
Comparing Cardiac Ejection Fraction Estimation Algorithms Without a Gold Standard¹

Matthew A. Kupinski, PhD, John W. Hoppin, PhD, Joshua Krasnow, MD, Seth Dahlberg, MD, Jeffrey A. Leppo, MD
Michael A. King, PhD, Eric Clarkson, PhD, Harrison H. Barrett, PhD

Rationale and Objectives. Imaging and estimation of left ventricular function have major diagnostic and prognostic importance in patients with coronary artery disease. It is vital that the method used to estimate cardiac ejection fraction (EF) allows the observer to best perform this task. To measure task-based performance, one must clearly define the task in question, the observer performing the task, and the patient population being imaged. In this report, the task is to accurately and precisely measure cardiac EF, and the observers are human-assisted computer algorithms that analyze the images and estimate cardiac EF. It is very difficult to measure the performance of an observer by using clinical data because estimation tasks typically lack a gold standard. A solution to this “no-gold-standard” problem recently was proposed, called regression without truth (RWT).

Materials and Methods. Results of three different software packages used to analyze gated, cardiac, and nuclear medicine images, each of which uses a different algorithm to estimate a patient’s cardiac EF, are compared. The three methods are the Emory method, Quantitative Gated Single-Photon Emission Computed Tomographic method, and the Wackers-Liu Circumferential Quantification method. The same set of images is used as input to each of the three algorithms. Data were analyzed from the three different algorithms by using RWT to determine which produces the best estimates of cardiac EF in terms of accuracy and precision.

Results and Discussion. In performing this study, three different consistency checks were developed to ensure that the RWT method is working properly. The Emory method of estimating EF slightly outperformed the other two methods. In addition, the RWT method passed all three consistency checks, garnering confidence in the method and its application to clinical data.

Key Words. Cardiac ejection fraction; image quality; gold standard.

© AUR, 2006

The American Heart Association estimates that more than 600,000 people in the United States died of causes related to coronary artery disease in 2000 (1). Imaging and estimation of left ventricular function and volumes have ma-

major diagnostic and prognostic importance in patients with coronary artery disease (2–4). It was shown by Sharir et al (5) that the cardiac death rate increased exponentially as poststress cardiac ejection fraction (EF) decreased. Thus, small errors in estimates of cardiac EF may result in improper treatment because of this exponential relationship. It is vital that the imaging system used to estimate cardiac EF allow the observer to best perform this task.

Cardiac EF is an example of an estimation task in medical imaging. Many imaging modalities can be used to estimate cardiac EF, including nuclear medicine, ultrasound, x-ray, and magnetic resonance imaging systems. In addition, many different computer algorithms and models exist for analyzing images produced by each modality.

Acad Radiol 2006; 13:329–337

¹ From the Optical Sciences Center, The University of Arizona, 1630 East University Blvd, Tucson, AZ 85721 (M.A.K., E.C., H.H.B.); Central Institute for Electronics, Forschungszentrum Jülich GmbH, Germany (J.W.H.); Department of Radiology, University of Massachusetts Medical School, Worcester, MA (J.K., S.D., J.A.L., M.A.K.). Received September 23, 2005; revision received December 1; revision accepted December 2. Supported under National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering grant no. R01-EB002146. **Address correspondence to:** M.A.K. e-mail: kupinski@radiology.arizona.edu

© AUR, 2006

doi:10.1016/j.acra.2005.12.005

Intermodality comparisons (comparisons of different imaging modalities) and interalgorithm comparisons (comparisons of different algorithms) are necessary to determine how to best estimate cardiac EF. We use task-based measures of image quality to compare imaging modalities and algorithms. To measure task-based performance, one must clearly define the task in question, the observer performing the task, and the patient population being imaged. In this report, the task is to accurately and precisely measure cardiac EF, and the observers are human-assisted computer algorithms that analyze the images and estimate cardiac EF. Finally, the patient population is at-risk patients who have had their cardiac EF estimated. The overall goal is to determine which method allows for the most accurate and precise measurement of a patient's cardiac EF.

It is very difficult to measure the performance of an observer by using clinical data. This difficulty arises because estimation tasks typically lack a gold standard (5–9). For example, a patient's true cardiac EF can never be known. However, the patient's cardiac EF can be estimated by using multiple imaging techniques (2,10–17). Researchers historically have been forced to compare estimates that result from one imaging modality with results from a different, more accepted imaging modality (18–21). This comparison is tantamount to correlating the estimates of the two modalities together. This type of analysis shows nothing about how both estimates relate to the gold standard.

