7.

Chapter 4

RESULTS OF SIMULATIONS USING RWT

In Chapter 3 we presented our method RWT along with an analysis of the bias and variance of the method using simulation studies. More specifically, we evaluated how well RWT could estimate the linear model parameters and the parameters of the assumed distribution for a given number of patients and modalities. In this chapter we analyze the performance of RWT with varying amounts of patients, modalities, and noise. Also in this chapter we investigate the impact of relaxing the linear model assumption. We conclude the chapter with a discussion of our findings.

4.1 Figures of Merit

The figure of merit in linear regression analysis is root mean-squared error (RMSE). When comparing estimation tasks with an available gold standard, RMSE is an estimate of the parameter σ_m. Without the x coordinates we are unable to calculate RMSE. We can, however, use the results of RWT to derive a comparable figure of merit to characterize the performance of a single application of RWT. RMSE for a given modality m is as follows,

$$\text{RMSE}_m = \sqrt{\frac{1}{P} \sum_{p=1}^{P} \left(\Theta_p - \frac{\theta_{pm} - \hat{b}_m}{\hat{a}_m} \right)^2}, \tag{4.1}$$

where $\hat{a}_m$ and $\hat{b}_m$ are estimates of the linear model parameters returned by RWT. We choose this figure of merit as it measures the difference between the gold standard (Θ_p) and the values found by adjusting the data (θ_{pm}) by the estimated linear model parameters.

We would like to note that this figure of merit cannot be used in practice due to a lack of a gold standard, but it provides an excellent technique to evaluate the method in simulation. In our results section we perform 50 simulations and average RMSE_m (denoted by $\overline{\text{RMSE}_m}$) as well as compute the standard error.

4.2 Simulations

In Chap. 2 we focused on estimating the linear model parameters for a fixed number of patients and modalities. We showed that RWT was accurate even when the true and assumed distributions were from different families. In this section we investigate the performance of RWT with varying number of patients, number of modalities, and noise. Furthermore, we compare the performance of RWT to conventional linear regression analysis under the unlikely assumption that a gold standard is available.

4.2.1 Implementation

The likelihood function was implemented and optimized on an 800 MHz Pentium III using Matlab. We used a quasi-Newton optimization method from Matlab to determine the maximum of the likelihood. We constrained the optimization to look for reasonable values of the parameters, *i.e.*, positive slopes and positive variances. We fixed the initial guess as the midpoint of the search space, a point *not* equal to the true values of the parameters. With the above constraints, the results of the optimization were not sensitive to the initial guess. The optimization took from a few seconds to a few minutes to run depending on the form of the assumed distribution used in the likelihood expression.

We performed numerous simulation studies in which we sampled cardiac ejection fractions (the gold standard) for a simulated patient population from a beta distribution with fixed parameters, *i.e.*, $pr_t(\Theta)$ was a beta PDF. We then adjusted this gold standard using linear models with known parameters a_m, b_m as well as a known noise level characterized by σ_m. This comprised the data that was input to RWT. The gold standard values *were not* input into the method. In computing the likelihood function, we not only need the data but we must also assume a functional form for the gold-standard density. We assumed a truncated normal distribution with varying mean and variance, *i.e.*, $pr_a(\Theta)$ was a truncated normal density characterized by μ_a and σ_a. Note that this distribution is *different* from the distribution that was actually used to generate the gold standard. This difference is meant

to simulate the real-world situation in which one would not know exactly how the gold standard was distributed.

Both the beta and the truncated normal distributions are bounded between zero and one. In this chapter we study the performance of RWT only with these bounded distributions. Difficulties that arise when extending the RWT method to distributions that span the entire real line will be discussed in Chap. 5.

In Fig. 4.1(a) we show that the $\overline{\text{RMSE}}$, as given in equ. 4.1, decreases as patient number increases. The variance of the noise σ_m was fixed for each modality in this experiment. In the limit of large patient numbers the three different curves (each representing a different modality), tend to a minimum value σ_m/a_m (See equ. 3.1 and equ. 4.1), in accordance with ML theory. Figure 4.1(b) compares the performance of conventional regression analysis to that of RWT. As one would expect, conventional regression analysis using the gold standard outperforms RWT. The difference between the two decreases as a function of patient population size.

Another, slightly different approach, to studying the relationship between bias and the number of patents is plotting the true and estimated linear model parameters as a function of the number of patients. In Fig. 4.3, we present the results of an experiment consisting of averaging the results of RWT over 150 runs using three modalities. We see from the figure that RWT is asymptotically unbiased, as predicted for an ML estimator.

It is not surprising that an increase in data yields more accurate results. An increase in the number of modalities is a somewhat less intuitive notion given the complexity of our ML estimator. In Fig. 4.2 we display a plot of $\overline{\text{RMSE}}$ versus number of modalities. One sees that after a few modalities the gain in accuracy is not substantial. Note that the performance of conventional linear regression analysis is independent of the number of modalities. The performance of RWT with one modality is very poor but relatively constant with two or more modalities.

Lastly, we look at the impact on $\overline{\text{RMSE}}$ of varying the parameter σ_m to understand what occurs to the accuracy as the noise in the data increases. The curves in Fig. 4.4(a) show
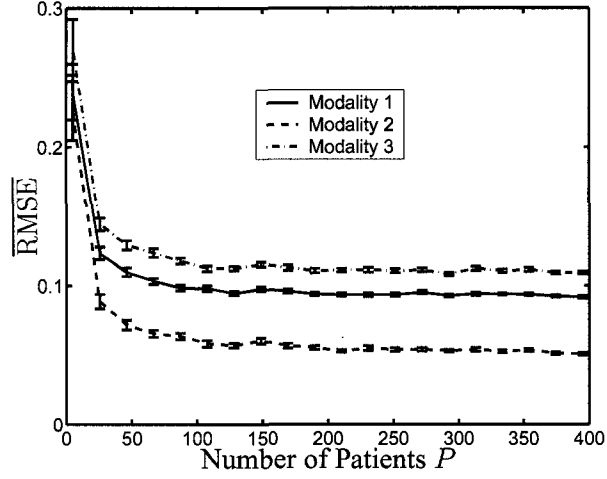
(a)
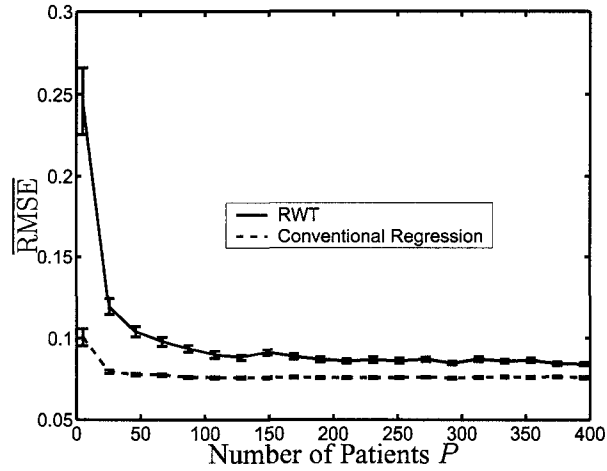
(b)

FIGURE 4.1. In (a), the $\overline{\text{RMSE}}$ for three different modalities is plotted versus the number of patients. As patients increase, RMSE_m converges to σ_m/a_m, by equ. 3.1 and equ. 4.1. In (b) a comparison is made between RWT and linear regression analysis with a gold standard. Note that the RMSE in (b) is also averaged over the three modalities. As expected, conventional regression analysis has lower $\overline{\text{RMSE}}$, but the performances of the two methods converge as the number of patients increases. For these experiments $\vec{a} = [0.6, 0.7, 0.8]$, $\vec{b} = [-0.1, 0.0, 0.1]$, $\vec{\sigma} = [0.05, 0.03, 0.08]$, and the error bars represent the standard error calculated over 50 independent experiments.
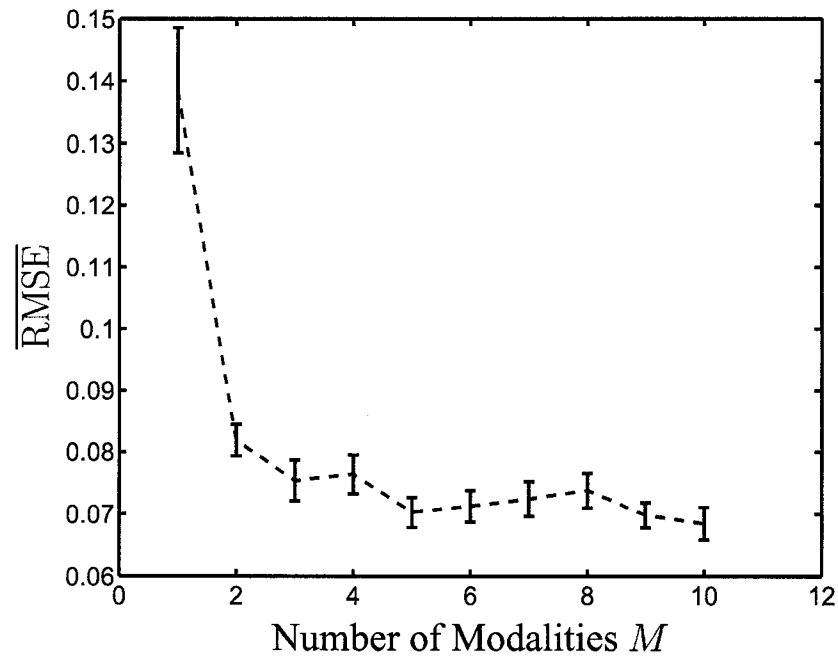
FIGURE 4.2. The above plot displays $\overline{\text{RMSE}}$ (averaged across simulations and modalities) versus the number of modalities used in a RWT experiment. There exists a sharp decline in $\overline{\text{RMSE}}$ from one to two modalities followed by a slow decline. One might expect this, especially since RWT cannot work properly with only one modality. The performance of conventional regression analysis is independent of the number of modalities. The same model parameters were used for all modalities in all experiments ($a_m = 1$, $b_m = 0.1$, $\sigma_m = 0.05$, $P = 100$)

(a)

(b)

(c)
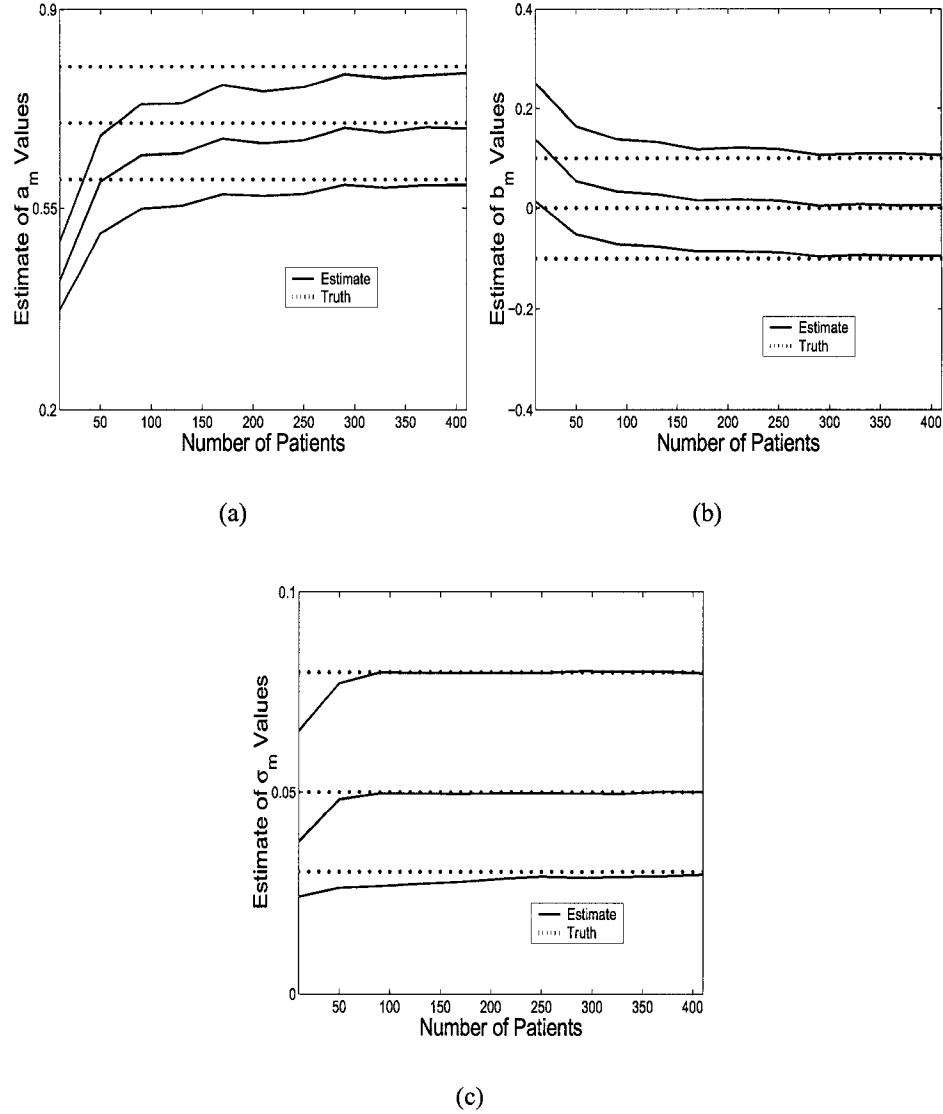
FIGURE 4.3. Plotting estimated and true linear model parameters as a function of the number of patients. The experiment consists of averaging the results of RWT over 150 runs using three modalities. Each plot has three dotted lines representing the true values and three solid lines representing the estimates. Notice that all estimates asymptote to the truth. In this experiment, the assumed and true distribution were both from the truncated normal family.
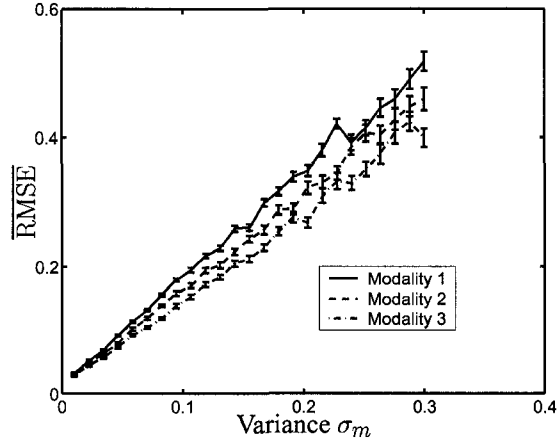
(a)
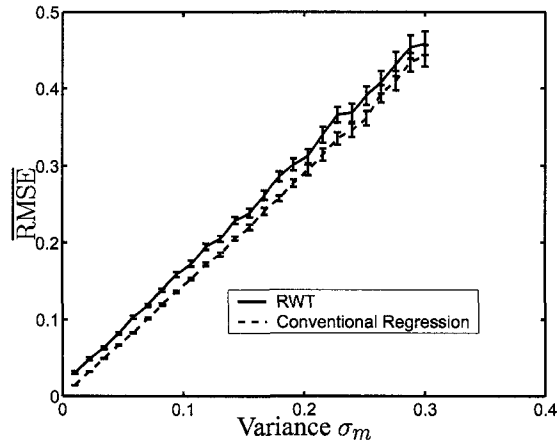
(b)

FIGURE 4.4. In (a), the $\overline{\text{RMSE}}$ for three different modalities is plotted versus variance of the noise σ_m. $\overline{\text{RMSE}}$ increases in accordance with $1/a_m$, by equ. 3.1 and equ. 4.1. In (b) a comparison is made between RWT and linear regression analysis with a gold standard. Note that the RMSE in (b) is also averaged over the three modalities. $\overline{\text{RMSE}}$ does not converge to zero for RWT as σ_m tends to zero. The parallel nature of the two graphs indicates that the comparative performance of RWT is independent of σ_m. For these experiments $\vec{a} = [0.6, 0.7, 0.8]$, $\vec{b} = [-0.1, 0.0, 0.1]$, $P = 100$, and the error bars represent the standard error calculated over 50 independent experiments.

that $\overline{\text{RMSE}}$ increases linearly with increases in σ_m. The slope of these lines are given by $1/a_m$ as predicted from equ. 3.1 and equ. 4.1. Figure 4.4(b) compares the performance of conventional regression analysis with that of RWT. We see that while the $\overline{\text{RMSE}}$ limits to zero as $\sigma_m \rightarrow 0$ for conventional regression analysis, RWT limits to a positive constant. This positive constant is a result of bias introduced by using different true and assumed distribution. Namely, in this simulation we sampled from a beta distribution and used a truncated normal distribution in our likelihood expression. The constant difference between the two plots in Fig. 4.4(b) indicates the independent relationship between the variance of the noise and the comparative performance of RWT and conventional regression analysis.

4.3 Nonlinear models

A clear limitation of what we have presented thus far is the strict assumption of a linear model governing the relationship between the gold standard and the individual modalities. To ease this assumption one can rewrite equ. 3.1 as,

$$\theta_{pm} = \mathcal{N}(\Theta_p; \vec{\nu}_m) + \epsilon_{pm}, \tag{4.2}$$

where $\mathcal{N}(\cdot)$ is some nonlinear function of the gold standard with model parameters $\vec{\nu}_m$.

In Fig. 4.4 we show the results of a single experiment using three modalities with a quadratic model for each modality. In modality one, Fig. 4.4(a), there is a nonlinear relationship between the gold standard and the estimate. Modality two, Fig. 4.4(b), has a weak nonlinear relationship. Finally, modality three, Fig. 4.4(c), is linear. RWT accurately fits all three modalities. The time required for the optimization procedure to converge is increased by the added parameters to be estimated. It should also be noted that, with too many parameters, regression analysis will eventually fit the noise in the data. While we have shown that the method can be extended to nonlinear models, there is extensive work that needs to be completed with the linear models before we can fully characterize the performance of this technique using nonlinear models.
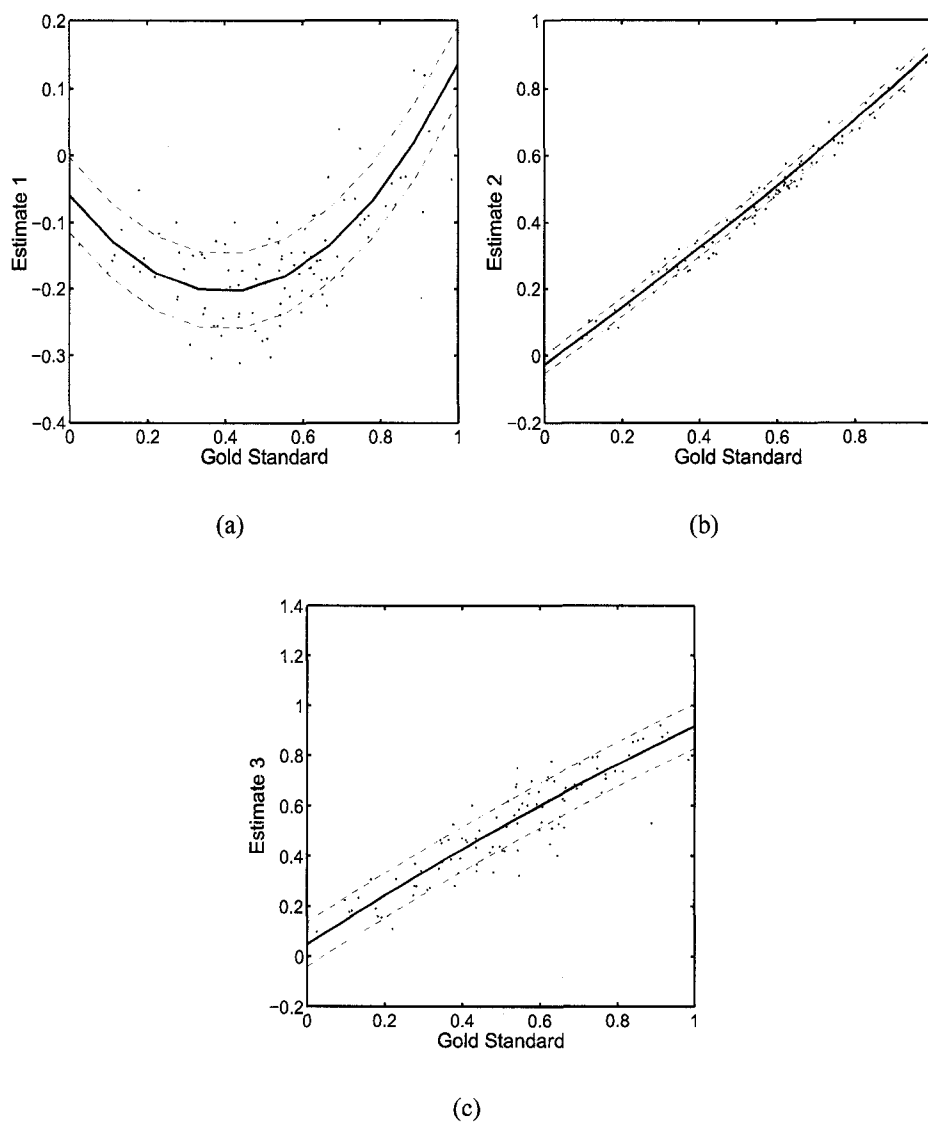
FIGURE 4.5. An application of RWT with a quadratic model. For modality one there existed a strong nonlinear relationship with the gold standard and a relatively large variance which qualitatively were discovered. Modality two was slightly nonlinear with small variance, while modality three was linear with large variance. Both were fit well by the quadratic RWT.

