Comparison of the Binormal and Lehman Receiver Operating Characteristic Curves

Tuesday 8th June, 2021

Abstract

In this paper, we compared the binormal and the Lehmann receiver operating characteristic (ROC) curves using simulation studies and real data sets related to HIV. We ran a large set of simulation studies with a wide range of distributions, varying sample sizes and different degrees of overlap between the diseased and non-diseased population distributions. Simulation results suggest that the binormal ROC model performs better for normal data. However, for non-normal data, the Lehman model outperforms the binormal model. The discrepancies between the performances of the two models were more apparent for larger sample sizes or large degree of separation between the distributions of the diseased and non-diseased subjects. Like the binormal model, the Lehman model provides smooth estimates of ROC curves and a closed form expression for the area under the ROC curve. However, unlike the binormal model, the Lehman model does not make any assumption about the underlying distribution of the biomarker. Further, the Lehmann model can be used to obtain covariate-adjusted ROC curves, and accommodates inference of correlated and longitudinal biomarkers. Thus, the Lehman model should be considered as an alternative when the restrictive distributional assumption of the binormal model are not met.

Key Words: ROC curve, Binormal, Lehmann, Cox PH, HIV/AIDS, Simulation.

1 Introduction

Let $D \in \{0, 1\}$ be a binary random variable denoting disease status of subjects. $D = 1$ for subjects in the diseased (case) population and $D = 0$ for subjects in the non-diseased (reference or control) population. Let Y be a positive random variable representing the value of a continuous biomarker that will be used to distinguish between the diseased and non-diseased population. We assume subjects with higher Y values are more likely to be diseased. For $t \in (-\infty, \infty)$ and $k \in \{0, 1\}$, let $S_k(t) = \Pr(Y > t|D = k)$ denote the survival function. The sensitivity (true positive rate) of the biomarker at cut-off t is given by $\Pr(Y > t|D = 1) = S_1(t)$. Similarly, the specificity (true negative rate) of the biomarker at cut-off t is given by $\Pr(Y \leq t|D = 0) = 1 - \Pr(Y > t|D = 0) = 1 - S_0(t)$. Note that $\lim_{t\to\infty} S_1(t) = \lim_{t\to\infty} S_0(t) = 0$ and $\lim_{t\to-\infty} S_1(t) = \lim_{t\to-\infty} S_0(t) = 1$. These two measures of diagnostic performance vary depending on the cut-off selected. However, the receiver operating characteristic (ROC) curve, which incorporates both of these measures, gives a full picture of the performance of the classifier biomarker at different thresholds. That is, unlike the single measurements of sensitivity and specificity, the ROC curve does not require selection of a particular threshold. The curve is the entire set of possible values of $S_0(t)$ (1-specificity)and $S_1(t)$ (sensitivity) obtained by dichotomizing Y with different values of $t \in (-\infty, +\infty)$. Graphically, the ROC curve is a plot of $S_0(t)$ versus $S_1(t)$ for every possible threshold t. The coordinates of a given point on the curve are ROC$(\cdot)$=$\{$(1-specificity(t), sensitivity(t))$=(S_0(t), S_1(t)), t \in (-\infty, \infty)\}$. Using a change of variable, if we denote $S_0(t)$ by $x \in (0, 1)$, $S_1(t) = S_1(S_0^{-1}(x))$ and then ROC$(\cdot)$ can also be expressed by $(x, S_1(S_0^{-1}(x)))$. Thus, the ROC is given by

$$\text{ROC}(x) = S_1(S_0^{-1}(x)) \quad \text{for } x \in (0, 1), \tag{1.1}$$

where $S_0^{-1}(\cdot)$ is the inverse of the function $S_0(\cdot)$.

The ROC curve has many advantages compared to the other measures of diagnostic perfromance. The ROC curve gives a complete picture of the performance of the classifier biomarker at different thresholds. Moreover, the ROC curve is invariant under monotone increasing transformations; that is, does not depend on the scale of the biomarker. In other

words, for any monotonic increasing function f, Y and $f(Y)$ has the same ROC curve. Further, the ROC curve facilitates meaningful comparisons of two or more biomarkers measured in different units or on different scales. Despite these advantages, it is useful to reduce the ROC curve to a single quantitative summary index representing overall performance of the classifier. Such quantitative summary index frequently used for an overall assessment of classifier performance is the area under the ROC curve (AUC), defined as AUC=$\int_0^1 \text{ROC}(x)dx = \int_0^1 S_1(S_0^{-1}(x))dx$ for $x \in (0,1)$. AUC equals the probability that a randomly selected subject from the diseased population would have a higher score than a randomly selected subject from the non-diseased population; that is Pr[Y1> Y0] [1, 2]. Because the ROC curve is bounded within the unit square, the AURC can range from 0 to 1. An area of 1 represents a perfect biomarker (sensitivity=specificity=1) which is rarely in clinical practice. An area of 0 represents an imperfect biomarker which is extremely unlikely happen in clinical practice. For this reason an area of 0.5 (diagonal line) is often considered as the minimum AUC. An area of 0.5 represents an non-informative or non-discriminatory biomarker which is no better than flipping of a coin.