A solution to this problem, called the “no-gold-standard problem,” recently was proposed by Kupinski et al (22) and Hoppin et al (23), called regression without truth (RWT). The proposed solution to the no-gold-standard problem was shown and studied in simulation (22,23). We endeavored to make our simulation studies as realistic as possible. However, simulation studies always give rise to questions about real data. We performed phantom studies of the method and found that RWT performs well by using real image data taken from a phantom and using a volume-estimation task (24). In this report, we compare results of three different software packages used to analyze gated, cardiac, and nuclear medicine images, each of which has an algorithm to estimate a patient's cardiac EF. The same set of images is used as input to each of the three algorithms. We analyze data from the three different algorithms by using RWT to determine which produces the best estimates of cardiac EF.

We briefly describe the three different computer algorithms and the RWT method and then present and discuss

our results. We show that RWT returns values consistent with the data, thus further validating the use of RWT. Questions still exist about the usefulness of RWT, which we address further in the Discussion section.

MATERIALS AND METHODS

Cardiac EF Estimation Algorithms

Left ventricular EF (LVEF) was estimated in 122 consecutive patients who underwent single-photon emission computed tomographic (SPECT) perfusion imaging immediately after either exercise or pharmacological stress (25). Perfusion imaging was performed by using 148 MBq (4 mCi) of thallium 201 and performed on a Philips Medical Systems Prism 3000XP SPECT (Amsterdam, The Netherlands) system equipped with low-energy high-resolution parallel-hole collimators on each of its three camera heads. Acquisitions were acquired as 64×64 frames with a pixel size of 6.34 mm at 3° intervals over 120° of rotation for each head. Acquisition time was 26 seconds at each projection angle, and this time was divided into eight intervals during the cardiac cycle by using electrocardiogram gating with tracking of the average gate interval from frame to frame. Projections were two-dimensionally prereconstruction filtered with a Butterworth filter with an order of 5.0 and a cutoff value of 0.2 of the sampling frequency. They were reconstructed by filtered backprojection by using the standard software of the Philips Medical Systems Odyssey Workstation to which the SPECT system was attached.

LVEF was obtained in automatic mode with no operator adjustment by using three commercially available programs. The first was the Quantitative Gated SPECT (QGS) program from Cedars-Sinai (26,27). With QGS, the location of center-of-mass left ventricular counts is detected automatically, radial sampling is used to fit an ellipsoid to the midmyocardial surface, then a cost function is used to determine the endocardial and epicardial surface points associated with the ellipsoid. The enclosed endocardial volumes at end-diastole and end-systole are used to calculate LVEF. The second program used to calculate LVEF was the Emory Tool Box (Emory) (13,28). The Emory program starts by automatically determining in short-axis slices the apex, basal valve plane, and long axis of the left ventricle. A cylindrical coordinate system is used to find the locations on the center of the heart wall in the basal and midventricular regions, and a spherical coordinate system is used to find the center of the

wall in the apical region. A 10-mm myocardial thickness is assumed at end-diastole, with one half of this on either side of the center locations. Wall thickening is determined during the cardiac cycle, and epicardial and endocardial surfaces are determined. LVEF is calculated again from end-diastolic and end-systolic volumes. The third program used to calculate LVEF was the Wackers-Liu Circumferential Quantification (WLCQ) from Yale (29,30). This program determines the location of maximum counts in circumferential count profiles for a series of short-axis slices containing the left ventricle. Then the location of the endocardial surface within this region is determined, the apical region is modeled as a semiellipsoid for which parameters are determined from eight central horizontal long-axis slices, and LVEF is calculated from the enclosed end-diastolic and end-systolic volumes.

Method of Comparison

The RWT method is described in detail in other publications (22,23). We present the method briefly for completeness. The RWT method assumes a parametric relationship between the gold standard EF Θ_p for patient p and the estimate of that patient's cardiac EF denoted by θ_{pm} for the p^{th} patient and m^{th} method of estimation. Although this relationship may be arbitrarily complicated, to date, we have assumed a linear relationship between the gold standard and estimates through the equation:

$$\theta_{pm} = a_m \Theta_p + b_m + \varepsilon_{pm} \quad (1)$$

where ε_{pm} is a zero-mean-noise term characterized by its SD σ_m , and the parameters a_m , b_m , and σ_m compose the linear model parameters that describe how the m^{th} method's estimates relate to the true gold standard. The same set of patients had their cardiac EFs measured by means of all methods. In general, one would not expect a highly nonlinear relationship between the gold standard and the estimation method unless the estimation method produced purely senseless results. Assuming a linear relationship is tantamount to claiming that the method requires some simple calibration to match, on average, the gold standard. We previously showed that a nonlinear relationship can be used with RWT (22).