4.4 Discussion

In Chap. 3 we studied the bias and variance of our linear model parameter estimates for a fixed number of patients and modalities. In this chapter we analyzed the performance of RWT through numerical simulation with varying amounts of patients, modalities, and noise using RMSE as a figure of merit. The purpose for doing so was to analyze the robustness of our method with respect to such quantities. We were able to use RMSE as a figure of merit because all studies were done in simulation and the gold standard was thus known. RMSE is *not* a viable figure of merit for comparing estimation techniques when a gold standard is not available. Furthermore, since the gold standard was available we were able to compare the performance of RWT with regression analysis with the truth or x axis.

The results of the simulation studies performed agreed with predictions from the theoretical analysis. Namely, as the number of patients was increased, the RMSE decreased rapidly and approached the performance of conventional regression analysis. We also found that RMSE declines as additional modalities are added to the analysis, though the advantage is not very significant after a few modalities. We return to this point in the next chapter. Another intuitive result was the linear increase in RMSE as a function of the modality noise parameter σ_m. Finally, in Sec. 4.3 we address the potential for a nonlinear model relating the gold standard to the estimate. We give the results of a simulation using a quadratic model that displays the effectiveness of RWT even in the nonlinear case.

Chapter 5

CALCULATING THE FISHER INFORMATION FOR RWT

In previous chapters we have presented a method for evaluating and comparing multiple estimation methods without the benefit of a gold standard. We have named it Regression Without Truth (RWT). We have evaluated the robustness of the method through numerous simulation studies. This chapter attempts to use Fisher information to better evaluate the performance of RWT.

As discussed, if a gold standard is available, evaluating an estimation method consists of analyzing scatter plots of data along with the fitted model, *e.g.*, estimated EF versus true EF. With our method, however, this type of analysis is not possible because we do not use the gold standard and assume it is unavailable. Thus we are unsure how well the method fit the model parameters, and we must resort to other means to assess the performance and reliability of RWT.

In this chapter we derive an expression for the Fisher information matrix for RWT, from which we ascertain the Cramér-Rao (CR) or lower bound for the variances of the estimated model parameters [30–32]. We also show that the variances calculated experimentally are very close to the CR bounds calculated using the Fisher information. In an attempt to further understand our method we also derive the Fisher information matrix for standard regression analysis and compare the resultant CR bound with that of RWT. Also in this chapter, we discuss the importance of having a bounded assumed distribution. We conclude the chapter with a discussion of our findings.

5.1 Vector Notation

Let us first address some notational issues for the sake of convenience. We began by assuming a linear relationship between the true (Θ_p) and estimated (θ_{pm}) EF for a given patient p

measured on modality m. We represent the relationship as

$$\theta_{pm} = a_m \Theta_p + b_m + \epsilon_{pm}, \tag{5.1}$$

where a_m and b_m are the linear model parameters and ϵ_{pm} is the noise term. For this chapter we would like to work with vector notation, therefore we write equ. 5.1 using the measured EFs for a given patient p from M different modalities as

$$\boldsymbol{\theta}_p = \boldsymbol{a}\Theta_p + \boldsymbol{b} + \boldsymbol{\epsilon}_p, \tag{5.2}$$

where $\boldsymbol{\theta}_p$ is an $M \times 1$ vector consisting of the measured EFs, $\boldsymbol{a}$ and $\boldsymbol{b}$ are $M \times 1$ vectors made up of the linear model parameters, and $\boldsymbol{\epsilon}_p$ is an $M \times 1$ vector whose elements are the noise terms.

We model the true distribution $pr_t(\Theta_p)$ with an assumed distribution, $pr_a(\Theta_p|\boldsymbol{r})$, which is parameterized by $\boldsymbol{r}$. The parameters $\boldsymbol{r}$ are added to the list of unknown parameters, allowing us to write a likelihood over all patients as

$$\Lambda = \prod_{p=1}^{P} pr(\boldsymbol{\theta}_p | \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{r}). \tag{5.3}$$

The log-likelihood is given by

$$\lambda = \ln(\Lambda) = -\frac{P}{2} \sum_{m=1}^{M} \ln\left(2\pi\sigma_m^2\right) +$$

$$\sum_{p=1}^{P} \ln\left[\int d\Theta_p \, pr_a(\Theta_p|\boldsymbol{r}) \exp\left(\sum_{m=1}^{M}(-\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2)\right)\right]. \tag{5.4}$$

5.2   Calculating Fisher information for RWT

Using the expression for the log-likelihood in equ. 5.4 we are able to calculate the Fisher information matrix for our estimator. The parameters we are estimating can be represented by the vector,

$$\boldsymbol{\eta} = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{r}\}, \tag{5.5}$$

where η consists of the linear model parameters as well as the parameters of our assumed distribution. For the example presented in Sec. 3.1, η consisted of 11 parameters; 3 linear model parameters from each of the 3 modalities along with the parameters ν and ω characterizing the beta density in the likelihood expression. Notice that the true EF is not in this list of parameters. As discussed in Chap. 3, we are *not* estimating the true EF. We are performing ML estimation and using *measured* EF as our data to estimate the linear model parameters and the parameters of the assumed distribution.

The likelihood Λ and the log-likelihood λ are functions of the parameters η, represented as

$$\lambda(\boldsymbol{\eta}|\{\boldsymbol{\theta}_p\}) = \ln[\Lambda(\boldsymbol{\eta}|\{\boldsymbol{\theta}_p\})] = \ln[pr(\{\boldsymbol{\theta}_p\}|\boldsymbol{\eta})]. \tag{5.6}$$

Using this notation, we write the ij^{th} component of the Fisher information matrix as

$$J_{ij} = \left\langle \frac{\partial\lambda(\boldsymbol{\eta}|\{\boldsymbol{\theta}_p\})}{\partial\eta_i} \frac{\partial\lambda(\boldsymbol{\eta}|\{\boldsymbol{\theta}_p\})}{\partial\eta_j} \right\rangle_{\{\boldsymbol{\theta}_p\}} \Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}_o}, \tag{5.7}$$

where $\boldsymbol{\eta}_o$ is the true value of $\boldsymbol{\eta}$ [30,31]. The expectation in the above equation is taken with respect to the density on the data at $\boldsymbol{\eta}_o$. Using the expression for the log-likelihood given in equ. 5.4, we are able to rewrite equ. 5.7 as

$$J_{ij} = P\int_R d^M\theta_p \left(pr(\boldsymbol{\theta}_p|\boldsymbol{\eta}) \frac{\partial \ln pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})}{\partial\eta_i} \frac{\partial \ln pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})}{\partial\eta_j} \right) \Bigg|_{\boldsymbol{\eta}=\boldsymbol{\eta}_o}, \tag{5.8}$$

where R is the range of θ_p. The diagonal element J_{ii}^{-1} is the CR bound for the corresponding parameter η_i.

Although J cannot be computed analytically for most choices of $pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})$, it can be readily approximated using Monte Carlo methods. The two partial derivatives in the above expression are also, in general, derivatives of definite integrals (see appendix B for examples). Both integral expressions, however, contain density functions that we can sample from and thus perform Monte Carlo integration. This integration was accomplished in practice by sampling 10,000 observations from $pr(\boldsymbol{\theta}_p|\boldsymbol{\eta}_o)$, and for each of these observations sampling another 10,000 obserervations from $pr(\Theta_p|\boldsymbol{r}_o)$ to compute the two partial

derivatives. This amounts to more than 1×10^8 samples to compute the elements of J. Using this many samples resulted in a reproducibility of 3%. Calculations were made using Matlab on a Pentium IV, 1.4 GHz machine running Linux. We show the results of these calculations in the next section.

5.3 Results

5.3.1 Evaluation of method

ML estimators are asymptotically efficient since they achieve the CR bound as the number of samples tends to infinity [54]. Equality holds in the CR bound irrespective of sample size if and only if the likelihood is a member of a dim(η)-parameter exponential family [54–56]. We are not able to show analytically that our likelihood is a member of an exponential family. Therefore, we are forced to rely on empirical studies. Furthermore, because large data sets are often difficult to obtain in medical imaging, we are inclined to evaluate the "efficiency" of RWT with small sample sizes (*i.e.*, few patients). To accomplish this evaluation, we calculate the sample variance of the estimated parameters for varying numbers of patients and compare these results to the CR lower bound of the variance as computed using eqn. 5.8.

We ran simulation studies in which we varied the number of patients P from 5 to 400 and measured the sample standard deviation of the estimated parameters using 100 independent runs of RWT. There were 3 modalities for each run, and the true gold standard was sampled from a truncated normal distribution. Furthermore, the assumed distribution used in the method was also a truncated normal distribution. In the example presented in Sec. 3.1 the true and assumed distributions were from different families. In our calculation of the Fisher information for RWT, we assume the model matches the data.

In Fig. 5.1 we plot the sample standard deviation along with the CR lower bound of the linear model parameter a_1 versus number of patients. Similar graphs were obtained for the other elements of a. In Figs. 5.2 and 5.3 we show the analogous results for b_1 and σ_1
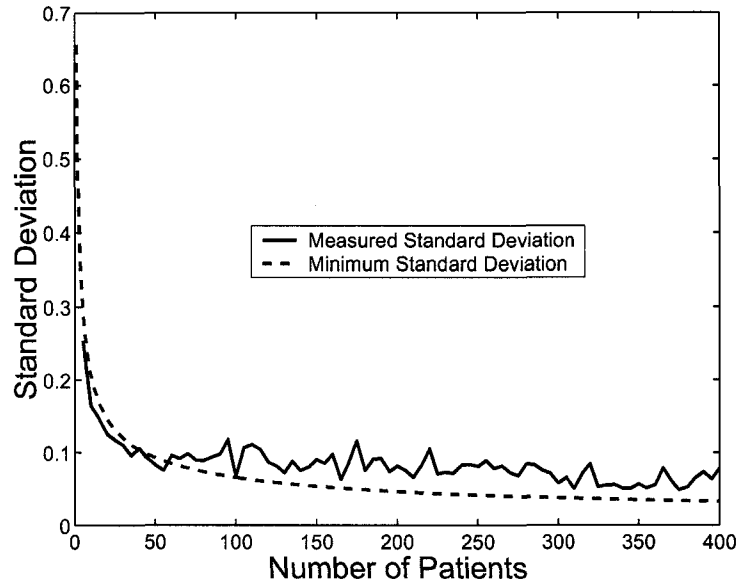
FIGURE 5.1. A comparison between the measured and minimum standard deviation for the estimated parameter a_1. The measured standard deviation was determined from 100 independent runs using three modalities. The measured standard deviation is lower than the minimum standard deviation for certain patient numbers due to sampling effects.

respectively. We see from these figures that even with small sample sizes our estimator is very close to the CR lower bound.

Although the CR bound will decrease with increasing numbers of patients P, the impact of the number of modalities on the CR bound is less obvious. In Fig. 5.4 we plot the CR bound as a function of patients for a varying number of modalities. As the number of modalities increases, the CR bound decreases. However, the added benefits of increasing the number of modalities is quickly saturated. In fact, the curves for 6 and 12 modalities are indistinguishable.

Thus far we have shown the effect that varying numbers of patients and modalities has on the CR bound for the various estimated parameters. A complex relationship exists between the true values of the parameters and the lower bound of the variance for the estimates of these parameters. This relationship is shown clearly in Fig. 5.5, where the true value of the slope for all three modalities was varied from 0 to 1.5 and the CR bound for
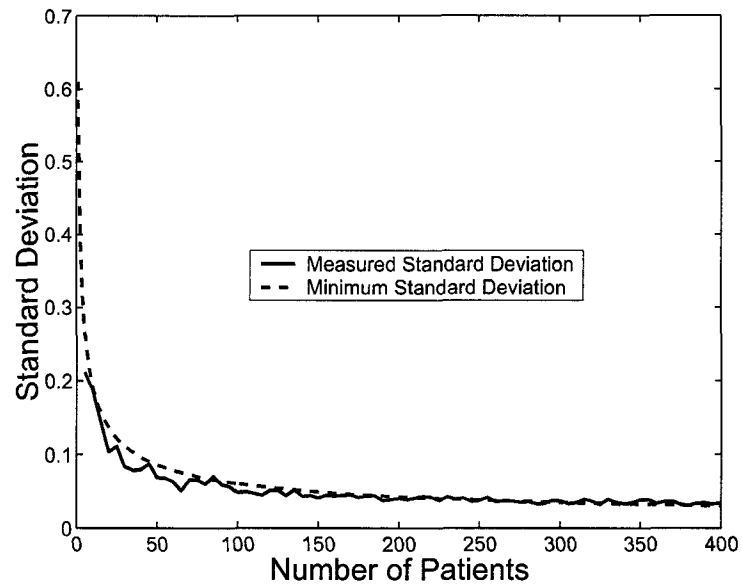
FIGURE 5.2. A comparison between the measured and minimum standard deviation for the estimated parameter b_1. The measured standard deviation was determined from 100 independent runs using three modalities.
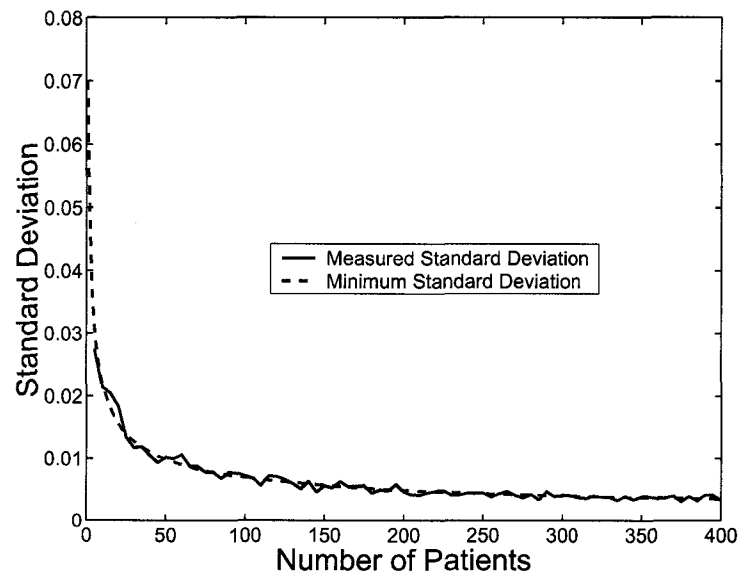
FIGURE 5.3. A comparison between the measured and minimum standard deviation for the estimated parameter σ_1. The measured standard deviation was determined from 100 independent runs using three modalities. Note the two curves are especially close.
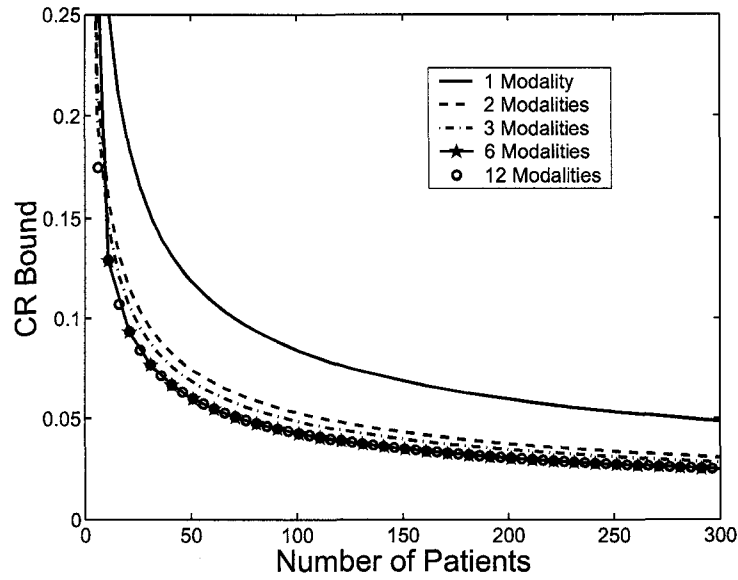
FIGURE 5.4. Comparing the CR bound of a_1 as a function of patients across a varying number of modalities. The linear model parameters were equal across all modalities.
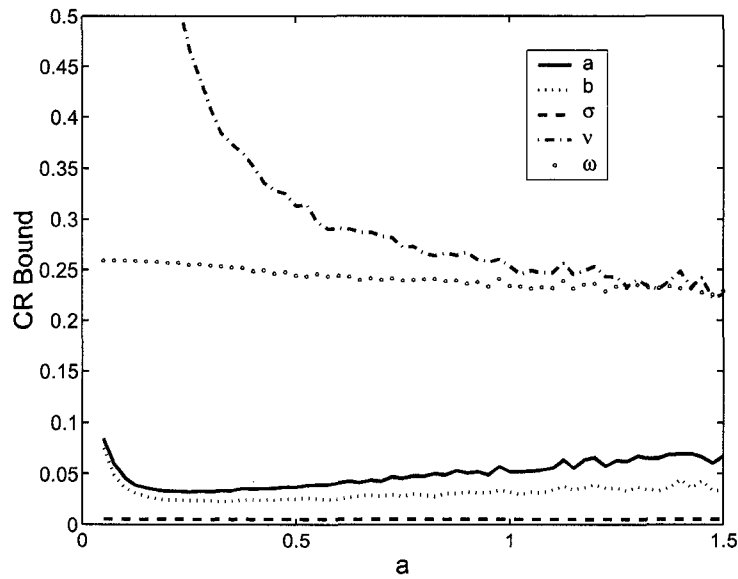
FIGURE 5.5. A comparison of the CR bounds in the beta case for the parameters a, b, σ, ν and ω as a function of a. Note that with the exception of ν the CR bound for each parameter varies slowly with a. The parameter ν tends to infinity as a limits to zero because less and less information is transmitted about the true density. Other than a, the true values of the parameters did not vary significantly from modality to modality.
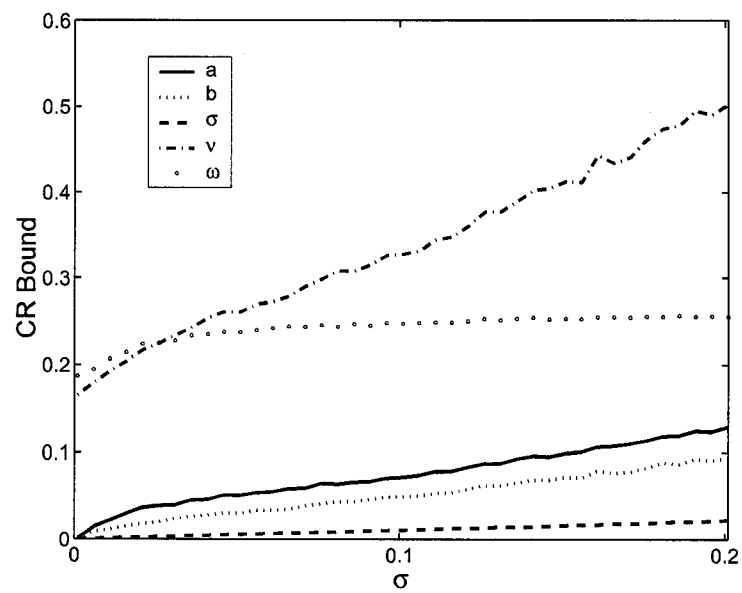
FIGURE 5.6. A comparison of the CR bounds in the beta case for the parameters a, b, σ, ν and ω as a function of σ. As the value of σ increases the noise grows larger and the other parameters become more difficult to estimate. Other than σ, the true values of the parameters did not vary from modality to modality.
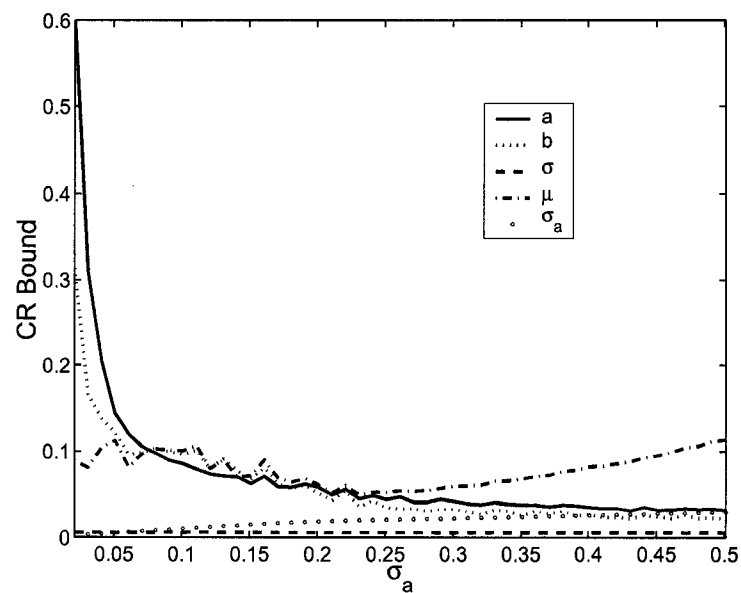
FIGURE 5.7. A comparison of the CR bounds in the truncated normal case for the parameters a, b, σ, μ and σ_a as a function of σ_a. The CR bound of the mean of the true distribution μ increases with σ_a corresponding to the "flattening" of the truncated normal density. The bounds for the parameters a and b tend to infinity as σ_a limits to zero, corresponding to the development of a sharp peak for the truncated normal.

each of the estimated parameters was plotted. With the exception of the beta parameter ν, the CR bounds appears to vary slowly with varying a. The CR bound on the parameter ν varies greatly due to the increase in information gained with increasing slope. For example, a slope of zero would give no information about the shape of the true distribution. The results obtained from varying σ_1 are shown in Fig. 5.6. The CR bounds for all of the estimated parameters are independent of the values of b.