We assume that the noise term ε_{pm} is distributed as a zero mean Gaussian with an SD of σ_m that is a characteristic of the method of estimation. Other noise models can be accounted for easily by using RWT. We must further assume that the gold standard Θ_p is sampled from a dis-

tribution for which the shape is characterized by the vector of parameters \vec{p} . We previously studied two different types of parameterized distribution: truncated-normal distribution and β distribution (23). The truncated-normal distribution is a Gaussian distribution with a mean μ_θ and a variance σ_θ (i.e. $\vec{p} = [\mu_\theta \ \sigma_\theta]$) that is truncated at less than zero and greater than one. This truncation allows this distribution to represent cardiac EFs that are bound between zero and one. The β distribution (Appendix) also is bound between zero and one, but unlike truncated-normal distribution, can have zero probability density at zero and one for certain values of \vec{p} .

Truncated-normal distribution has the unfortunate property that there is always a non-zero-probability density near a cardiac EF of one and zero (and hence a non-zero probability of obtaining an EF in a differential interval around both zero and one). Both these situations should have a zero-probability density. Never will a patient have a cardiac EF close to one or close to zero. Based on this information and the work of Sharir et al (5) showing distributions of estimated cardiac EFs, we postulate that the β distribution is the more appropriate distribution for our study.

Using the linear model and one of the two distributions characterizing the gold standard, we are able to form a likelihood expression (22) that can be used to perform maximum-likelihood estimation. Maximum-likelihood estimation will return estimates of a_m , b_m , and σ_m for all methods m , as well as an estimate of \vec{p} that characterizes how the gold standard is distributed. Parameters a_m and b_m can be used to "adjust" the estimates θ_{pm} to be closer, on average, to the gold standard. Furthermore, the term σ_m/a_m characterizes the reproducibility (refer to equation 1) of the method and provides a figure of merit used for comparing the different methods. One can argue that the method that produces the most reproducible estimates after being adjusted to be unbiased is the best method.

Analysis

In our previous works (22–24), we were able to compare our estimates of linear model parameters with true linear model parameters because we had the gold standard available to us, either through simulation or from a phantom study. For this study, we are forced to address our lack of a gold standard, as would any researcher using RWT on real data. Therefore, we cannot compare our estimates with the truth because the truth is not available to us. Instead, we are limited to performing consistency checks on the estimates returned by RWT. These consis-

tency checks will not guarantee that the method is working properly. However, if the method does not pass these consistency checks, we know the RWT method is not working properly. Therefore, although we cannot state with certainty that RWT is working properly, we can run numerous studies to ascertain whether the method is not working properly. We run three different consistency checks, labeled consistency check 1 (CC1), CC2, and CC3, described in detail next.

CC1.—The RWT method returns information relating the gold standard to estimates returned by a method (equation 1). However, two different estimation methods share the same gold standard because the same patient population is used in both. Thus, we can use results returned by RWT to represent the relationship between two estimates. Mathematically, this relationship is determined by solving for the gold standard in equation 1 for each of the estimates θ_{pi} and θ_{pj} , resulting in:

$$\theta_{pi} = \frac{a_i}{a_j} \theta_{pj} + \left(b_i - \frac{a_i b_j}{a_j} \right). \tag{2}$$

If the parameters returned by the RWT method accurately relate the gold standard to estimates returned by the various methods, the slopes and intercepts determined by equation 2 should accurately relate the two estimates to each other. The converse of the previous statement is not necessarily true. However, the contrapositive of this statement is true and allows us to state that if the relationship between the estimates is not determined accurately, the relationship between the gold standard and the estimates also is not accurately determined by means of RWT.

CC2.—Another consistency check we can perform involves comparing the distribution of the gold standard returned by means of RWT $pr(\Theta | \vec{p})$ to the histogram of the raw data θ_{pm} for a given method m calibrated to match the gold standard. Specifically, we compare the gold standard density to a histogram of:

$$\frac{\theta_{pm} - b_m}{a_m} \tag{3}$$

for all patients p and for a given method m . These data should not match exactly because equation 3 has extra noise added onto it from the ϵ_{pm} term. However, if the variance of ϵ_{pm} is small for method m , the histograms should match the gold-standard density returned by means

Table 1
Values for σ_m , b_m , σ_m , and σ_m/a_m Estimated by Using RWT and EF Data Sets Estimated by Using the QGS, Emory, and WLCQ Techniques

Estimates Assuming	QGS	Emory	WLCQ
<i>a_m</i>			
β Distribution	1.1537	1.0309	1.1555
Truncated-normal distribution	0.7547	0.6759	0.7560
<i>b_m</i>			
β Distribution	-0.1356	0.0414	-0.0838
Truncated-normal distribution	0.1446	0.2909	0.1968
σ_m			
β Distribution	0.0587	0.0470	0.0769
Truncated-normal distribution	0.0590	0.0464	0.0770
σ_m/a_m			
β Distribution	0.0509	0.0456	0.0665
Truncated-normal distribution	0.0781	0.0686	0.1018

Results are given for RWT assuming either β or truncated-normal distribution. Note that although values of estimates differ under the two different assumptions, ranking of the techniques is conserved.

of RWT. As we show next, all variance terms for the three methods we are evaluating are small.

CC3.—A final consistency check we developed tests the sample covariance of the EF estimates returned by means of the various methods to what is predicted by means of the model used (equation 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 estimates returned by means of different modalities. Using these two aspects of our RWT model, it can be shown that:

$$Cov(\theta_i, \theta_j) = \hat{a}_i \hat{a}_j Var(\Theta), \tag{4}$$

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 means of RWT, $Var(\Theta)$ is the variance of the gold-standard parameter, and $Cov(\theta_i, \theta_j)$ is the covariance of the estimates returned by means of method i and the estimates returned by means of method j . The expectations needed to compute the covariance and variance in equation 4 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 side of equation 4 can be estimated by using the sample covariance and data given. However, the right side of equation 4 is determined completely by

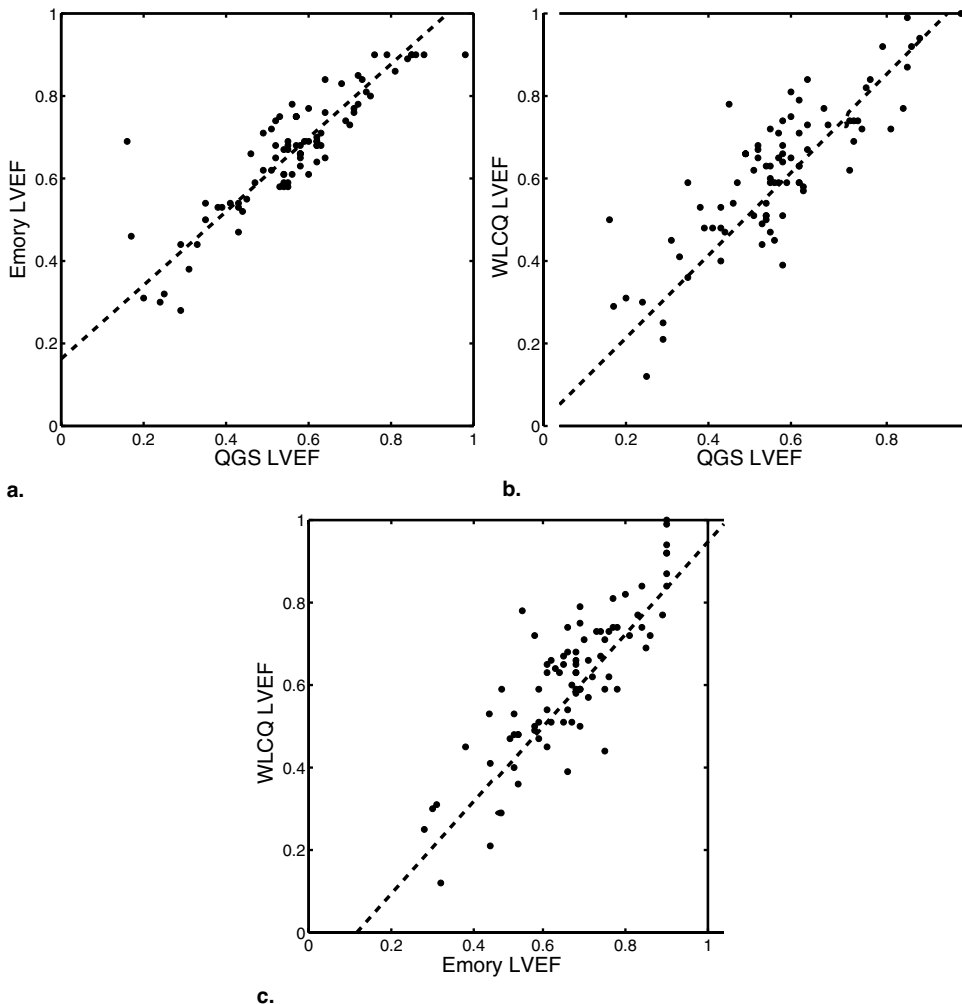


Figure 1. Results of CC1 applied to the cardiac EF estimation algorithm data set. Plots consist of EF estimates of the three different methods plotted against each other. The regression lines shown are NOT the result of conventional regression analysis; rather, they are results of the slopes and intercepts in equation 2 estimated by using results of RWT. The plot shown was generated using results of RWT performed with a β -assumed distribution.

the slopes returned by means of RWT, and the $\vec{\rho}$ determines the variance on the gold standard. Hence, we are relating a data measure to a value determined completely by means of RWT and using the assumptions made by means of RWT. Again, this check can only tell us if the method is not working or our assumptions are invalid; it cannot inform us if the method is working properly.