In Fig. 5.7 we compare the CR bounds of the various parameters in the case when the assumed and true distributions are truncated normal distributions. More specifically, we compare the bounds as a function of the standard deviation σ_a of the assumed distribution $pr_a(\Theta_p|\{\mu, \sigma_a\})$. Increasing σ_a results in an increase of the width of the truncated normal density function. This in turn increases the difficulty of estimating the mean μ as shown in Fig. 5.7. As σ_a tends to zero, the CR bounds for a_1 and b_1 increase dramatically. This increase occurs because a small value of σ_t corresponds to a very tightly packed set of EFs sampled from $pr_t(\Theta)$, a data set for which even conventional regression analysis falters.

5.3.2 With truth

It is informative to calculate the Fisher information for the case with a gold standard and compare the CR bound to that of RWT. Calculating the Fisher information in the case when a gold standard exists is a much simpler problem. Since the Θ_p values are now known there is no need to marginalize, nor is it necessary to use multiple modalities to estimate the linear model parameters . The log-likelihood can be written as

$$\lambda(a, b, \sigma) = \sum_{p=1}^{P} ln\,[pr(\theta_p, \Theta_p | a, b, \sigma)] \tag{5.9}$$

$$= \sum_{p=1}^{P} ln\,[pr(\theta_p | a, b, \sigma, \Theta_p) pr(\Theta_p | \boldsymbol{\rho})] \tag{5.10}$$

where a, b, and σ are now scalars and the components of $\boldsymbol{\rho}$ characterize the density on Θ_p. Using this expression for λ and equ. 5.8, we are able to represent the Fisher information

matrix as

$$
J = \begin{bmatrix} \langle \frac{\partial \lambda}{\partial a} \frac{\partial \lambda}{\partial a} \rangle & \langle \frac{\partial \lambda}{\partial a} \frac{\partial \lambda}{\partial b} \rangle & \langle \frac{\partial \lambda}{\partial a} \frac{\partial \lambda}{\partial \sigma} \rangle \\[2mm] \langle \frac{\partial \lambda}{\partial b} \frac{\partial \lambda}{\partial a} \rangle & \langle \frac{\partial \lambda}{\partial b} \frac{\partial \lambda}{\partial b} \rangle & \langle \frac{\partial \lambda}{\partial b} \frac{\partial \lambda}{\partial \sigma} \rangle \\[2mm] \langle \frac{\partial \lambda}{\partial \sigma} \frac{\partial \lambda}{\partial a} \rangle & \langle \frac{\partial \lambda}{\partial \sigma} \frac{\partial \lambda}{\partial b} \rangle & \langle \frac{\partial \lambda}{\partial \sigma} \frac{\partial \lambda}{\partial \sigma} \rangle \end{bmatrix}, \tag{5.11}
$$

where the expectations $\langle \cdot \rangle$ are taken with respect to θ_p and Θ_p. Using equ. 5.10 and equ. 5.11 we write the Fisher information matrix as

$$
J = \frac{P}{\sigma^2} \begin{bmatrix} \langle \Theta^2 \rangle_{\{\Theta_p\}} & \langle \Theta \rangle_{\{\Theta_p\}} & 0 \\[2mm] \langle \Theta \rangle_{\{\Theta_p\}} & 1 & 0 \\[2mm] 0 & 0 & 2 \end{bmatrix}, \tag{5.12}
$$

where $\langle \Theta \rangle_{\{\Theta_p\}}$ is the expected value of the true EF data.

We compare the CR bounds of a_1, b_1, and σ_1 for RWT to those of conventional regression analysis with truth. In Figs. 5.8 and 5.9 we show the difference between the computed CR bounds in the cases with and without truth as a function of the number of patients for the linear model parameters a_1, b_1, and σ_1. Note that when we plot the CR bound we are plotting the lower bound of the standard deviation. The underlying true distributions $pr_t(\Theta_p)$ in Figs. 5.8 and 5.9 are beta and truncated normal, respectively. One can see that the CR bounds approach each other as the number of patients increases since all tend to zero as $1/\sqrt{P}$. Plots for the other components of a, b, and σ are very similar, and thus not shown here.

5.4 Normal assumed distribution

Thus far we have discussed calculating the Fisher information matrix for RWT numerically. In this section we discuss an analytic solution. Let us take a closer look at the likelihood expression given in eqn. 5.3. When $pr_a(\Theta_p|r)$ is normal over the real line, the likelihood
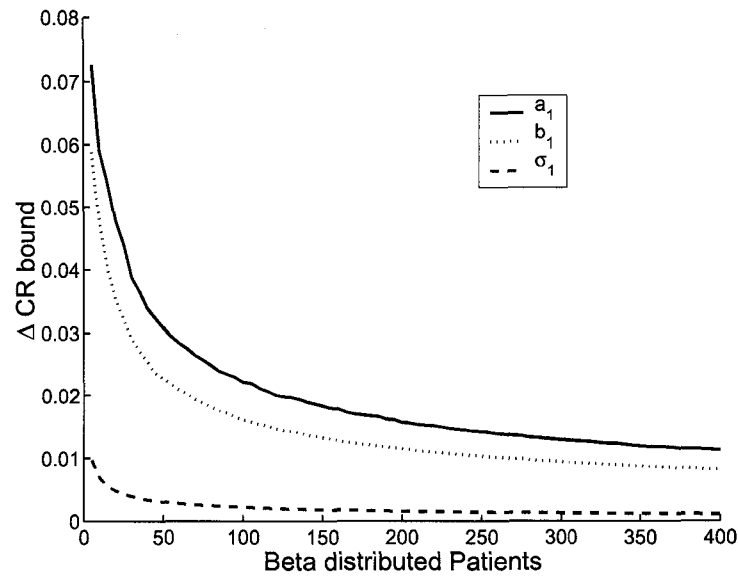
FIGURE 5.8. The difference of the CR bounds between our method and conventional regression analysis with truth when the true density is beta distributed.
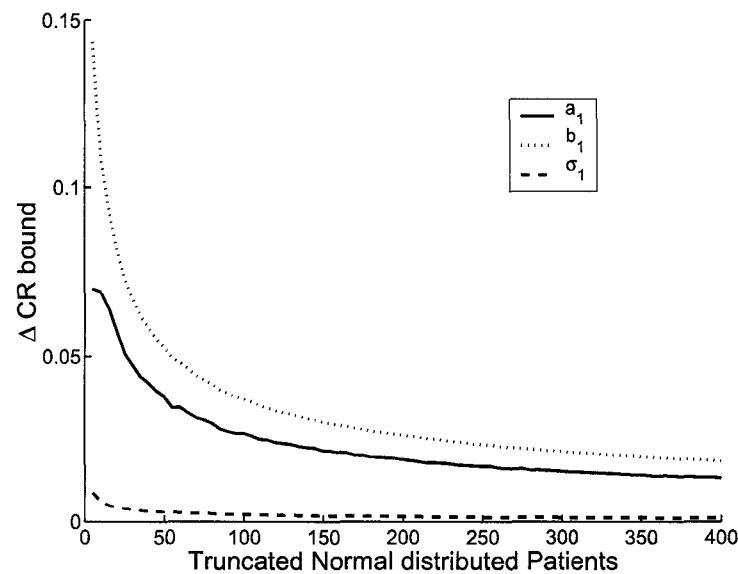
FIGURE 5.9. The difference of the CR bounds between RWT and conventional regression analysis with truth when the true density is truncated normal distributed.

takes on the form

$$\Lambda = \prod_{p=1}^{P} \left(\prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \int_{-\infty}^{\infty} d\Theta_p \left(\frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp\left(\frac{1}{2\sigma_\alpha^2}(\Theta_p - \mu)^2 \right) \right) \right.$$
$$\left. \cdot \exp\left(-\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2 \right) \right) \tag{5.13}$$

where σ_α and μ are the standard deviation and mean, respectively, for the assumed distribution. While this choice of distribution does not make sense for the EF problem, it very well may arise in different applications of RWT. This integral has the following closed-form solution

$$\Lambda = \prod_{p=1}^{P} \left(\prod_{m=1}^{M} \frac{1}{2} \frac{1}{\sqrt{A\sigma_m^2\sigma_\alpha^2}} \exp\left(\frac{B^2 - AC}{A} \right) \right), \tag{5.14}$$

where A, B, and C are given as

$$A = \sum_{m=1}^{M} \frac{a_m^2}{2\sigma_m^2} + \frac{1}{2\sigma_\alpha^2} \tag{5.15}$$

$$B = \sum_{m=1}^{M} \frac{a_m b_m - \theta_{pm} a_m}{2\sigma_m^2} - \frac{\mu}{2\sigma_\alpha^2} \tag{5.16}$$

$$C = \sum_{m=1}^{M} \frac{(\theta_{pm} - b_m)^2}{2\sigma_m^2} + \frac{\mu^2}{2\sigma_\alpha^2}. \tag{5.17}$$

While this expression is complicated, it is simply a multivariate Gaussian that can be written as

$$\Lambda = \prod_{p=1}^{P} \left(N \exp\left(-\frac{1}{2}(\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \boldsymbol{K}^{-1}(\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \right) \right), \tag{5.18}$$

where N is the normalizing constant and $\boldsymbol{K}$ and $\overline{\boldsymbol{\theta}}_p$ are the covariance and mean of the vector $\boldsymbol{\theta}_p$. An example for the case when $M = 3$ is given in Appendix C.

We calculate the components of the Fisher information matrix for RWT using the fol-

lowing form

$$J_{ij} = -\left\langle \frac{\partial^2}{\partial \eta_i \partial \eta_j} \lambda \right\rangle_{\{\theta_p\}} \Bigg|_{\eta=\eta_o} \tag{5.19}$$

$$= -P.\left(\frac{\partial^2}{\partial \eta_i \partial \eta_j} (\log(N)) \right. \tag{5.20}$$

$$\left. -\frac{1}{2} \left\langle \frac{\partial^2}{\partial \eta_i \partial \eta_j} \left((\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \, \boldsymbol{K}^{-1} \, (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \right) \right\rangle_{\theta_p} \right) \Bigg|_{\eta=\eta_o}. \tag{5.21}$$

which we may rewrite as

$$J_{ij} = -P \left(\frac{\partial^2}{\partial \eta_i \partial \eta_j} (\log(N)) \right. \tag{5.22}$$

$$\left. -\frac{1}{2} \mathrm{tr} \left(\boldsymbol{K} \frac{\partial^2}{\partial \eta_i \partial \eta_j} (\boldsymbol{K}^{-1}) \right) - \frac{\partial}{\partial \eta_i} (\overline{\boldsymbol{\theta}}_p) (\boldsymbol{K}^{-1}) \frac{\partial}{\partial \eta_j} (\overline{\boldsymbol{\theta}}_p) \right) \Bigg|_{\eta=\eta_o}. \tag{5.23}$$

The derivation of this expression is given in appendix D.

5.4.1 Singular Fisher information matrices

The Fisher information matrix is singular when the assumed distribution $pr_a(\Theta|\{\mu, \omega\})$ is unbounded. This singularity can be seen by looking at likelihood function used in RWT which is of the form

$$L(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{r}|\boldsymbol{\theta}_p) = \int d\Theta_p \, pr_a(\Theta_p|\boldsymbol{r}) \, pr(\boldsymbol{\theta}_p|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\sigma}, \Theta_p). \tag{5.24}$$

If Θ_p is unbounded and if both $pr_a(\Theta_p|\boldsymbol{r})$ and $pr(\boldsymbol{\theta}_p|\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\sigma}, \Theta_p)$ are members of families with shift and scale parameters, then the value of L will remain constant for specific shifts and scales in the above integral. These shifts and scales result in a singularity in the Fisher information matrix brought on by the constant value of L.

This singularity is displayed nicely in our Gaussian example by looking at the integral in equ. 5.13. We may rewrite the exponent in the integrand as

$$\frac{1}{2\sigma_\alpha^2}(\Theta_p - \mu)^2 - \sum_{m=1}^{M} \frac{a_m^2}{2\sigma_m^2} \left(\Theta_p - \frac{\theta_{pm} - b_m}{a_m} \right)^2.$$

Looking at this expression, we notice that the value of the integral, and thus its likelihood, is unchanged by the following shift,

$$\mu \to \mu + \Delta, b \to b - a\Delta. \tag{5.26}$$

In other words, the likelihood is constant along this direction in parameter space. The likelihood is also unchanged by the following scaling,

$$\mu \to \tau\mu, \sigma_\alpha^2 \to \tau^2\sigma_\alpha^2, a \to \frac{1}{\tau}a \tag{5.27}$$

Both correspond to singularities in the Fisher information matrix.

The problem of singular Fisher information matrices is dealt with rigorously in a work by Stoica and Marzetta [57] in which they derive an expression for the CR bound of the form

$$C \geq UJ^+U + \gamma U_{null}U_{null}^t, \tag{5.28}$$

where, in the case of an unbiased estimator, U is an orthonormal matrix, U_{null} is the component of U projecting to the null space of J, J^+ is a Moore-Penrose pseudoinverse of J, γ is a free parameter, and C is the covariance matrix for the parameters. In the example presented here J does have a non-trivial null component, which follows directly from the shift described in eqn. 5.26 and scaling described in eqn. 5.27 above. The existence of these null components implies C is unbounded as γ tends to infinity. While the dimension of the null space for the Fisher information is only two, the direction of the vector involves $2M+2$ parameters. As a result the CR bound for these parameters is infinite.

In terms of the log-likelihood function this shift and scale can be written as

$$\lambda(\eta|\{\theta_p\}) = \lambda(T_{\Delta,\alpha}\eta|\{\theta_p\}), \tag{5.29}$$

where $T_{\Delta,\alpha}$ is a linear operator that shifts the parameters according to eqs. 5.26 and 5.27. One can view this singularity of the Fisher information matrix as a result of the invariance of the log-likelihood under the two-parameter group of shifts and scales in eqn. 5.29.

5.5 Discussion

We have evaluated our method (RWT) for comparing and ranking estimation methods without the use of a gold standard. We found experimentally, in Chap. 3, that RWT produces estimates with very small bias when the data match the model. We also found in Chap. 3 that the bias was larger, but reasonable, when the data does not match the model [28]. The focus of our work here was to quantify the efficiency of our estimator. We have not yet proved RWT to be efficient. However, we have shown above that, experimentally, RWT is close to efficient.

We have also used Fisher information to analyze a case in which RWT does not work. The particular difficulty that arises in this case can be avoided by choosing an assumed distribution with compact support. We looked at the difference in the CR bounds for RWT and conventional regression analysis with truth and found that this difference decreased for all model parameters as the number of patients increased. Furthermore, the CR bound for RWT decreased as the number of modalities increased. However, this decrease quickly levels off at a value above the CR bound for the conventional regression analysis, indicating that there is a limit to how much information can be gained by increasing the number of modalities. Lastly, we showed that the CR bound is largely independent of the linear model parameters.

Because RWT is close to efficient, the Fisher information matrix for RWT is a valuable tool for analyzing and understanding how well our method will work for a given estimation task. For example, we could calculate the Fisher information for reasonable values of the true model parameters for a particular problem and use the corresponding CR bounds as measures of the reliability of RWT. A potential shortcoming of our analysis is the assumption that the data matches the model. This information is typically unavailable. Despite this problem, Fisher information has proven to be helpful in evaluating the reliability of RWT and our understanding of the method.

Chapter 6
VALIDATING THE NO-GOLD-STANDARD THEORY USING VOLUME ESTIMATION

6.1 Introduction

In Chapters 3 and 4 we performed extensive studies using simulated data to better understand the performance of RWT. In Chap. 5 we further investigated the performance of RWT by calculating the Fisher information matrix for the linear model parameters and analyzing the corresponding Cramèr-Rao bounds. These studies have largely been successful, yet often received the usual, and justified, skepticism associated with simulation studies. Thus, to address this skepticism we have performed a phantom study involving volume estimation using both computed tomography (CT) and single photon emission computed tomography (SPECT). We chose volume as our parameter of interest largely because estimating cardiac EF is a volume estimation task.

In this chapter we present the design and the results of our volume estimation experiment. We performed our experiment using the dual-modality imaging system developed here at the Center for Gamma Ray Imaging (CGRI) [58], a system we describe in detail in Sec. 6.2. In Sec. 6.3 we present the results of three different volume estimation experiments we performed by comparing the performance of RWT with that of conventional regression analysis. We conclude the chapter with a discussion of our findings.