RESULTS

Recall that the RWT method assumes a linear relationship between the gold standard and the estimates returned by each method. The RWT method also uses a parametric

Table 2
Values for slopes and intercepts of regression plots calculated by using Equation 2 shown in Figure 1, along with the slope and intercepts calculated by using conventional regression analysis

Estimated Using:	Emory vs QGS	WLCQ vs QGS	WLCQ vs Emory
Slopes			
Regression analysis	0.7772	0.8758	0.9994
RWT: β	0.8936	1.0016	1.1209
RWT: truncated-normal	0.8956	1.0017	1.1185
Intercepts			
Regression analysis	0.2278	0.1225	-0.0496
RWT: β	0.1625	0.0519	-0.1302
RWT: truncated-normal	0.1614	0.0519	-0.1286

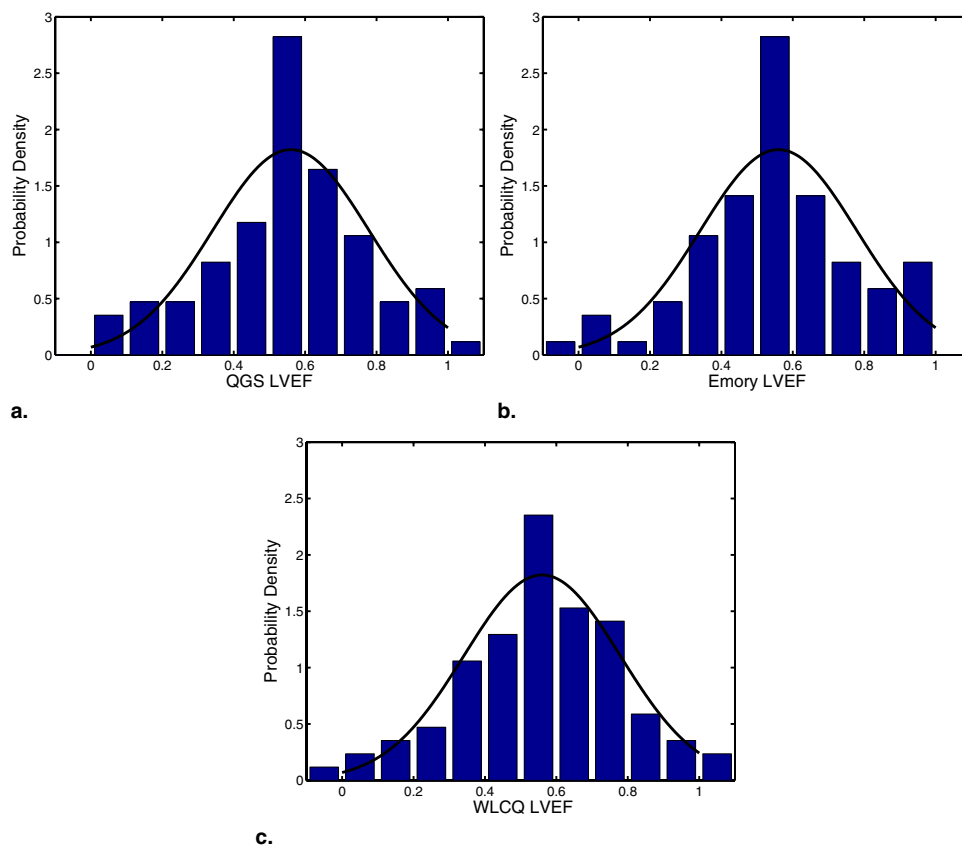


Figure 2. Results of CC2 applied to the cardiac EF estimation algorithm data set by using truncated-normal distribution. Plots consist of the truncated-normal PDF returned by means of RWT (same in each graph), along with the histogram of each data set adjusted according to equation 3.

distribution to characterize the randomness in patients' cardiac EFs. We study two different types of parameterized distribution: truncated-normal distribution and β distribution (23). We performed our RWT analysis on the data set of 85 patients twice; once by using the truncated-normal distribution and once by using the β distribution. Table 1 lists results returned by means of RWT for both studies.