6.2 Design of Experiment

As mentioned above, our experiment to validate RWT consists of estimating multiple volumes in a phantom using a dual-modality (CT/SPECT) imaging system developed by our group shown in Fig. 6.1. The CT component of the dual-modality system is comprised of an
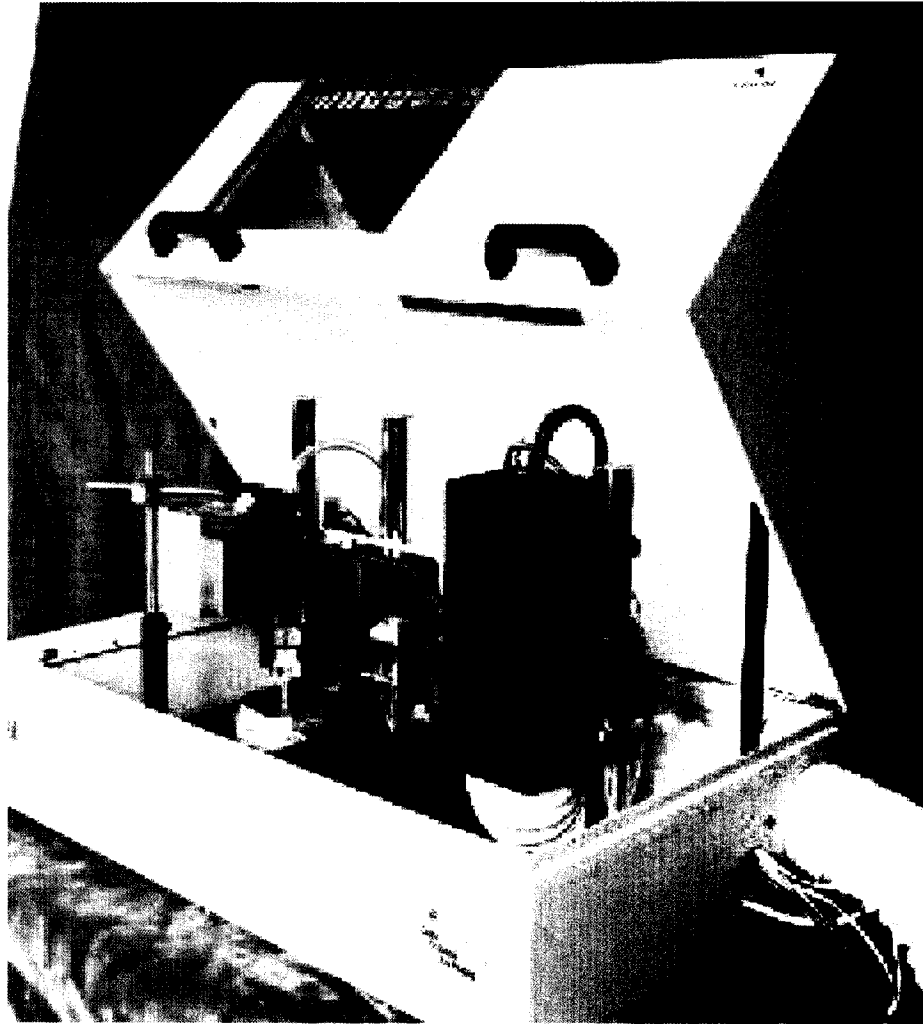
FIGURE 6.1. A photograph of the dual-modality imaging system.

Top View

X-Ray Tube

CdZnTe Detector

Mouse Holder and
Rotation Stage

CCD Detector
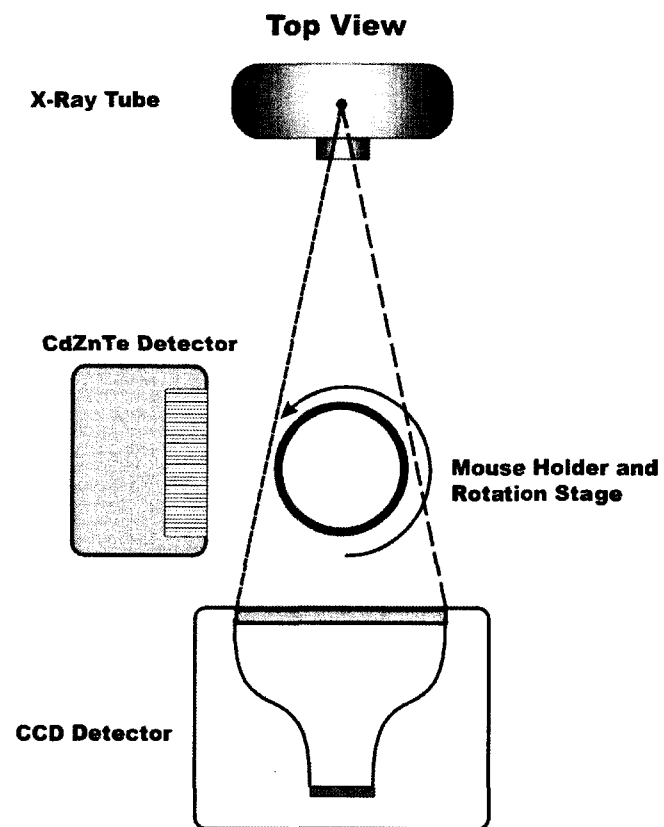
FIGURE 6.2. A schematic diagram for the dual-modality imaging system.

FIGURE 6.3. A photograph of the phantom used in the volume estimation experiment.

Oxford Instruments (XTF5000/75) x-ray tube and a Kodak KAF-1001E series 1024 × 1024 pixel CCD array with an active area of 5.0 × 5.0cm² [46]. The SPECT component of the dual-modality system consists of a compact Cadmium Zinc Teluride (CdZnTe) semiconductor camera with field of view 2.5 × 2.5cm² developed previously at the CGRI [59–61]. Note that the tomographic data in both systems are generated by rotating the object rather than the camera. A schematic diagram of the system is given in Fig. 6.2.

We fabricated the phantom by drilling out an asymmetric pattern in a 2.5cm diameter plexiglass cylinder. A photograph of the phantom is shown in Fig. 6.3. A 3D reconstruction of the phantom imaged on the CT system is given in Fig. 6.4 (note that the reconstruction is inverted in an attempt to better display the complex nature of the phantom). The phantom has a volume of approximately 4ml. We used a solution consisting of 25% ^{99m}Tc-pertechnetate solution (typically 8mCi/ml), 5% omnipaque (an x-ray contrast agent), and 70% water. We imaged 25 volumes with values we sampled from a truncated normal distribution with lower and upper bounds of 0.5 and 3.5ml, respectively. The phantom was filled to the predetermined volumes using a pipette accurate to ± 2μl. Given the accuracy
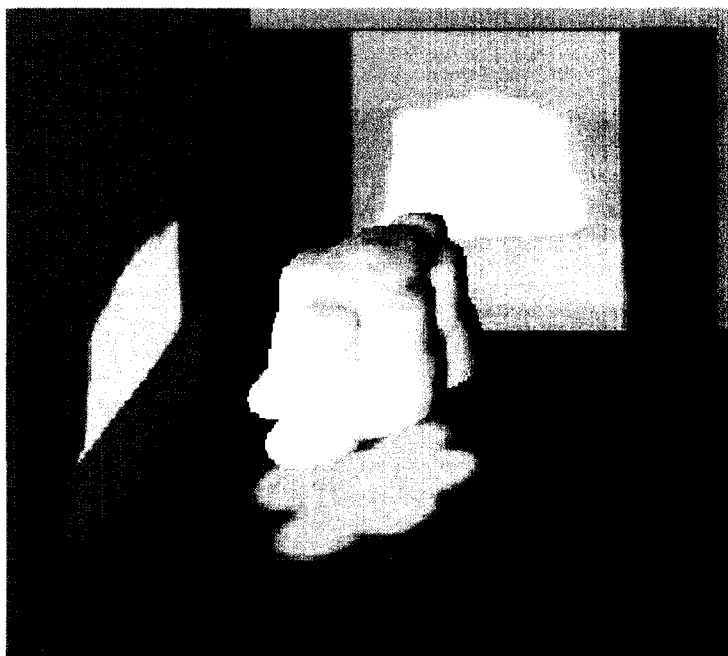
FIGURE 6.4. A 3D Reconstruction of the phantom filled with 3.06ml of solution imaged using the CT system. Note that the reconstruction is inverted in an attempt to better display the complex nature of the phantom.

of the pipette, it is a gold standard of volume for this experiment. Thus we will be able to compare RWT with conventional regression analysis.

Image data were acquired at 180 projection angles with 1 second exposures on the CT system. The SPECT data were taken at 60 projection angles each with 35 seconds of exposure. Data were collected for forty-two different volumes, totaling approximately 16 Gigabytes of projection data. The data collected using the CT system were reconstructed on a 64 × 64 × 32 voxelized grid, while data collected using the SPECT system were reconstructed on a 64×64×64 voxelized grid. All CT data were reconstructed using filtered back projection (FBP) [62]. A slice of a CT reconstruction is shown in Fig. 6.5. SPECT data were reconstructed using either FBP, as shown in Fig. 6.6, or maximum-likelihood expectation maximization (MLEM) [63] as shown in Fig. 6.7. We thresholded voxel values in order to segment out the solution in the image reconstructions. We will discuss our
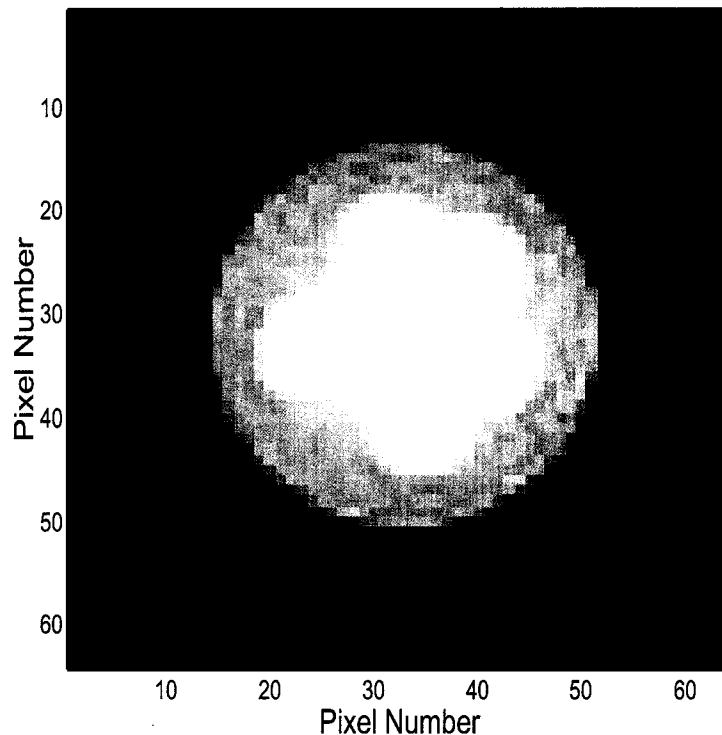
FIGURE 6.5. A 2D slice of the 3D Reconstruction of the phantom imaged with 3.06ml of solution using the X-ray CT.

approach to reconstruction and thresholding in more detail in the results section.

6.3 Results

In this section we discuss the results of three separate volume estimation experiments. In section 6.3.1 we present the results of Experiment I involving twenty-five different volumes and three separate volume estimation techniques, the results of which were presented at the 2003 SPIE Conference on Medical Imaging [64]. Experiment II involves more data, forty-two volumes, and only two estimation techniques. In Experiment III we investigate the impact of a sub-par estimator on RWT. Image reconstruction in Experiments I and II were performed using FBP and the entire set of projection data. In Experiment III, however, we use the MLEM algorithm to reconstruct the SPECT data set and reconstruct the CT data
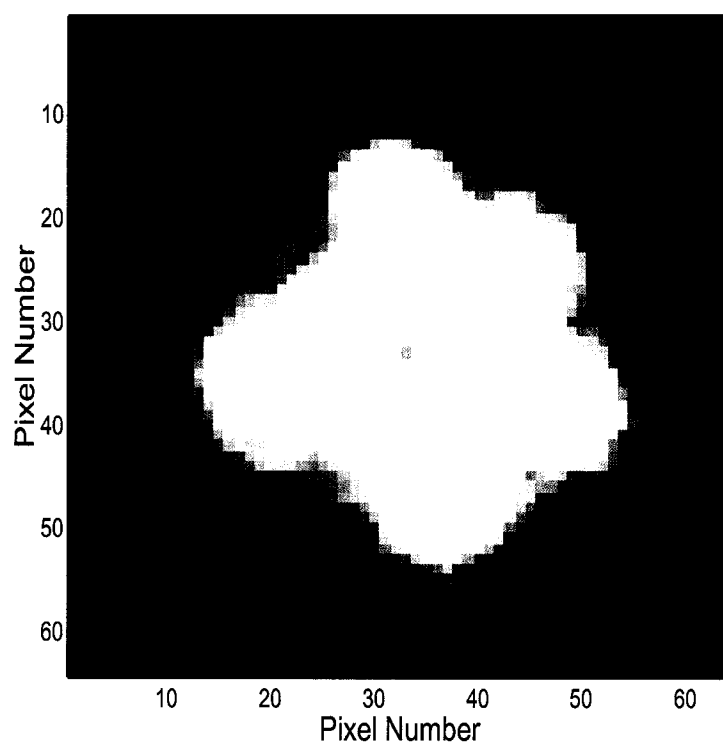
FIGURE 6.6. A 2D slice of the 3D Reconstruction of the phantom imaged with 3.06ml of solution using the Spot Imager.
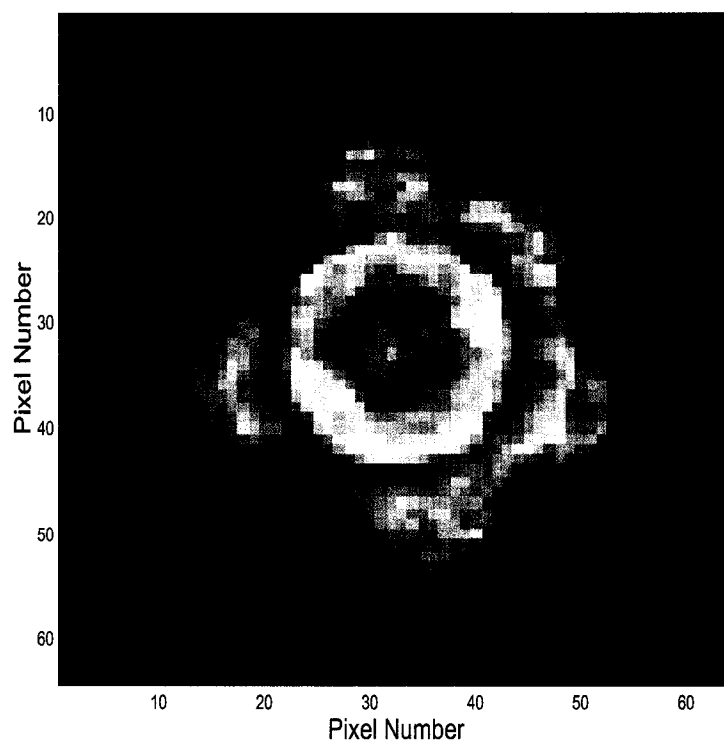
FIGURE 6.7. A 2D slice of the 3D EM Reconstruction of the phantom imaged with 3.06ml of solution using the Spot Imager.

with a limited number of projection angles.

6.3.1 Experiment I

In this section we present the results of an experiment to validate RWT involving twenty-five different volumes imaged using the dual-modality system. We generated two sets of volume estimates using the CT data and a third using the SPECT data. For the SPECT reconstructions we chose our threshold values manually using a gray-level histogram for each image. We generated two sets of volume estimates using the CT data. The first estimation approach, CTI, consisted of manual thresholding and included magnification correction. The second estimation approach, CTII, used a fixed threshold and did not account for magnification. Thus the relationship between the estimates obtained using CTII and the gold standard is quantified with a slope greater than one. In Fig. 6.8 we display a histogram of voxel values from a CT reconstruction.

We applied RWT to the three sets of volume estimates obtained in the experiment resulting in estimated slopes, intercepts, and noise terms relating the volume estimates to the gold standard. This analysis did not use the known gold standard (*i.e.*, pipette values) to determine this relationship. We also performed conventional regression analysis using the gold standard (*i.e.*, pipette values) for comparison. In Table 6.1 we summarize these results. Note that there are differences between the slopes, intercepts, and noise terms obtained from these two methods. However, the ordering of the slopes and noise terms is the same between the two methods, as predicted by the simulation in Sec. 3.4.2. Regression analysis performs better than RWT because it has access to the x coordinates (*i.e.*, the gold standard).

In Fig. 6.9 we plot the volume estimates obtained using the three aforementioned techniques versus the gold standard. Also shown in Fig. 6.9 are the lines representing the linear model parameters estimated using RWT. The results shown in Fig. 6.9 are somewhat misleading given that we use the gold standard in the plots; an advantage RWT does not have.
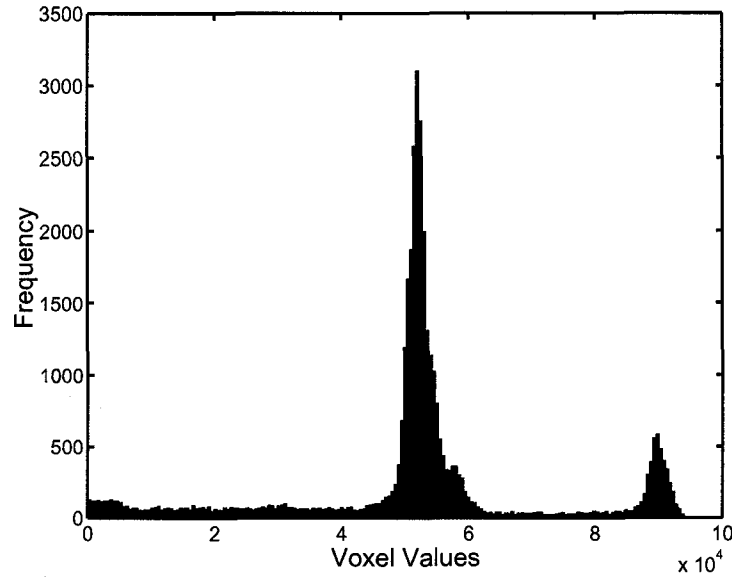
FIGURE 6.8. A histogram of positive voxel values from a CT reconstruction of the phantom imaged with 3.06ml of solution. The two peaks consist of voxel values corresponding to plexiglass ($\sim$50,000) and solution ($\sim$90,000). A majority of the voxel values corresponding to air are negative, and thus not shown. Note that the voxel values are not given in Hounsfield units.

TABLE 6.1. Estimates of the linear model parameters using regression analysis with and without truth. Note that the CTI estimates were obtained using manual thresholding and magnification correction while the CTII estimates were obtained using a fixed threshold and no magnification correction.

Estimation Techniques	SPECT	CTI	CTII
	$\hat{a}_{SI}$	$\hat{a}_{CTI}$	$\hat{a}_{CTII}$
Estimates from regression analysis	0.9387	1.0462	1.6135
Estimates using RWT	0.8091	0.9032	1.3947
	$\hat{b}_{SI}$	$\hat{b}_{CTI}$	$\hat{b}_{CTII}$
Estimates from regression analysis	0.0110	-0.0684	-0.0210
Estimates using RWT	0.0253	0.0026	0.3572
	$\hat{\sigma}_{SI}$	$\hat{\sigma}_{CTI}$	$\hat{\sigma}_{CTII}$
Estimates from regression analysis	0.0351	0.0428.	0.0646
Estimates using RWT	0.0478	0.0512	0.1003

This explains the noticeable imperfections in the plots.

As discussed previously, RWT also returns estimates of the parameters defining the underlying distribution of the gold standard. Because we generated the gold standard from a known distribution, we can, again, evaluate the performance of RWT. Fig 6.10 contains plots of the true and estimated densities along with a histogram of the data used in the experiment.

6.3.2 Experiment II

In this experiment we evaluate the performance of RWT using forty-two data sets and only two estimation techniques. The two volume estimation techniques we evaluated are SPECT with manual thresholding and CT with magnification correction and manual thresholding. We display the numerical results of our linear model parameter estimation in Table 6.2 and the graphical results in Fig. 6.11. Fig 6.12 contains plots of the true and estimated densities along with a histogram of the data used in the experiment. Notice that the results of RWT for Experiment II are better than in Experiment I. We will address this improvement further in the discussion.

6.3.3 Experiment III

The primary goal of Experiment III is to investigate the impact of a noisy modality on RWT for a case with three modalities. To generate a noisy modality we first studied the impact decreasing the number of projection angles has on volume estimation. We were able to do this quantitatively by plotting the volume estimate as a function of the number of projection angles as shown in Fig. 6.13. From this plot and a qualitative assessment of the histograms corresponding to different numbers of projection angles, we decided to reconstruct the CT data set using only fifteen projection angles for each volume.