CC1

CC1 checks that the linear model parameters returned by means of RWT are consistent with comparisons made between the estimates from two different methods. This analysis is particularly interesting because researchers often perform regression analysis when developing a new technique. Here, we did not perform regression analysis; we determined the relationship among the three estimation methods and the gold standard and used that information to determine

the relationship among the various estimation methods. Figure 1 shows these results for β distribution. Results for truncated-normal distribution are nearly identical (ie, the RWT-fit lines are indistinguishable from those shown in Figure 1). Thus, both methods return estimates of linear model parameters that are consistent with CC1, although the linear model parameters determined by means of the two methods are different. Table 2 lists these results and the slopes and intercepts returned by using conventional regression analysis. One would not expect results obtained by using RWT to match exactly results obtained by using conventional regression analysis; however, we found that results were close for both the β and truncated-normal distributions.

CC2

CC2 checks the distribution of the gold standard determined by both the model used (β or truncated-normal)

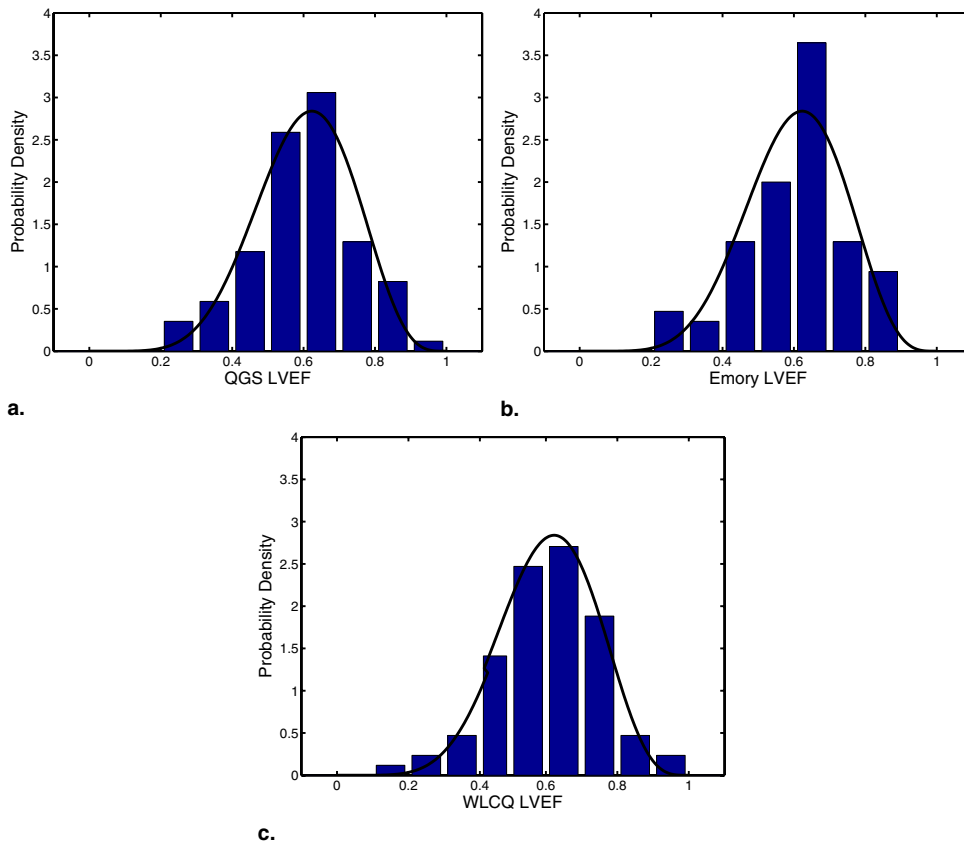


Figure 3. Results of CC2 applied to the cardiac EF estimation algorithm data set by using β distribution. Plots consist of the β PDF returned by means of RWT (same in each graph), along with the histogram of each data set adjusted according to equation 3.

and parameters returned by means of RWT that characterize the shape of these distributions. CC1 validated the estimated linear model parameters. This consistency check attempts to validate the estimated parameters characterizing the gold-standard distribution. Figures 2 and 3 show the gold-standard densities using the estimated parameters and the histograms of the raw data adjusted by using equation 3 for the truncated-normal distribution and β distribution, respectively. It is clear from Figure 2 that the method returns a density that has a substantial non-zero-probability density at a cardiac EF of zero and one. Thus, there is a finite probability of obtaining an ejection fraction close to zero or one. Both these cases clearly are not possible. Although the densities and histograms appear to be consistent, the resultant distributions do not agree with our prior knowledge of the range of cardiac EF values. Using β distribution (Figure 3), the method fares much better. Ranges on the density by using β distribution are consistent with what measured cardiac EFs should be. Thus, the normal distribution, although it

Table 3
Results of the Consistency Check for the Sample Covariance

Sample Covariances	Emory vs QGS	WLCQ vs QGS	WLCQ vs Emory
Estimated directly from data	0.0210	0.0236	0.0210
Estimated using RWT: β	0.0212	0.0237	0.0212
Estimated using RWT: truncated-normal	0.0245	0.0274	0.0245

seems to pass CC2, does not agree with our prior knowledge of the range of EF values.