For our "good" estimation techniques we used the same CT estimates as generated in Experiment II and our SPECT data reconstructed using the MLEM algorithm. MLEM has

(a)

(b)

(c)
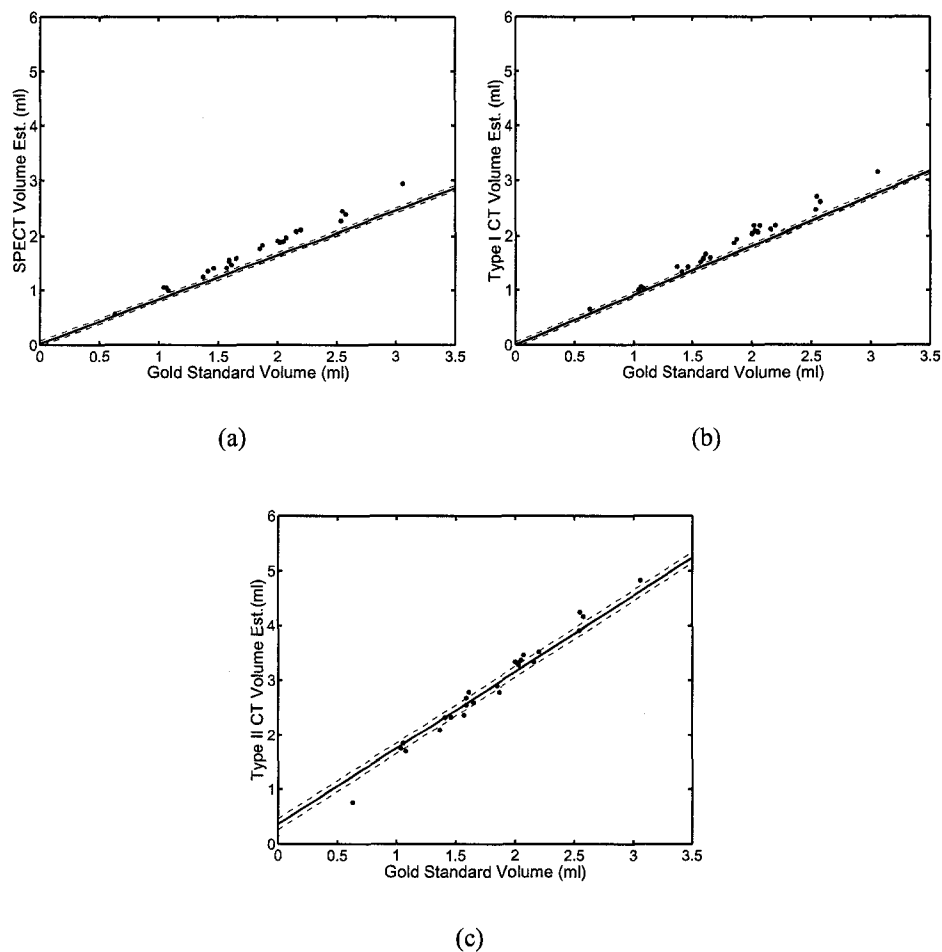
FIGURE 6.9. Experiment I: The results of a phantom study for estimating volumes from three estimation techniques. Twenty-five different volumes were imaged on two different modalities (SPECT, CT). In each graph we have plotted the true volume against the estimates from three different estimation techniques ((a)SPECT, (b)CTI, (c)CTII). The solid lines were generated using the estimated linear model parameters. The dashed lines denote the estimated standard deviations for each estimation technique. Values for these parameters are shown in Table 6.1. Note that although we have plotted the true volumes on the x axis of each graph, this information was not used in computing the linear model parameters
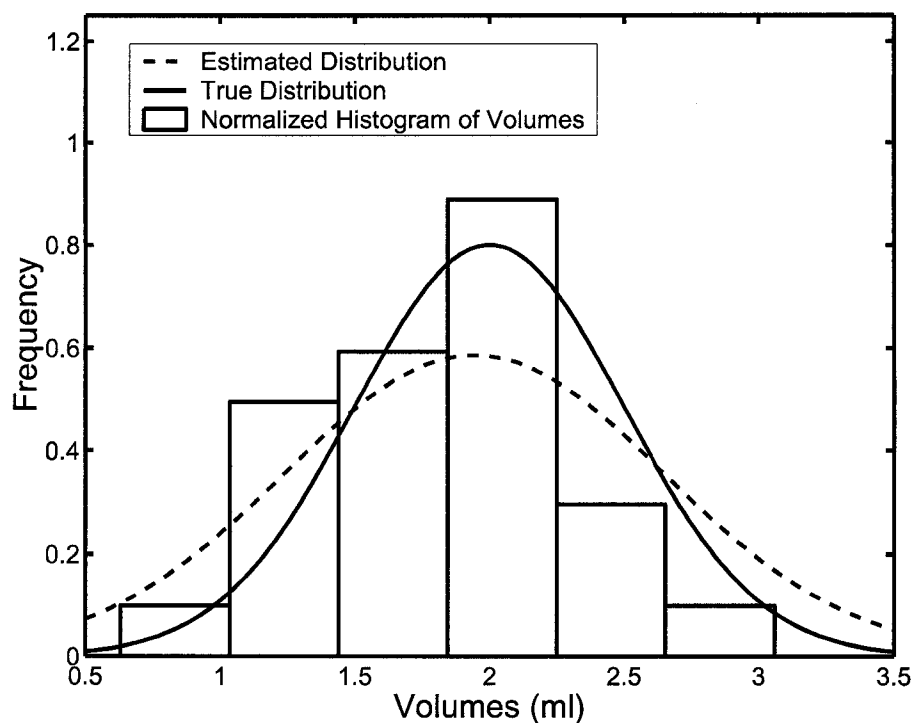
FIGURE 6.10. A comparison of the normalized histograms for the underlying volumes with the parameters returned by RWT estimating the mean and variance of the underlying gold-standard distribution. The true volumes were sampled from a truncated-normal distribution with a mean of 2ml and standard deviation of 0.5ml. The 25 volumes themselves had a sample mean of 1.8016ml and sample standard deviation 0.5607ml. Our no-gold-standard analysis predicted a mean of 1.9444ml and a standard deviation of 0.7057ml.
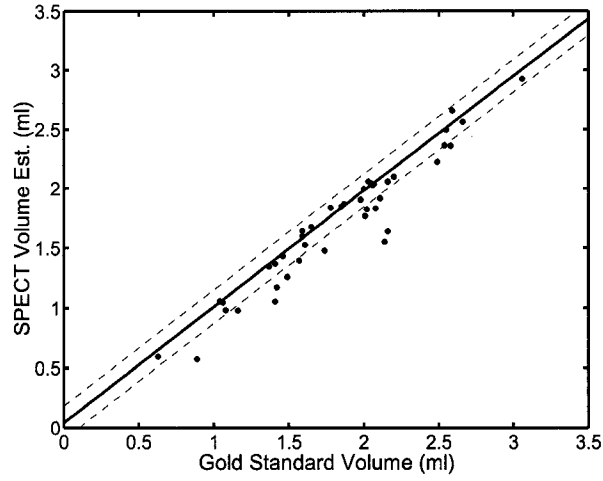
(a)
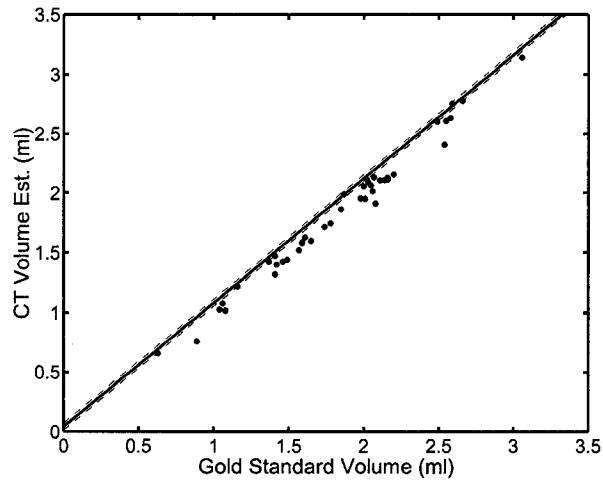
(b)

FIGURE 6.11. Experiment II: The results of a phantom study for estimating volumes with two estimation techniques. Forty-two different volumes were imaged on two different modalities (SPECT, CT). In each graph we have plotted the true volume against the estimates from two different estimation techniques ((a)SPECT, (b)CT). The solid lines were generated using the estimated linear model parameters. The dashed lines denote the estimated standard deviations for each estimation technique. Values for these parameters are shown in Table 6.2.
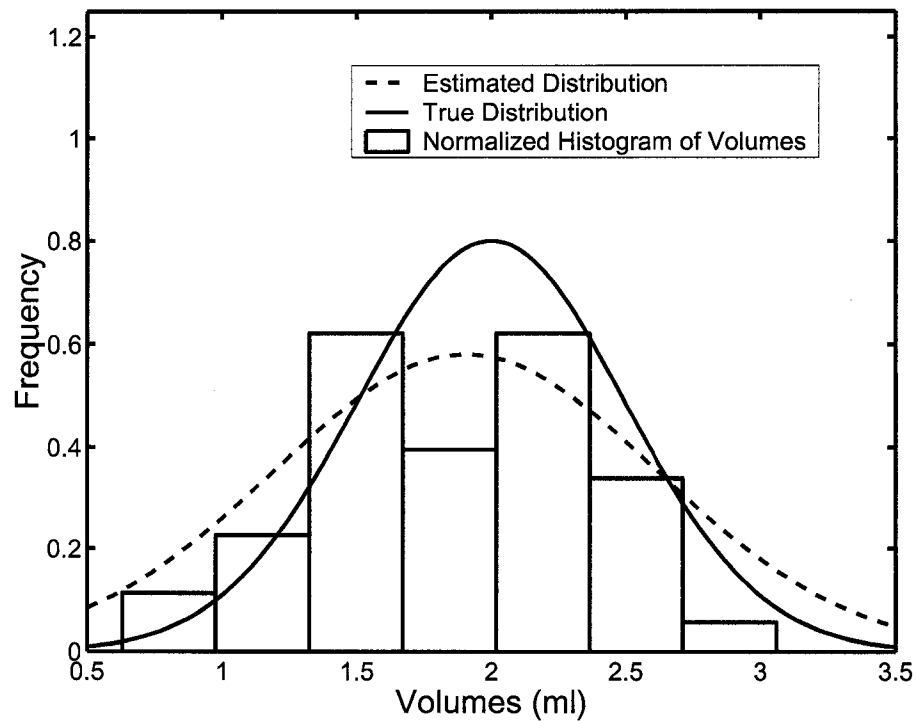
FIGURE 6.12. Experiment II: A comparison of the normalized histograms for the underlying volumes with the parameters returned by RWT estimating the mean and variance of the underlying gold-standard distribution. The true volumes were sampled from a truncated-normal distribution with a mean of 2ml and standard deviation of 0.5ml. The 25 volumes themselves had a sample mean of 1.8383ml and sample standard deviation 0.5301ml. Our no-gold-standard analysis predicted a mean of 1.9062ml and a standard deviation of 0.7141ml.
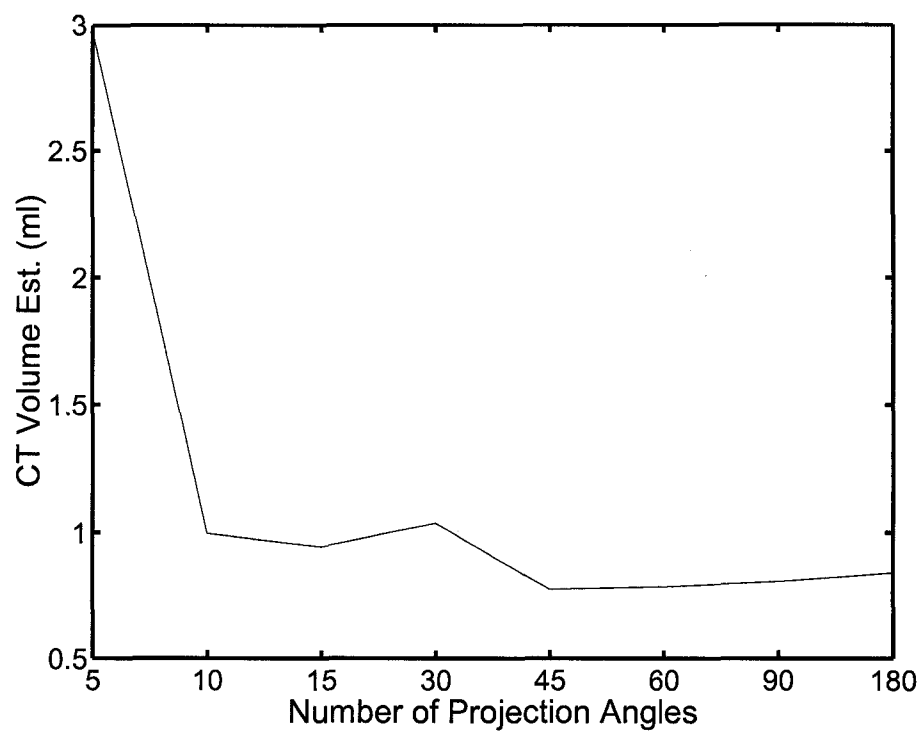
FIGURE 6.13. A plot of estimated volume using CT as a function of the number of projection angles used in the reconstruction. All reconstructions were performed using FBP.

TABLE 6.2. Experiment II: Estimates of the linear model parameters using regression analysis with and without truth with 42 data sets and two estimation techniques.

Estimation Techniques	SPECT	CT
	$\hat{a}_{SI}$	$\hat{a}_{CT}$
Estimates from regression analysis	0.9708	1.0462
Estimates using RWT	0.9306	0.9999
	$\hat{b}_{SI}$	$\hat{b}_{CT}$
Estimates from regression analysis	-0.0691	-0.0743
Estimates using RWT	0.0489	0.0476
	$\hat{\sigma}_{SI}$	$\hat{\sigma}_{CTI}$
Estimates from regression analysis	0.1521	0.0657
Estimates using RWT	0.1386	0.0257
	$\hat{\sigma/a}_{SI}$	$\hat{\sigma/a}_{CTI}$
Estimates from regression analysis	0.1567	0.0628
Estimates using RWT	0.1489	0.0257

a free parameter, namely stopping point, which has been studied extensively in our group [65, 66] . In order to determine a proper stopping point for the algorithm, we evaluated the relationship between our volume estimates and the iteration number. In Fig. 6.14 we plot our volume estimates as a function of iteration number using a fixed threshold for our segmentation technique. In Fig. 6.15 we plot our volume estimates as a function of iteration number using manual thresholding. The different segmentation approaches have a dramatic impact on the shape of the plots as a function of iteration number.

Unlike the first two experiments, in this experiment we also perform RWT using a beta density as our assumed distribution. Namely, we evaluate the performance of RWT when the data do not match the model. We display the numerical results of our linear model parameter estimation using both beta and truncated normal families as the assumed distributions in Table 6.3. We display the graphical results of RWT using the truncated normal distribution as the assumed distribution in Fig. 6.16. Fig 6.17 contains plots of the true and estimated densities along with a histogram of the data used in the experiment. The
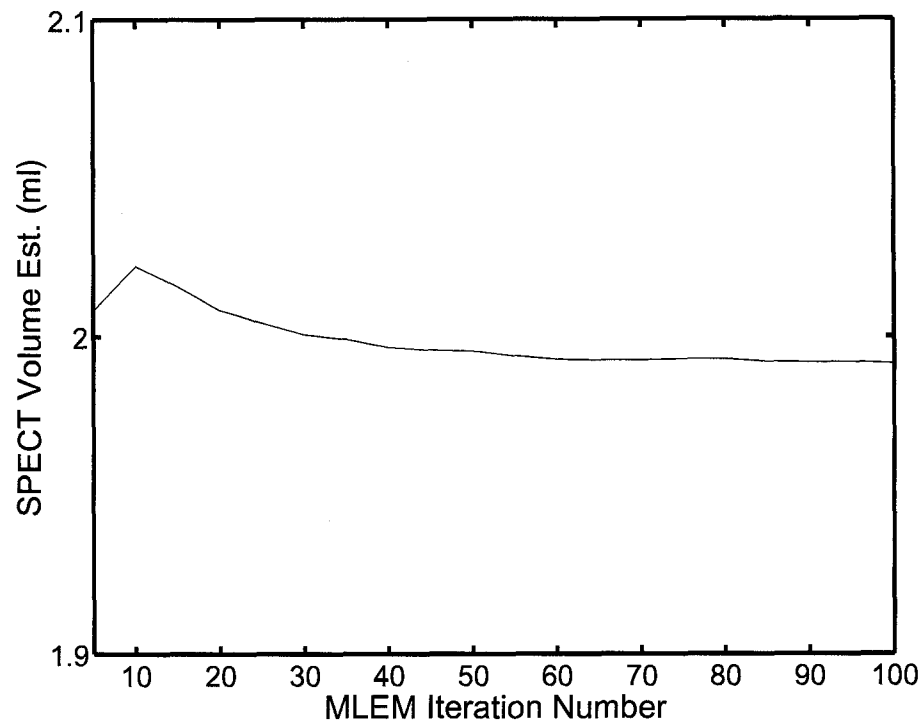
FIGURE 6.14. A plot of estimated volume using SPECT as a function of the iteration number for the MLEM reconstruction algorithm. Note that the gold standard volume was 1.87 ml. A fixed threshold was used for the volume estimation.
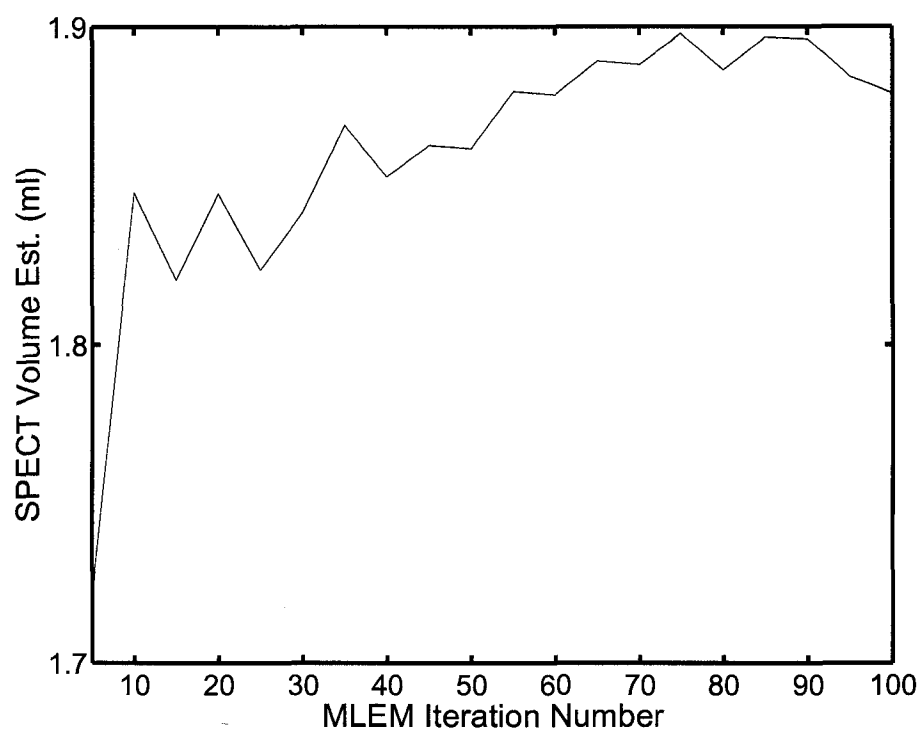
FIGURE 6.15. A plot of estimated volume using SPECT as a function of the iteration number for the MLEM reconstruction algorithm. The gold standard volume was 1.87 ml. For our reconstruction scheme, 1.87 ml corresponds to approximately 34,080 voxels.

TABLE 6.3. Estimates of the linear model parameters using regression analysis with and without truth with 42 data sets and three estimation techniques.

Estimation Techniques	CT	SPECT-EM	CT-15
	$\hat{a}_{CT}$	$\hat{a}_{EM}$	$\hat{a}_{LA}$
Estimates from regression analysis	1.0405	0.9203	0.9811
Estimates using RWT: Truncated Normal	0.9635	0.8882	0.9392
Estimates using RWT: Beta	1.3008	1.1775	1.2664
	$\hat{b}_{CT}$	$\hat{b}_{EM}$	$\hat{b}_{LA}$
Estimates from regression analysis	-0.0743	0.0341	0.2155
Estimates using RWT: Truncated Normal	-0.0976	-0.0579	0.1238
Estimates using RWT: Beta	-0.7106	-0.5813	-0.4708
	$\hat{\sigma}_{CT}$	$\hat{\sigma}_{EM}$	$\hat{\sigma}_{LA}$
Estimates from regression analysis	0.0802	0.1410	0.4147
Estimates using RWT: Truncated Normal	0.1148	0.0256	0.3828
Estimates using RWT: Beta	0.0897	0.0653	0.3778

estimates of the linear model parameters using the beta assumed distribution were not very accurate, this is not a surprise given the number of patients and modalities (see Sec. 4.2).

6.4 Discussion

We have further evaluated our method (RWT) for comparing estimation methods without the use of a gold standard by performing volume estimation using a phantom and multiple imaging systems. We have found that our method does, in fact, allow for the comparison of estimation techniques without the use of a gold standard. Specifically, the estimates of the linear model parameters obtained using RWT are closely correlated with those obtained through standard regression analysis using the x axis. The errors observed in our estimates of the linear model parameters are consistent with the results of simulation studies presented in Chaps. 3, 4, and 5.

The estimation techniques we employed in Experiment I (SPECT, CTI, and CTII) for volume estimation were not particularly noisy, as can be seen in Fig. 6.9. However, the

(a)

(b)

(c)
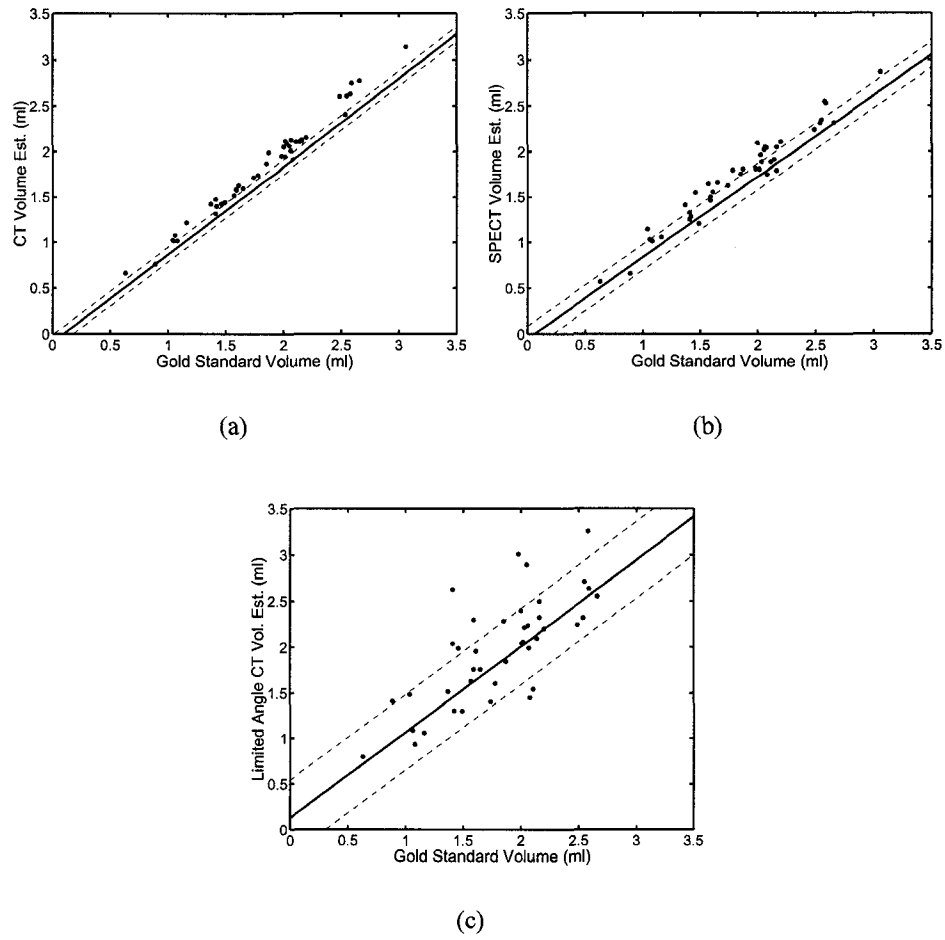
FIGURE 6.16. Experiment III: The results of a phantom study for estimating 42 volumes using three estimation techniques. The estimates given in (a) are the same as the CT estimates in Experiment II. The estimates in (b) were generated using SPECT data reconstructed with the MLEM algorithm carried out to 100 iterations. The estimates in (c) were generated using only fifteen of the 180 projection angles of the CT data set.
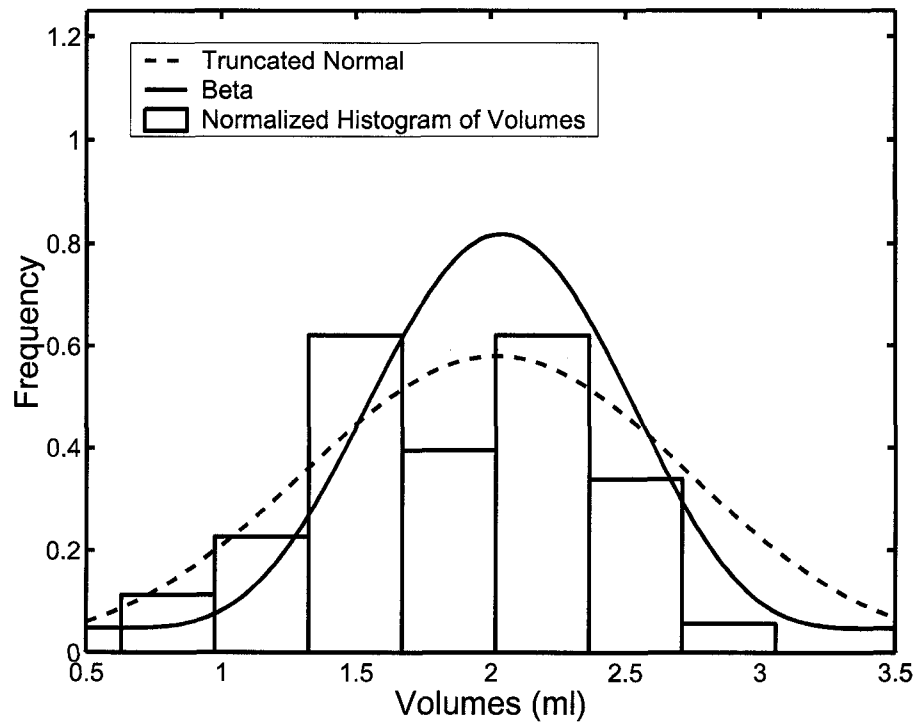
FIGURE 6.17. Experiment III: A comparison of the normalized histograms for the underlying 42 volumes with the parameters returned by RWT estimating the mean and variance of the underlying gold-standard distribution. The true volumes were sampled from a truncated-normal distribution with a mean of 2ml and standard deviation of 0.5ml. The 42 volumes themselves had a sample mean of 1.8383 and sample standard deviation 0.5301ml. Our no-gold-standard analysis predicted a mean of 2.0206ml and a standard deviation of 0.7137ml. When the assumed distribution was from the beta family, RWT returned beta parameter values of $\nu = 4.1794$ and $\omega = 4.3558$. A beta distribution characterized by these values has a mean of 1.9691ml and a standard deviation of 0.4866ml.

slope of CTII is substantially greater than one due to magnification in our CT system. RWT accurately determined this increased slope (Fig. 6.9(c)). The lack of noise present in the three volume estimation techniques used in our experiment led to differences between the estimates of $\hat{\sigma}_m/\hat{a}_m$ for each technique that were not statistically significant. This result implies that the three volume estimation techniques we used performed equally well.

In Experiment II we added seventeen gold-standard volumes and removed an estimation technique. The results of the experiment were noticeably better than those of Experiment I. These results agree well with the simulation results presented in Sec. 5.3.1. Most importantly, RWT returns an accurate estimate of $\hat{\sigma}_m/\hat{a}_m$ both techniques.

In Experiment III we performed RWT with 42 volumes estimates using three different techniques. We studied the impact of using MLEM rather than FBP to reconstruct the SPECT data as well as the impacting of removing projection angles prior to reconstructing the CT data. No significant gain in volume estimation accuracy was achieved using MLEM. Removing projection angles, on the other hand, produced an expected drop off in accuracy. The volume estimates using only 15 of the 180 projection angles produced an extremely noisy modality. The significant increase in σ for a given estimation technique is very likely the cause for the somewhat inaccurate estimates of the linear model parameters given in Table 6.3. This result follows the findings of the simulation study shown in Fig. 4.3, in which RMSE increased as a function of σ_m.

In conclusion, we have further evaluated our method (RWT) as applied to a real data set rather than a data set generated numerically in simulation. The results of RWT as calculated in our volume estimation experiment match very well with the simulation studies performed in previous chapters. As discussed, the performance of RWT suffered with the introduction of a very noisy estimation technique in Experiment III. In practice, we recommend researchers remove the noisy data set and perform RWT again with just two estimation techniques as is done in Experiment II.

Chapter 7

CONCLUSION

7.1 Introduction

In this chapter we summarize the results of the dissertation, discuss the practical implemen-
tation of RWT, and present problems worth investigating in the future. In Sec. 7.2 and 7.3
we discuss the advantages and limitations of RWT, respectively. In Sec. 7.4 we discuss a
set of consistency checks researchers can use when performing RWT to gain confidence in
their results. We conclude the chapter with future work section made up of results from a
preliminary study in which we perform the consistency checks using a real data set without
a gold standard along with a discussion of some areas of potential research for RWT.

7.2 Discussion of RWT

We have presented and evaluated a method for ranking estimation methods in medical
imaging without the use of a gold standard. Our model consists of assuming that each
estimation method is simply an affine transformation of the gold standard with Gaussian
noise. Our method, RWT, then uses a maximum-likelihood approach to estimate the pa-
rameters defining this mapping. These parameters, which we have called the linear model
parameters, can then be used to rank the different estimation methods.

We showed through numerical simulations in Chaps. 3 and 4 that RWT is able
to accurately estimate the linear model parameters under varying conditions. In Chap. 5 we
calculated the Fisher information for RWT using our likelihood expression and we were
able to show numerically that our estimator achieves the Cramér-Rao bound, indicating we
have an efficient estimator of the linear model parameters. In Chap. 6 we moved away from
simulation studies by performing a volume estimation experiment using a phantom imaged
via SPECT and x-ray CT here at the Center for Gamma Ray Imaging. RWT performed

extremely well with both simulated and real data. We strongly recommend that RWT be used in the future by researchers comparing estimation techniques. A noteworthy aspect of RWT is the exploitation of information previously unused. We have shown that we can succeed without the x-axis if we have measurements on multiple modalities for a common group of patients.

7.3 Limitations of RWT

The key advantage of RWT over conventional regression analysis is that it does not require a gold standard. The performance of RWT, however, is hindered by this lack of information. Furthermore, like conventional regression analysis, RWT assumes a known functional form for the relationship between the gold standard and the data. However, unlike conventional regression analysis, this relationship cannot be visually assessed without the gold standard. We must also assume a functional form of the gold standard density $pr(\Theta)$, an assumption we cannot actually check. The model does, however, permit the parameters characterizing the shape of this density to vary in RWT. We have assumed a Gaussian noise model, that is also implicit in conventional regression analysis, but other noise models are easy to implement in the likelihood expression. We will further discuss independence assumptions in the future work section.

Another limitation of RWT is the requirement (see Sec. 5.4.1) that the assumed distribution be bounded. Often these bounds can be drawn with confidence based upon some physical properties of the forward problem, $i.e.$ EF, though for certain parameters of interest choice of bounds could be cause for argument.

Possibly the largest limitation of RWT is our inability to measure its performance. In all of the studies performed here we have had a gold standard, either through simulation or in the case of the phantom study through the use of a pipette. Such checks will not exist in problems to which we would like to apply RWT. In an attempt to better facilitate the use of RWT we present a few consistency checks in the next section in an attempt to help

researchers gain (or lose) confidence in RWT.

7.4 Consistency Checks

In this work we have been able to compare our estimates of the linear model parameters and the models of the gold standard distribution to their true values since all studies were done with ground truth available. Namely, either we generated the truth in simulation or we used a legitimate gold standard technique (*i.e.*pipette) for comparison. Without a gold standard it is impossible to verify if RWT has accurately estimated the linear model parameters and the parameters of the gold standard distribution. As a result, we have developed a set of consistency checks for researchers to invoke when performing RWT. These consistency checks will not guarantee researchers that RWT is working properly, though they can determine if the method is NOT working properly. We derive three such consistency checks below and in the next section we present preliminary results of the checks as applied in a study we have begun comparing cardiac EF estimation algorithms.

7.4.1 Consistency Check 1

The RWT method returns information relating the gold standard to the estimates returned by a method (equ. 3.1). Two different estimation methods, however, share the same gold standard because the same patient population is employed in both. Thus, we can use the results returned by RWT to represent the relationship between two estimates. Mathematically, this relationship is determined by solving for the gold standard in equ. 3.1 for each of the estimates θ_{pi} and θ_{pj} resulting in,

$$\theta_{pi} = \frac{a_i}{a_j}\theta_{pj} + \left(b_i - \frac{a_i b_j}{a_j}\right). \tag{7.1}$$

If the parameters returned by the RWT method accurately relate the gold standard to the estimates returned by the various methods, then the slopes and intercepts determined by equ. 7.1 should accurately relate the two estimates to each other. The converse of the

previous statement is not necessarily true. Namely, if the relationship between the estimates is accurately determined by RWT, then the relationship between the gold standard and the estimates is not necessarily accurate. However, the contrapositive of the above statement is true and allows us to state that if the relationship between the estimates is not accurately determine then the relationship between the gold standard and the estimates is also not accurately determined by RWT.

7.4.2 Consistency Check 2

Another consistency check that we can run involves comparing the distribution of the gold standard returned by RWT $pr(\Theta|\vec{\rho})$ to the histogram of the raw data θ_{pm} for a given method m calibrated to match the gold standard. Specifically, we will compare the gold standard density to a histogram of

$$\frac{\theta_{pm} - b_m}{a_m} \tag{7.2}$$

for all patients p and for a given method m. These data should not match exactly because equ. 7.2 has extra noise added onto to it from the ϵ_{pm} term. However, if the variance of ϵ_{pm} is small for method m, then the histograms should match the gold-standard density returned by RWT.

7.4.3 Consistency Check 3

A final consistency check we have developed tests the sample covariance of the EF estimates returned by the various methods to what is predicted by the model employed (equ. 3.1). This test incorporates two very important aspects of the RWT model; the first is the linear relationship we assume, and the second is the independence of the noise in the estimates returned by different modalities. Using these two aspects of our RWT model, it can be shown that

$$\text{Cov}(\theta_i, \theta_j) = \hat{a}_i \hat{a}_j \text{Var}(\Theta), \tag{7.3}$$

where θ_i and θ_j are the estimates of EF from the different methods, $\hat{a}_i$ and $\hat{a}_j$ are the estimates of the slopes returned by RWT, and $\mathrm{Var}(\Theta)$ is the variance of the gold standard parameter, and $\mathrm{Cov}(\theta_i, \theta_j)$ is the covariance of the estimates returned by method i and the estimates returned by method j. The expectations needed to compute the covariance and variance in equ. 7.3 are expectations taken over the population of patients, hence, the single subscript denoting the method on the estimates θ_i and the lack of a subscript on the gold standard Θ. The left-hand side of equ. 7.3 can be estimated using the sample covariance and the data given. However, the right-hand side of equ. 7.3 is completely determined by the slopes returned by RWT along with the $\vec{\rho}$ returned by RWT that definines the gold standard density. Hence, we are relating a data-measure to a value determined completely by RWT and using the assumptions that make-up RWT. Once again, this check can only tell us if the the method is not working or our assumptions are invalid; it cannot inform us if the method is working properly.

7.5 Future Work

7.5.1 Comparing Cardiac Ejection Fraction Estimation Algorithms

In this section we present preliminary results of our consistency checks applied to a data set consisting of multiple cardiac EF estimation algorithms used in SPECT. As discussed in Sec. 2.1.3, EF is typically estimated in nuclear medicine using planar imaging. Recently researchers have attempted to estimate EF using SPECT, an approach that presents a more complicated segmentation problem. Many approaches exist for estimating EF from SPECT data [67–71]. Our data set consists of 85 EFs estimated using Quantitative Gated SPECT (QGS) developed by Germano *et al.* [68] at Cedars-Sinai Medical Center, the Emory Cardiac Toolbox (ECT) developed by Faber *et al.* [69] at Emory University, and Wackers-Liu Cardiac Quantification (WLCQ) developed by Wackers and Liu [71] at Yale University.

Nakajima *et al.* [67] performed a study that compared cardiac EF estimation algorithms using the approaches discussed in Sec. 2.2.2. In this section we show the preliminary results

TABLE 7.1. Values for a_m, b_m, σ_m, and σ_m/a_m estimated using RWT and the EF data sets estimated using the QGS, Emory, and WLCQ techniques. Results are given for RWT assuming either a beta or a truncated normal distribution. Note that while the values of the estimates are differ under the two different assumptions, the ranking of the techniques is conserved.

a_m	QGS	Emory	WLCQ
Estimates assuming a beta dist.	1.1537	1.0309	1.1555
Estimates assuming a trunc. normal dist.	0.7547	0.6759	0.7560
b_m	QGS	Emory	WLCQ
Estimates assuming a beta dist.	-0.1356	0.0414	-0.0838
Estimates assuming a trunc. normal dist.	0.1446	0.2909	0.1968
σ_m	QGS	Emory	WLCQ
Estimates assuming a beta dist.	0.0587	0.0470	0.0769
Estimates assuming a trunc. normal dist.	0.0590	0.0464	0.0770
σ_m/a_m	QGS	Emory	WLCQ
Estimates assuming a beta dist.	0.0509	0.0456	0.0665
Estimates assuming a trunc. normal dist.	0.0781	0.0686	0.1018

of a similar study in which we use RWT to compare the different techniques. We present the results of linear model parameters estimated using RWT in Table 7.1 along with the results of the three aforementioned consistency checks in Fig. 7.1, Fig. 7.2, and Table 7.2. Our estimates using RWT performed very well in all three consistency checks. Specifically, RWT performed very well when we assumed a beta density in our model. This result fits well with previous working claiming EF to be beta distributed [7,46].