CC3

The final consistency check (CC3) uses both the estimated linear model parameters and estimated parameters characterizing the gold standard to compare theoretically determined covariances (equation 4) with covariances

measured from the data. Table 3 lists results of CC3. It is clear that the estimates returned from the β distribution very closely match the measured covariances determined by the data. The differences are in the third significant figure. Truncated-normal values for the covariances are close, but not as close as values returned when using β distribution. Again, when we used β distribution, the method clearly passes CC3. Results are less clear for truncated-normal distribution.

DISCUSSION

The RWT method is meant to be used on data for which a gold standard is not available. However, to date, we performed simulation studies and phantom studies for which the gold standard was available to us. Because the gold standard is unavailable in practice, one can never be completely ensured that RWT is working properly. However, in this report, we developed three consistency checks that can be performed on RWT results to garner confidence in the results of RWT. These three consistency checks analyze the linear model parameters returned by means of RWT, as well as parameters characterizing the gold-standard distribution.

Using cardiac EF data derived from gated SPECT studies, we applied RWT and performed our three consistency checks. Using β distribution for the gold standard, we found that results of RWT passed all three checks. Furthermore, we determined that all three methods (Emory, QGS, and WLCQ) have similar performances, although the Emory method appears to slightly outperform the other two methods.

The reader may have noticed a large discrepancy between slopes and intercepts returned by means of RWT using β and truncated-normal as assumed distributions. RWT using β distribution performed better in consistency checks than RWT using truncated-normal distribution. However, it is difficult to determine which model is more appropriate for these data by using results of the consistency checks. This question of model comparison may be addressed by using the maximum value of the likelihood function for each assumed distribution. The model with the higher likelihood is the more appropriate model. This procedure is analogous to using maximum-likelihood theory to choose the model. The log-likelihood value for RWT by using β distribution was -229 when the log-likelihood was -279 for RWT by using truncated-normal distribution. This would imply that β distribution is more

appropriate for these data, which we suspected from the consistency check results. Although this analysis may indicate that the linear model parameters returned by means of RWT are more accurate with β distribution, it should be noted that the rank order of the systems remained the same using either model. We do not suggest that one use this model-comparison technique for a large number of possible models with different numbers of parameters. This type of analysis could return misleading results.

The three consistency checks presented cannot conclusively validate the results of RWT; they can conclusively invalidate these results if the method has not performed well or the model used is inappropriate. RWT is a method that directly addresses the no-gold-standard problem, although without a gold standard, it is difficult to validate results of RWT. The consistency checks we developed enable researchers to garner confidence in the results of RWT.

REFERENCES

1. American Heart Association. Heart Disease and Stroke Statistics—A 2003 Update. Dallas, TX: American heart Association, 2003.
2. Cwajg E, Cwajg J, He Z-X, et al. Gated myocardial perfusion tomography for the assessment of left ventricular function and volumes: comparison with echocardiography. *J Nucl Med* 1999; 40:1857-1865.
3. White HD, Cross DB, Elliot JM, Norris RM, Yee TW. Long-term prognostic importance of patency of the infarct-related coronary artery after thrombolytic therapy for acute myocardial infarction. *Circulation* 1994; 89:61-67.
4. Ritchie JL, Bateman TM, Bonow RO. Guidelines for clinical use of cardiac radionuclide imaging. Report of the American College of Cardiology/American Heart Association Task Force on the Assessment of Diagnostic and Therapeutic Cardiovascular Procedures. *J Am Coll Cardiol* 1995; 25:521-547.
5. Sharir T, Germano G, Kang X, et al. 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. *J Nucl Med* 2001; 42:831-837.
6. Faraone SV, Tsuang MT. Measuring diagnostic accuracy in the absence of a gold standard. *Am J Psychiatry* 1984; 151:650-657.
7. Henkelman RM, Kay K, Bronskill MJ. Receiver operator characteristic (ROC) analysis without truth. *Med Decis Making* 1990; 10:24-29.
8. Sharir T, Germano G, Kavanagh P, et al. Incremental prognostic value of post-stress left ventricular ejection fraction and volume by gated myocardial perfusion single photon emission computed tomography. *Circulation* 1999; 100:1035-1042.
9. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method-comparison studies. *Clin Chem* 1973; 19:49-57.
10. Abe M, Kazatani Y, Fukuda H, Tatsuno H, Habara H, Shinbata H. 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. *J Nucl Cardiol* 2000; 7:569-574.
11. Achtert A, King MA, Dahlberg ST, Pretorius PH, LaCroix KH, Tsui BMW. An investigation of the estimation of ejection fractions and cardiac volumes by a quantitative gated SPECT software package in simulated spect images. *J Nucl Cardiol* 1998; 5:144-152.
12. Bellenger NG, Burgess MI, Ray SG, et al. Comparison of left ventricular ejection fraction and volumes in heart failure by echocardiography, ra-

- dionuclide ventriculography and cardiovascular magnetic resonance. *Eur Heart J* 2000; 21:1387–1396.
13. Faber TL, Vansant J, Pettigrew RI, et al. Evaluation of left ventricular endocardial volumes and ejection fractions computed from gated perfusion SPECT with magnetic resonance imaging: comparison of two methods. *J Nucl Cardiol* 2001; 8:645–651.
 14. He Z, Cwajg E, Presian JS, Mahmarian JJ, Verani MS. 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. *J. Nucl Cardiol* 1999; 6:412–417.
 15. Rumbereger JA, Behrenbeck T, Bell MR, et al. Determination of ventricular ejection fraction: a comparison of available imaging methods. *Mayo Clin Proc* 1997; 72:860–870.
 16. Vaduganathan P, He Z, Vick III, W, Mahmarian JJ, Verani MS. Evaluation of left ventricular wall motion, volumes, and ejection fraction by gated myocardial tomography with technetium 99m-labeled tetrofism: a comparison with cine magnetic imaging. *J Nucl Cardiol* 1999; 6:3–10.
 17. Vanhove C, Franken PR. Left ventricular ejection fraction and volumes from gated blood pool tomography: comparison between two automatic algorithms that work in three-dimensional space. *J Nucl Cardiol* 2001; 8:466–471.
 18. Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983; 32:307–313.
 19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307–310.
 20. Albert PS, McShane LM, Shih JH. Latent class modeling approaches for assessing diagnostic error without a gold standard: with applications to p53 immunohistochemical assays in bladder tumors. *Biometrics* 2001; 57:610–619.
 21. Qu Y, Tan M, Kutner MH. Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics* 1996; 52: 797–810.
 22. Kupinski MA, Hoppin JW, Clarkson E, Barrett HH. Estimation in medical imaging without a gold standard. *Acad Radiol* 2002; 9:290–297.
 23. Hoppin JW, Kupinski MA, Kastis G, Clarkson E, Barrett HH. Objective comparison of quantitative imaging modalities without the use of a gold standard. *IEEE Trans Med Imaging* 2002; 21:441–449.
 24. Hoppin J, Kupinski MA, Wilson DW, et al. Evaluating estimation techniques in medical imaging without a gold standard: experimental validation. *Med Imaging 2003 Image Percept Performance Proc* 2003; 5034:230–237.
 25. Krasnow J, Trask I, Sahlberg S, Margulis G, Leppo JA. Automatic determination of left ventricular function (LVEF) by gated SPECT: comparison of four quantitative software programs. *J Nucl Cardiol* 2001; 8(suppl):S138.
 26. Germano G, Erel J, Kiat H, Kavanagh PB, Berman DS. Quantitative LVEF and qualitative regional function from gated thallium-201 perfusion SPECT. *J Nucl Med* 1997; 38:749–754.
 27. Germano G, Kiat H, Kavanagh PB, et al. Automatic quantification of ejection fraction from gated myocardial SPECT. *J Nucl Med* 1995; 36: 2138–2147.
 28. Faber TL, Cooke CD, Folks RD, et al. Left ventricular function and perfusion from gated SPECT perfusion images: an integrated method. *J Nucl Med* 1999; 40:650–659.
 29. Liu YH, Khaimov D, Sinusas AJ, Wackers FJ. Gated SPECT quantification for myocardial thickness, left ventricular volumes and ejection fraction: methodology and phantom validation. Published by IEEE, New York, NY. Proceedings of the 2003 IEEE Nuclear Science and Medical Imaging, 2004; M11–M299.
 30. Liu YH, Khaimov D. A geometry and count-based method for gated SPECT quantification of myocardial thickness, left ventricular volume and ejection fraction. *IEEE Trans Nucl Sci* 2005; 52:159–165.

APPENDIX A

β Distribution

The probability density function for β distribution has the form:

$$pr(\Theta_p | v, \omega) = \frac{\Theta_p^{v-1} (1 - \Theta_p)^{\omega-1}}{B(v, \omega)} \quad (A1)$$

where

$$B(v, \omega) = \int_0^1 d\Theta_p \Theta_p^{v-1} (1 - \Theta_p)^{\omega-1}. \quad (A2)$$