7.5.2 Evaluating Assays Used to Identify Human Herpesvirus 8

The methodology presented in this dissertation is applicable to numerous estimation problems outside of medical imaging. For example, human herpesvirus 8 (HHV-8) is a recently discovered virus [72] that is running rampant throughout the world, especially Africa. The means of transfer of HHV-8 are unknown and the percentage of the population infected is

(a)

(b)

(c)

FIGURE 7.1. The results of consistency check 1 applied to the cardiac EF estimation algorithm data set. The plots consist of the EF estimates of the three different methods plotted against each other. The regression lines shown are NOT the result of conventional regression analysis. Rather they are the results of the slopes and intercepts in Equ. 7.1 estimated using the results of RWT. The plot shown was generated using results of RWT performed with a beta assumed distribution, plots using a truncated normal assumed distribution are nearly identical.
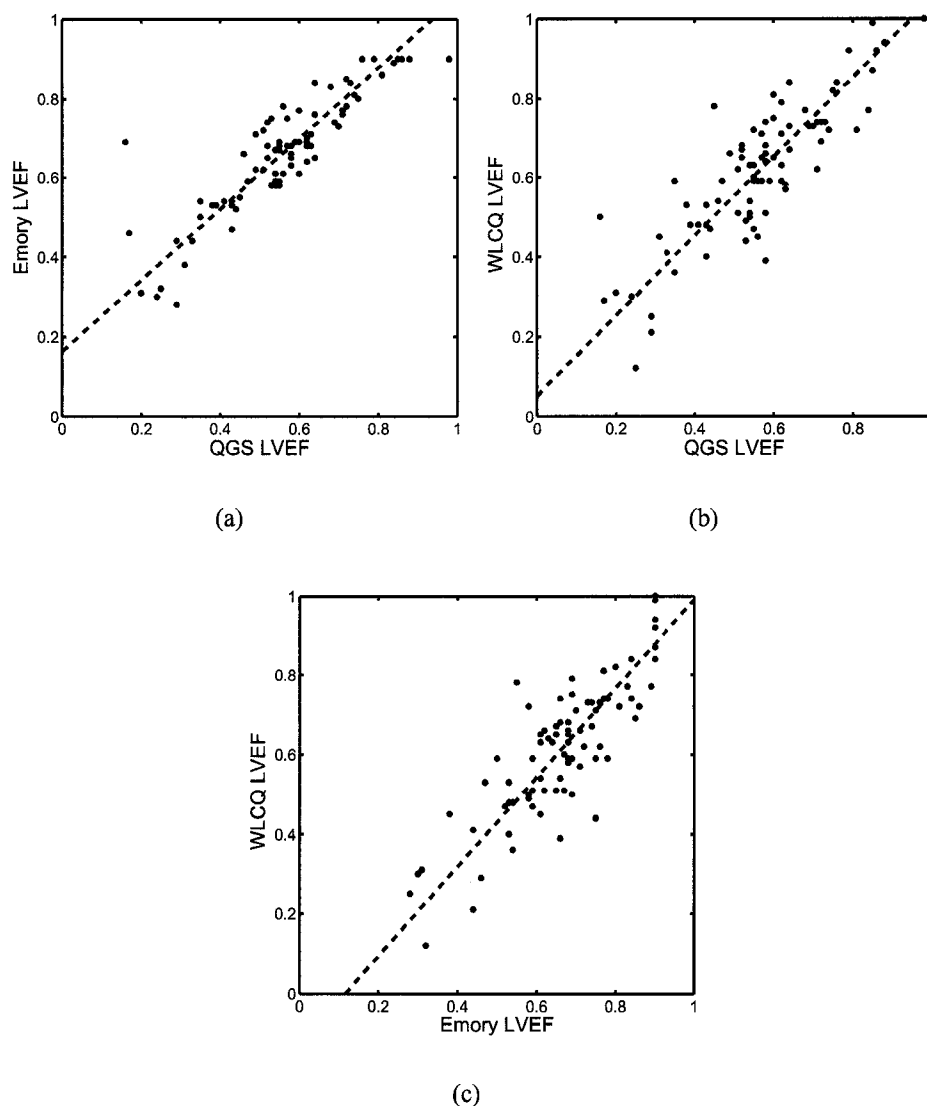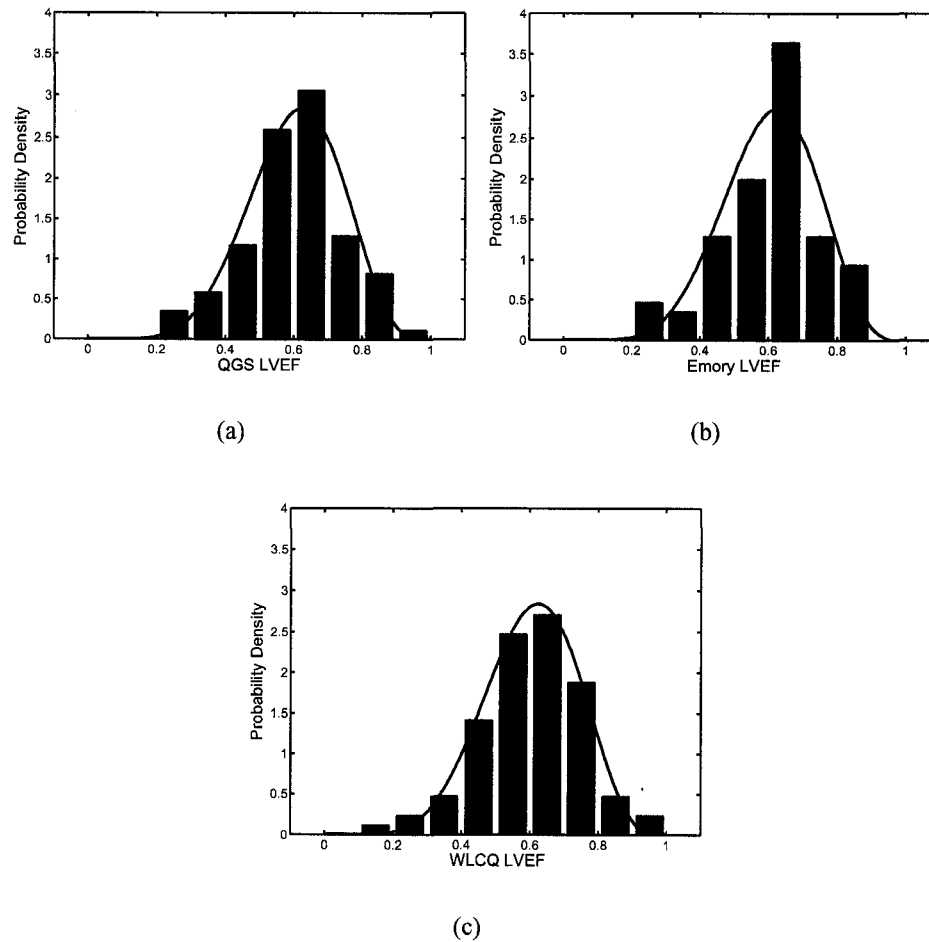
FIGURE 7.2. The results of consistency check 2 applied to the cardiac EF estimation algorithm data set. The plots consist of the beta PDF returned by RWT along with the histogram-med data sets adjusted according to 7.2.

TABLE 7.2. Results of the consistency check for the sample covariance using the results of RWT applied to the EF data gathered using QGS, Emory, and WLCQ techniques.

Sample Covariances	Emory vs. QGS	WLCQ vs. QGS	WLCQ vs Emory
Est. directly from data	0.0210	0.0236	0.0210
Est. w/RWT: beta	0.0212	0.0237	0.0212
Est. w/RWT: trunc. norm.	0.0245	0.0274	0.0245

estimated to be somewhere between 20 and 80 percent in Africa [73]. All data indicate that the presence of HHV-8 in an individual is a necessary, though not sufficient, condition for the onset of a cancer called Kaposi's sarcoma (KS). KS is the disease responsible for the dark skin lesions associated with AIDS.

A problem confronting epidemiologists studying the HHV-8 is the lack of a gold standard assay test to determine the presence of HHV-8 [74–76]. The goal of such assays is to classify blood samples as virus present or virus absent. The assays, however, are not binary decision variables, rather they give an estimate of the level of infection for a given patient represented by the continuous variable "viral load". Thus HHV-8 assays can be viewed as both an estimation and a classification task.

A study was performed using blood samples taken from KS patients and a control group and the classification task performance of multiple assays were assessed using AUC as the figure of merit [75]. We would like to rank the different assays "viral load" estimation performance using RWT and compare this ranking to the ranking arrived at using classification task performance. A potential goal of epidemiologists at the National Cancer Institute is to design a classification scheme consisting of multiple assays. Recently, models have been developed defining classifiers that take into account multiple covariates [77]. We are interested in defining a classifier based on information gathered from performing RWT and comparing the two.

Appendix A

A.1 The beta distributions

The PDF for the beta distribution has the form

$$pr(\Theta_p | \nu, \omega) = \frac{\Theta_p^{\nu-1}(1 - \Theta_p)^{\omega-1}}{B(\nu, \omega)} \tag{A.1}$$

where

$$B(\nu, \omega) = \int_0^1 d\Theta_p \Theta_p^{\nu-1}(1 - \Theta_p)^{\omega-1}. \tag{A.2}$$

A.2 The truncated normal distribution

The PDF for the truncated normal has the form

$$pr(\Theta_p | \mu, \sigma_a) = A(\mu, \sigma_a) \exp\left(-\frac{1}{2\sigma_a^2}(\Theta_p - \mu)^2\right) \Pi(x), \tag{A.3}$$

where $\Pi(x)$ is an indicator function with value one on the interval $[0, 1]$ and zero elsewhere. We use the subscript a in equ. A.3 so as to differentiate this σ from the linear model parameters $\boldsymbol{\sigma}$. The normalizing constant $A(\mu, \sigma_a)$ is given as

$$A(\mu, \sigma_a) = \sqrt{\frac{\pi}{2}} \sigma_a \left(erf\left(\frac{1-\mu}{\sqrt{2}\sigma_a}\right) + erf\left(\frac{\mu}{\sqrt{2}\sigma_a}\right) \right) \tag{A.4}$$

and

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt \exp\left(-t^2\right). \tag{A.5}$$

Appendix B

B.1 Derivatives of the log-likelihood

As derived in Sec. 5.2, the components of the Fisher information matrix for our problem are given by

$$J_{ij} = P \int_R d^M \theta_p \left(\frac{1}{pr(\theta_p|\eta)} \frac{\partial pr(\theta_p|\eta)}{\partial \eta_i} \frac{\partial pr(\theta_p|\eta)}{\partial \eta_j} \right) \Bigg|_{\eta=\eta_o},$$

where

$$pr(\theta_p|\eta) = \int_R pr(\Theta_p|r) \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_p - a_m\Theta_p - b_m)^2 \right).$$

The expressions for the partial derivatives of $pr(\theta_p|\eta)$ with respect to the linear model parameters are as follows:

$$\frac{\partial pr(\theta_p|\eta)}{\partial a_i} = \int_R d\Theta_p pr(\Theta_p|r) \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2 \right)$$
$$\cdot \left(\frac{\Theta_p}{\sigma_i^2}(\theta_{pi} - a_i\Theta_p - b_i) \right)$$

$$\frac{\partial pr(\theta_p|\eta)}{\partial b_i} = \int_R d\Theta_p pr(\Theta_p|r) \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2 \right)$$
$$\cdot \left(\frac{1}{\sigma_i^2}(\theta_{pi} - a_i\Theta_p - b_i) \right)$$

$$\frac{\partial pr(\theta_p|\eta)}{\partial \sigma_i} = \int_R d\Theta_p pr(\Theta_p|r) \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2 \right)$$
$$\cdot \left(\frac{1}{\sigma_i^3}(\theta_{pi} - a_i\Theta_p - b_i)^2 - \frac{1}{\sigma_i} \right).$$

For the linear model parameters these calculations were straightforward. The calculations are more involved in the case of the parameters of the assumed distribution. Below we show the partial derivatives with respect to r; the parameters of the assumed distribution. When the assumed distribution is from the beta family, the partial derivative of the beta PDF with respect to ν is

$$\frac{\partial pr(\Theta_p|\nu,\omega)}{\partial \nu} = pr(\Theta_p|\nu,\omega)\left(\ln(\Theta_p) - \frac{\partial B(\nu,\omega)}{\partial \nu} \Big/ B(\nu,\omega)\right),$$

and thus

$$\frac{\partial pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})}{\partial \nu} = \int_R d\Theta_p\, pr(\Theta_p|\nu,\omega) \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2\right)$$
$$\cdot \left(\ln(\Theta_p) - \frac{\partial B(\nu,\omega)}{\partial \nu} \Big/ B(\nu,\omega)\right).$$

The partial derivative of the normalizing constant with respect to ν is given by

$$\frac{\partial B(\nu,\omega)}{\partial \nu} = \frac{\partial}{\partial \nu} \int_0^1 d\Theta_p\, \Theta_p^{\nu-1}(1-\Theta_p)^{\omega-1}$$
$$= \int_0^1 \ln(\Theta_p)\Theta_p^{\nu-1}(1-\Theta_p)^{\omega-1},$$

which is an integral we solve numerically. In similar fashion, the partial derivative with respect to ω is given as

$$\frac{\partial pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})}{\partial \omega} = \int_R d\Theta_p\, pr(\Theta_p|\nu,\omega) \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp\left(\sum_{m=1}^{M} -\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2\right)$$
$$\cdot \left(\ln(1-\Theta_p) - \frac{\partial B(\nu,\omega)}{\partial \omega} \Big/ B(\nu,\omega)\right),$$

where

$$\frac{\partial B(\nu,\omega)}{\partial \omega} = \frac{\partial}{\partial \omega} \int_0^1 d\Theta_p\, \Theta_p^{\nu-1}(1-\Theta_p)^{\omega-1}$$
$$= \int_0^1 \ln(1-\Theta_p)\Theta_p^{\nu-1}(1-\Theta_p)^{\omega-1}.$$

We solve this integral numerically as well.

The partial derivatives in the case when the assumed distribution is from the truncated normal family follows in much the same way as those in the case of the beta. The two parameters of interest are the mean μ and the standard deviation σ_a. Differentiating the truncated normal with respect to μ yields

$$\frac{\partial pr(\Theta_p|\mu,\sigma_a)}{\partial \mu} = pr(\Theta_p|\mu,\sigma_a)\left(\frac{\partial A(\mu,\sigma_a)}{\partial \mu}\bigg/ A(\mu,\sigma_a) + \frac{1}{\sigma_a^2}(\Theta_p - \mu)\right),$$

from which it follows that

$$\begin{aligned}\frac{\partial pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})}{\partial \mu} &= \int_R d\Theta_p pr(\Theta_p|\mu,\sigma_a)\prod_{m=1}^{M}\frac{1}{\sqrt{2\pi\sigma_m^2}}\exp\left(\sum_{m=1}^{M}-\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2\right)\\ &\cdot \left(\frac{\partial A(\mu,\sigma_a)}{\partial \mu}\bigg/ A(\mu,\sigma_a) + \frac{1}{\sigma_a^2}(\Theta_p - \mu)\right).\end{aligned}$$

The partial derivative of the normalizing constant $A(\mu,\sigma_a)$ with respect to μ is given by

$$\frac{\partial A(\mu,\sigma_a)}{\partial \mu} = \exp\left(-\left(\frac{\mu}{\sqrt{2}\sigma_a}\right)^2\right) - \exp\left(-\left(\frac{1-\mu}{\sqrt{2}\sigma_a}\right)^2\right).$$

For σ_a we have

$$\begin{aligned}\frac{\partial pr(\boldsymbol{\theta}_p|\boldsymbol{\eta})}{\partial \sigma_a} &= \int_R d\Theta_p pr(\Theta_p|\mu,\sigma_a)\prod_{m=1}^{M}\frac{1}{\sqrt{2\pi\sigma_m^2}}\exp\left(\sum_{m=1}^{M}-\frac{1}{2\sigma_m^2}(\theta_{pm} - a_m\Theta_p - b_m)^2\right)\\ &\cdot \left(\frac{\partial A(\mu,\sigma_a)}{\partial \mu}\bigg/ A(\mu,\sigma_a) + \frac{1}{\sigma_a^3}(\Theta_p - \mu)^2\right)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial A(\mu,\sigma_a)}{\partial \sigma_a} &= \sqrt{\frac{\pi}{2}}\left(erf\left(\frac{1-\mu}{\sqrt{2}\sigma_a}\right) + erf\left(\frac{\mu}{\sqrt{2}\sigma_a}\right)\right)\\ &+ \frac{1}{\sigma_a}\left(\exp\left(-\left(\frac{1-\mu}{\sqrt{2}\sigma_a}\right)^2\right)(\mu - 1) - \exp\left(-\left(\frac{\mu}{\sqrt{2}\sigma_a}\right)^2\right)\mu\right).\end{aligned}$$

B.2 The Likelihood, Λ, in the Case of a Normal Assumed Distribution for M=3

As derived earlier, the likelihood, Λ, in the case of a normal assumed distribution may be written as

$$\Lambda = \prod_{p=1}^{P}\left(N\exp\left(-\frac{1}{2}(\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \boldsymbol{K}^{-1}(\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)\right)\right).$$

In the case when $M=3$

$$N = \frac{1}{(2\pi)^{\frac{3}{2}} \det{(K)}^{\frac{1}{2}}},$$

$$K^{-1} = \begin{bmatrix} \frac{Aa_1^2}{2\sigma_1^2} - \frac{1}{\sigma_1^2} & \frac{Aa_1a_2}{2\sigma_1^2\sigma_2^2} & \frac{Aa_1a_3}{2\sigma_1^2\sigma_3^2} \\[2mm] \frac{Aa_1a_2}{2\sigma_1^2\sigma_2^2} & \frac{Aa_2^2}{2\sigma_2^2} - \frac{1}{\sigma_2^2} & \frac{Aa_2a_3}{2\sigma_2^2\sigma_3^2} \\[2mm] \frac{Aa_1a_3}{2\sigma_1^2\sigma_3^2} & \frac{Aa_2a_3}{2\sigma_2^2\sigma_3^2} & \frac{Aa_3^2}{2\sigma_3^2} - \frac{1}{\sigma_3^2} \end{bmatrix}$$

and

$$\overline{\theta}_p = \frac{K\beta}{4},$$

where β is a 3×1 vector given by

$$\beta = \frac{1}{A} \begin{bmatrix} \frac{\mu a_1}{2\sigma_a\sigma_1^2} - \frac{a_1^2 b_1}{2\sigma_1^4} - \frac{a_1a_2b_2}{2\sigma_1^2\sigma_2^2} - \frac{a_1a_3b_3}{2\sigma_1^2\sigma_3^2} + \frac{Ab_1}{\sigma_1^2} \\[2mm] \frac{\mu a_2}{2\sigma_a\sigma_2^2} - \frac{a_2^2 b_2}{2\sigma_2^4} - \frac{a_1a_2b_1}{2\sigma_1^2\sigma_2^2} - \frac{a_2a_3b_3}{2\sigma_2^2\sigma_3^2} + \frac{Ab_2}{\sigma_2^2} \\[2mm] \frac{\mu a_2}{2\sigma_a\sigma_3^2} - \frac{a_3^2 b_3}{2\sigma_3^4} - \frac{a_1a_3b_1}{2\sigma_1^2\sigma_3^2} - \frac{a_2a_3b_3}{2\sigma_2^2\sigma_3^2} + \frac{Ab_3}{\sigma_3^2} \end{bmatrix}.$$

The expression for A is given in equ. 5.15.

B.3 Derivation of J_{ij} in the case of a normal assumed distribution

In this appendix we calculate the components of the Fisher information for our problem when the assumed distribution is a normal.

$$\begin{aligned} J_{ij} &= -\left\langle \frac{\partial^2}{\partial\eta_i\partial\eta_j}\lambda \right\rangle_{\{\theta_p\}}\Bigg|_{\eta=\eta_o} \\[3mm] &= -\left\langle \frac{\partial^2}{\partial\eta_i\partial\eta_j}\left(\sum_{p=1}^{P}\log\left(N\exp\left(-\frac{1}{2}\left(\theta_p-\overline{\theta}_p\right)^t K^{-1}\left(\theta_p-\overline{\theta}_p\right)\right)\right)\right)\right\rangle_{\{\theta_p\}}\Bigg|_{\eta=\eta_o} \\[3mm] &= -P\left(\frac{\partial^2}{\partial\eta_i\partial\eta_j}(\log(N)) - \frac{1}{2}\left\langle \frac{\partial^2}{\partial\eta_i\partial\eta_j}\left(\left(\theta_p-\overline{\theta}_p\right)^t K^{-1}\left(\theta_p-\overline{\theta}_p\right)\right)\right\rangle_{\theta_p}\right)\Bigg|_{\eta=\eta_o} \end{aligned}$$

The partial derivative of the remaining expected value consists of nine different terms, six of them are equal to zero. For example,

$$\left\langle \frac{\partial}{\partial \eta_i} \left((\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \right) \frac{\partial}{\partial \eta_j} \left(\boldsymbol{K}^{-1} \right) (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \right\rangle_{\boldsymbol{\theta}_p} = \frac{\partial}{\partial \eta_i} \left(\overline{\boldsymbol{\theta}}_p^t \right) \frac{\partial}{\partial \eta_j} \left(\boldsymbol{K}^{-1} \right) \left\langle (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \right\rangle_{\boldsymbol{\theta}_p}$$
$$= 0$$

There are two non-zero terms of the form

$$\left\langle \frac{\partial}{\partial \eta_i} \left((\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \right) \left(\boldsymbol{K}^{-1} \right) \frac{\partial}{\partial \eta_j} (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \right\rangle_{\boldsymbol{\theta}_p} = \frac{\partial}{\partial \eta_i} \left(\overline{\boldsymbol{\theta}}_p^t \right) \left(\boldsymbol{K}^{-1} \right) \frac{\partial}{\partial \eta_j} \left(\overline{\boldsymbol{\theta}}_p \right) ,$$

and the final non-zero term is given by

$$\left\langle (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \frac{\partial^2}{\partial \eta_i \partial \eta_j} \left(\boldsymbol{K}^{-1} \right) (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \right\rangle_{\boldsymbol{\theta}_p}$$
$$= \int_{-\infty}^{\infty} d^M \theta_p \, (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \frac{\partial^2}{\partial \eta_i \partial \eta_j} \left(\boldsymbol{K}^{-1} \right) (\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p) \, pr(\boldsymbol{\theta}_p | \boldsymbol{\eta})$$
$$= \int_{-\infty}^{\infty} d^M \theta_p \mathrm{tr} \left((\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)(\boldsymbol{\theta}_p - \overline{\boldsymbol{\theta}}_p)^t \frac{\partial^2}{\partial \eta_i \partial \eta_j} \left(\boldsymbol{K}^{-1} \right) \right) pr(\boldsymbol{\theta}_p | \boldsymbol{\eta})$$
$$= \mathrm{tr} \left(\boldsymbol{K} \frac{\partial^2}{\partial \eta_i \partial \eta_j} \left(\boldsymbol{K}^{-1} \right) \right) .$$

Thus our expression for J_{ij} is

$$J_{ij} = -P \left(\frac{\partial^2}{\partial \eta_i \partial \eta_j} (\log(N)) \right.$$
$$\left. - \frac{1}{2} \mathrm{tr} \left(\boldsymbol{K} \frac{\partial^2}{\partial \eta_i \partial \eta_j} \left(\boldsymbol{K}^{-1} \right) \right) - \frac{\partial}{\partial \eta_i} \left(\overline{\boldsymbol{\theta}}_p \right) \left(\boldsymbol{K}^{-1} \right) \frac{\partial}{\partial \eta_j} \left(\overline{\boldsymbol{\theta}}_p \right) \right) \Bigg|_{\boldsymbol{\eta} = \boldsymbol{\eta}_o} .$$

REFERENCES

[1] H. H. Barrett, "Objective assessment of image quality: Effects of quantum noise and object variability," *Journal of the Optical Society of America A*, vol. 7, no. 7, pp. 1266–1278, 1990.

[2] H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, "Objective assessment of image quality. ii. Fisher information, Fourier crosstalk, and figures of merit for task performance," *Journal of the Optical Society of America A*, vol. 12, no. 5, pp. 834–852, 1995.

[3] H. H. Barrett, C. K. Abbey, and E. Clarkson, "Objective assessment of image quality: III. ROC metrics, ideal observers, and likelihood-generating functions," *Journal of the Optical Society of America A*, vol. 15, no. 6, pp. 1520–1535, 1998.

[4] C. E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, no. 9, pp. 720–733, 1986.

[5] W. Sturtridge, B. Lentle, and D. A. Hanley, "Prevention and management of osteoporosis: Consensus statements from the scientific advisory board of the osteoporosis society of canada," *Canadian Medical Association Journal*, vol. 155, pp. 924–929, 1996.

[6] H. Al-Hallaq, J. N. River, M. Zamora, H. Oikawa, and G. S. Karczmar, "Correlation of magnetic resonance and oxygen microelectrode measurements of carbogen-induced changes in tumor oxygenation," *International Journal of Radiotion Oncology, Biology, and Physics*, vol. 41, no. 1, pp. 151–159, 1998.

[7] T. Sharir, G. Germano, X. Kang, H. C. Lewin, R. Miranda, I. Cohen, R. D. Agafitei, J. D. Friedman, and D. S. Berman, "Prediction of myocardial infarction versus cardiac death by gated myocardial perfusion SPECT: Risk stratification by the amount of stress-induced ischemia and the poststress ejection fraction," *Journal of Nuclear Medicine*, vol. 42, no. 6, pp. 831–837, 2001.

[8] A. Achtert, M. A. King, S. T. Dahlberg, P. H. Pretorius, K. H. LaCroix, and B. M. W. Tsui, "An investigation of the estimation of ejection fractions and cardiac volumes by a quantitative gated SPECT software package in simulated SPECT images," *Journal of Nuclear Cardiology*, vol. 5, pp. 144–152, March/August 1998.

[9] C. Vanhove and P. R. Franken, "Left ventricular ejection fraction and volumes from gated blood pool tomography: Comparison between two automatic algorithms that work in three-dimensional space," *Journal of Nuclear Cardiology*, vol. 8, pp. 466–471, July/August 2001.

[10] J. A. Rumberger, T. Behrenbeck, M. R. Bell, J. F. Breen, D. L. Johnston, D. R. Holmes, and M. Enriquez-Sarano, "Determination of ventricular ejection fraction: A comparison of available imaging methods," in *Mayo Clinic Proceedings*, vol. 72, pp. 860–870, September 1997.

[11] S. A. Feig, "Estimation of currently attainable benefit from mammographic screening in women aged 40-49," *Cancer*, vol. 75, pp. 2412–2419, 1995.

[12] S. D. Walter and L. M. Irwig, "Estimation of test error rates, disease prevalence, and relative risk from misclassified data: A review," *Journal of Clinical Epidemiology*, vol. 41, pp. 923–937, 1988.

[13] R. M. Henkelman, I. Kay, and M. J. Bronskill, "Receiver operator characteristic (ROC) analysis without truth," *Medical Decision Making*, vol. 10, pp. 24–29, 1990.

[14] S. V. Beiden, G. Campbell, K. L. Meier, and R. F. Wagner, "On the problem of ROC analysis without truth: The EM algorithm and the information matrix," in *Medical Imaging 2000: Image Perception and Performance*, vol. 3981, pp. 126–134, SPIE, 2000.

[15] P. S. Albert, L. M. McShane, and J. H. Shih, "Latent class modeling approaches for assessing diagnostic error without a gold standard: With applications to p53 immunohistochemical assays in bladder tumors," *Biometrics*, vol. 57, pp. 610–619, June 2001.

[16] H. L. K. M. Polansky, "Comparing observer performance with mixture distribution analysis when there is no external gold standard," in *Medical Imaging 1998: Image Perception* (H. L. Kundel, ed.), vol. 3340, pp. 78–84, SPIE, 1998.

[17] H. L. Kundel and M. Polansky, "Mixture distribution and receive operating characteristic analysis of bedside chest imaging with screen-film and computed radiography," *Academic Radiology*, vol. 4, pp. 1–7, January 1997.

[18] H. L. Kundel, M. Polansky, and M. Phelan, "Evaluating imaging systems in the absence of truth: A comparison of ROC and mixture distribution analysis in computer aided diagnosis in mammography," in *Medical Imaging 2001: Image Perception and Performance* (E. A. Krupinski, ed.), vol. 4324, pp. 153–158, SPIE, 2001.

[19] S. V. Faraone and M. T. Tsuang, "Measuring diagnostic accuracy in the absence of a "gold standard"," *American Journal of Psychiatry*, vol. 151, pp. 650–657, May 1984.

[20] Y. Qu, M. Tan, and M. H. Kutner, "Random effects models in latent class analysis for evaluating accuracy of diagnostic tests," *Biometrics*, vol. 52, pp. 797–810, September 1996.

[21] M. Abe, Y. Kazatani, H. Fukuda, H. Tatsuno, H. Habara, and H. Shinbata, "Left ventricular volumes, ejection fraction, and regional wall motion calculated with gated technetium-99m tetrofosmin SPECT in reperfused acute myocardial infarction at super-acute phase: Comparison with left ventriculography," *Journal of Nuclear Cardiology*, vol. 7, pp. 569–574, November/December 2000.

[22] E. Cwajg, J. Cwajg, Z.-X. He, W. S. Hwang, F. Keng, S. F. Nagueh, and M. S. Verani, "Gated myocardial perfussion tomography for the assessment of left ventricular function and volumes: Comparison with echocardiography," *Journal of Nuclear Medicine*, vol. 40, no. 11, pp. 1857–1865, 1999.

[23] T. L. Faber, J. Vansant, R. I. Pettigrew, J. R. Galt, M. Blais, G. Chatzimavroudis, C. D. Cooke, R. D. Folks, S. M. Waldrop, E. Guartler-Krawczynska, M. D. Wittry, and E. V. Garcia, "Evaluation of left ventricular endocardial volumes and ejection fractions computed from gated perfusion SPECT with magnetic resonance imaging: Comparison of two methods," *Journal of Nuclear Cardiology*, vol. 8, pp. 645–651, November/December 2001.

[24] P. Vaduganathan, Z. He, W. V. III, J. J. Mahmarian, and M. S. Verani, "Evaluation of left ventricular wall motion, volumes, and ejection fraction by gated myocardial tomography with technitium 99m-labeled tetrofism: A comparison with cine magnetic imaging," *Journal of Nuclear Cardiology*, vol. 6, pp. 3–10, January/February 1999.

[25] Z. He, E. Cwajg, J. S. Presian, J. J. Mahmarian, and M. S. Verani, "Accuracy of left ventricular ejection fraction determined by gated myocardial perfusion SPECT with Tl-201 and Tc-99m sestamibi: Comparison with first-pass radionuclide angiography," *Journal of Nuclear Cardiology*, vol. 6, pp. 412–417, July/August 1999.

[26] N. G. Bellenger, M. I. Burgess, S. G. Ray, A. Lahiri, A. J. S. Coats, J. G. F. Cleland, and D. J. Pennell, "Comparison of left ventricular ejection fraction and volumes in heart failure by echocardiography, radionuclide ventriculography and cardiovascular magnetic resonance," *European Heart Journal*, vol. 21, pp. 1387–1396, August 2000.

[27] J. Hoppin, M. Kupinski, G. Kastis, E. Clarkson, and H. H. Barrett, "Objective comparison of quantitative imaging modalities without the use of a gold standard," in *Lecture Notes in Computer Science: Information Processing in Medical Imaging* (M. Insana and R. Leahy, eds.), Springer, 2001.

[28] J. W. Hoppin, M. A. Kupinski, G. A. Kastis, E. Clarkson, and H. H. Barrett, "Objective comparison of quantitative imaging modalities without the use of a gold standard," *IEEE Transactions on Medical Imaging*, vol. 21, pp. 441–450, May 2002.

[29] M. A. Kupinski, J. W. Hoppin, E. Clarkson, H. H. Barrett, and G. A. Kastis, "Estimation in medical imaging without a gold standard," *Academic Radiology*, vol. 9, pp. 290–297, March 2002.

[30] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946.

[31] C. R. Rao, *Linear Statistical Inference and Its Applications*. John Wiley and Sons, second ed., 1973.

[32] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Part I*. New York: John Wiley & Sons, 1968.

[33] R. Lacayo, "Heart murmurs." Time.com:Pacific, December 2000.

[34] "American Heart Association website: Cardiovascular disease statistics," 2000.

[35] D. G. Greene, R. Carlisle, C. Grant, and I. L. Bunnell, "Estimation of left ventricular volume by one-plane cineangiography," *Circulation*, vol. 35, pp. 61–69, 1967.

[36] B. R. Chaitman, H. DeMots, D. Bristow, J. Rösch, and S. H. Rahimtoola, "Objective and subjective analysis of left ventricular angiograms," *Circulation*, vol. 52, pp. 420–425, September 1975.

[37] P. F. Cohn, J. A. Levine, G. A. Bergeron, and R. Gorlin, "Reproducibility of the angiographic left ventricular ejection fraction in patients with coronary artery disease," *American Heart Journal*, vol. 88, pp. 713–720, December 1974.

[38] H. T. Dodge, H. Sandler, D. W. Ballew, and H. D. Lord, "The use of biplane angiocardiography for the measurement of left ventricular volume in man," *American Heart Journal*, vol. 60, pp. 762–776, 1960.

[39] A. F. Amico, G. S. Lichtenberg, S. A. Reisner, C. K. Stone, R. G. Schwartz, and R. S. Meltzer, "Superiority of visual versus computerized echocardiographic estimation of radionuclide left ventricular ejection fraction," *American Heart Journal*, vol. 118, pp. 1259–1265, 1989.

[40] N. B. Schiller, P. M. Shah, M. Crawford, A. DeMaria, R. Devereux, H. Feigenbaum, H. Gutgesell, N. Reicheck, D. Sahn, I. Schnittger, N. H. Silverman, and A. J. Tajik, "Recommendations for quantitation of the left ventricle by two-dimensional echocardiography," *Journal of the American Society of Echocardiography*, vol. 2, pp. 358–367, 1989.

[41] D. Patton. Personal Communications, April 2003.

[42] D. P. Boyd and M. J. Lipton, "Cardiac computed tomography," in *Proceedings of the IEEE*, vol. 71, pp. 298–307, IEEE, 1983.

[43] D. G. Altman and J. M. Bland, "Measurement in medicine: the analysis of method comparison studies," *The Statistician*, vol. 32, pp. 307–313, 1983.

[44] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *Lancet*, vol. i, pp. 307–310, 1986.

[45] J. O. Westgard and M. R. Hunt, "Use and interpretation of common statistical tests in method-comparison studies," *Clinical Chemistry*, vol. 19, no. 1, pp. 49–57, 1973.

[46] G. Kastis, *Multi-Modality Imaging of Small Animals*. PhD thesis, University of Arizona, 2002.

[47] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, eds., *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, NY, 1995.

[48] R. L. Burden and J. D. Faires, *Numerical Analysis*. PWS Publishing Company, fifth ed., 1993.

[49] S. Kullback, *Information Theory and Statistics*. Dover Publications, Mineola, NY, 1968.

[50] H. Scheffé, *The Analysis of Variance*. John Wiley & Sons, Inc., 1959.

[51] G. Verbeke, G. Molenberghs, L. Duchateau, P. Janssen, L. Bijnens, and D. Shaw, *Linear Mixed Models in Practice: A SAS Oriented Approach*. No. 126 in Lecture Notes in Statistics, New York: Springer-Verlag, 1997.

[52] B. S. Everitt, *An Introduction to Latent Variable Models*. Chapman and Hall, 1984.

[53] A. L. McCutcheon, *Latent Class Analysis*. Quantitative Applications in the Social Sciences, Sage, 1987.

[54] J. Shao, *Mathematical Statistics*. Springer, 1999.

[55] U. Müller-Funk, F. Pukelsheim, and H. Witting, "On the attainment of the Cramér-Rao bound in L_r-differentiable families of distributions," *The Annals of Statistics*, vol. 17, no. 4, pp. 1742–1748, 1989.

[56] J. L. Denny. Personal Communication, May 2002.

[57] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular Fisher information matrices," *IEEE Transactions on Signal Processing*, vol. 49, pp. 87–90, 2001.

[58] G. A. Kastis, L. R. Furenlid, D. W. Wilson, T. E. Peterson, H. B. Barber, and H. H. Barrett, "Compact CT/SPECT small-animal imaging system," in *Nuclear Science Symposium and Medical Imaging Conference*, IEEE, November 2002.

[59] G. A. Kastis, H. B. Barber, H. H. Barrett, S. J. Balzer, D. Lu, D. G. Marks, G. Stevenson, J. M. Woolfenden, M. Appleby, and J. Tueller, "Gamma-ray imaging using a cdznte pixel array and a high-resolution, parallel-hole collimator," *IEEE Transactions on Nuclear Science*, vol. 47, pp. 1923–1927, December 2000.

[60] G. A. Kastis, M. C. Wu, S. J. Balzer, D. W. Wilson, L. R. Furenlid, G. Stevenson, H. B. Barber, H. H. Barrett, J. M. Woolfenden, P. Kelly, and M. Appleby, "Tomographic small-animal imaging using a high-resolution semiconductor camera," *IEEE Transactions on Nuclear Science*, vol. 49, pp. 172–175, February 2002.

[61] S. J. Balzer, "A portable gamma-ray imager for small animal studies," Master's thesis, University of Arizona, 2002.

[62] H. H. Barrett and W. Swindell, *Radiological Imaging: The Theory of Image Formation, Detection, and Processing*. San Diego: Academic Press, 1991.

[63] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Transactions on Medical Imaging*, vol. MI-1, pp. 113–122, October 1982.

[64] J. Hoppin, M. A. Kupinski, D. W. Wilson, T. Peterson, B. Gershman, G. Kastis, E. Clarkson, L. Furenlid, and H. H. Barrett, "Evaluating estimation techniques in medical imaging without a gold standard: Experimental validation," in *Medical Imaging 2003: Image Perception and Performance Proceedings*, pp. 230–237, SPIE, 2003.

[65] H. H. Barrett, D. W. Wilson, and B. M. Tsui, "Noise properties of the EM algorithm: I. theory," *Physics in Medicine and Biology*, vol. 39, pp. 833–846, May 1994.

[66] D. W. Wilson, B. M. Tsui, and H. H. Barrett, "Noise properties of the EM algorithm: II. Monte Carlo simulations," *Physics in Medicine and Biology*, vol. 39, pp. 847–871, May 1994.

[67] K. Nakajima, T. Higuchi, J. Taki, M. Kawano, and N. Tonami, "Accuracy of ventricular volume and ejection fraction measured by gated myocardial SPECT: comparison of 4 software programs," *The Journal of Nuclear Medicine*, vol. 42, pp. 1571–1578, 2001.

[68] G. Germano, H. Kiat, P. B. Kavanagh, M. Moriel, M. Mazzanti, H. Su, K. V. Train, and D. S. Berman, "Automatic quantification of ejection fraction from gated myocardial spect," *The Journal of Nuclear Medicine*, vol. 36, no. 11, pp. 2138–2147, 1995.

[69] T. L. Faber, M. S. Akers, R. M. Peshock, and J. R. Corbett, "Three-dimensional motion and perfusion quantification in gated single-photon emission computed tomograms," *The Journal of Nuclear Medicine*, vol. 32, pp. 2311–2317, 1991.

[70] E. Vallejo, D. P. Dione, A. J. Sinusas, and F. J. T. Wackers, "Assessment of left ventricular ejection fraction with quantitative gated SPECT: Accuracy and correlation with first-pass radionuclide angiography," *Journal of Nuclear Medicine*, vol. 41, pp. 874–882, 2000.

[71] S. Kirac, F. J. T. Wackers, and Y.-H. Lui, "Validation of the yale circumferential quantification method using 201 Tl and 99mTc: A phantom study," *Journal of Nuclear Medicine*, vol. 41, pp. 1436–1441, 2000.

[72] Y. Chang, E. Cesarman, M. S. Pessin, F. Lee, J. Culpeper, D. M. Knowles, and P. S. Moore, "Identification of herpesvirus-like DNA sequences in AIDS-associeated Kaposi's sarcoma," *Science*, vol. 266, pp. 1865–1869, 266.

[73] S. M. Mbulaiteye, R. J. Biggar, P. M. Bakaki, R. Pfeiffer, D. Whitby, A. M. Owor, E. Katongole-Mbidde, J. J. Goedert, C. M. Ndugwa, and E. A. Engels, "Human herpesvirus 8 infection and transfusion history in children with sickle cell disease in Uganda," *Unknown*, 2003. submitted.

[74] E. A. Engels, M. D. Sinclair, R. J. Biggar, D. Whitby, P. Ebbesen, J. J. Goedert, and J. L. Gastwirth, "Latent class analysis of human herpesvirus 8 assay performance and infection prevalence in sub-Saharan Africa and Malta," *Int. J. Cancer*, vol. 88, pp. 1003–1008, 2000.

[75] E. A. Engels, D. Whitby, P. B. Goebel, A. Stossel, D. Waters, A. Pintus, L. Contu, R. J. Biggar, and J. J. Goedert, "Identifying human herpesvirus 8 infection: Performance characteristics of serologic assays," *Journal of Acquired Immune Deficiency Syndrome*, vol. 23, pp. 346–354, 2000.

[76] R. J. Biggar, E. A. Engels, D. Whitby, D. H. Kedes, and J. J. Goedert, "Antibody reactivity to latent and lytic antigens to human herpesvirus-8 in longitudinally followed homosexual men," *The Journal of Infectious Diseases*, vol. 187, pp. 12–18, 2003.

[77] R. Pfeiffer, M. H. Gail, and L. Brown, "A mixture model for the distribution of IgG antibodies to helicobacter pylori: application to studying factors that affect prevalence," *Journal of Epidemiology and Biostatistics*, vol. 5, no. 5, pp. 267–275, 2000.