The visual nature of the curve and the simplicity of the numeric summary index (AUC) have made ROC analysis popular in many disciplines. Due to its widespread usefulness, various nonparametric and parametric methods have been so far developed for estimating the ROC curve [2, 3, 4]. In this paper, we investigated the comparative performance of a semi-parametric method. In particular, we assessed the diagnostic performance of the Lehmann ROC [6, 5] model compared to the commonly used binormal and ROC curve. We run a large set of simulation studies to evaluate the comparative performance of Lehmann model for a wide range of distributions and varying sample sizes. In each simulation, we considered different degrees of overlap between the diseased and non-diseased population distributions. The remainder of the paper is organized as follows: In section 2, we introduce the binormal and Lehman ROC curves. Simulation studies and results of the simulation studies are summarized in section 3. In section 4, we introduce the population, design, and variables of the studies used in our data analyses and summarize the results of our data analyses. Finally, the conclusion section features interpretations of our findings and concluding remarks.

2 Methods

Let Y_0 and Y_1 be two independent random variables, respectively denoting the value of a biomarker for a non-diseased and for a diseased population. In other words, Y_0 and Y_1 denote values of the continuous biomarker Y when $D = 0$ and $D = 1$, respectively. Let S_0 (h_0) and S_1 (h_1) be the survival (hazard) functions of the random variables Y_0 and Y_1, respectively. Let m and n are the number of subjects in the non-diseased and diseased population, then $y_1^0, y_2^0, y_3^0, \ldots, y_m^0$ and $y_1^1, y_2^1, y_3^1, \ldots, y_n^1$ represent a random sample of biomarker values from the non-diseased and diseased population, respectively. There are various methods for calculating the ROC curve depending on the distributional assumptions of $S_0(t)$ and $S_1(t)$ [2, 3, 4]. We provide here a brief description of the parametric, and semiparametric methods we used for the estimation of the ROC curve.

2.1 Parametric Method

Parametric modeling of an ROC curve assumes that the classifying bimarker follows the same known distribution in both non-diseased and diseased subjects. Parametric methods have an advantage over over non-parametric methods. They provide estimates of smooth ROC curves and a closed form expression for the area under the ROC curve. While there are many parametric models of the ROC curve, in this paper we focus on the binormal model. The binormal model, the most frequently used parametric model, assumes that the biomarker or a transformation of the biomarker follows normal distribution in both populations. In particular, under the binormal assumption, Y_0 and Y_1 are independent normal random variables with means $\mu_0 \leq \mu_1$ and variances σ_0^2 and σ_1^2, respectively. The assumption $\mu_0 \leq \mu_1$ follows from the convention that larger biomarker values are more indicative of disease. For a given threshold $t \in (-\infty, \infty)$, the sensitivity and 1-specificity can be expressed as $S_1(t) = 1 - \Phi\left(\frac{t-\mu_1}{\sigma_1}\right) = \Phi\left(\frac{\mu_1 - t}{\sigma_1}\right)$ and $S_0(t) = 1 - \Phi\left(\frac{t-\mu_0}{\sigma_0}\right) = \Phi\left(\frac{\mu_0 - t}{\sigma_0}\right)$ where Φ is the standard normal cumulative distribution function. Note that $1 - \Phi(t) = \Phi(-t)$. If we denote $S_0(t)$ by $x \in (0, 1)$ as in section 1, then $t = S_0^{-1}(x) = \mu_0 - \sigma_0 \Phi^{-1}(x)$ and $S_1(t) = \Phi\left(\frac{\mu_1 - (\mu_0 - \sigma_0 \Phi^{-1}(x))}{\sigma_1}\right)$. Using equation (1.1), the ROC of the binormal model can be expressed as:

$$\begin{aligned} \text{ROC}(x) &= S_1(S_0^{-1}(x)) \\ &= \Phi\left(\frac{\mu_1 - \mu_0 + \sigma_0\Phi^{-1}(x))}{\sigma_1}\right) \\ &= \Phi\left(\alpha + \beta\Phi^{-1}(x)\right), \end{aligned} \tag{2.1}$$

where $\alpha = \frac{\mu_1-\mu_0}{\sigma_1}$ and $\beta = \frac{\sigma_0}{\sigma_1}$ are the binormal parametrs. Equation (2.1) yields a closed form expression for the area under the binormal ROC curve which is given by

$$\begin{aligned} \text{AUC} &= \int_0^1 \Phi\left(\alpha + \beta\Phi^{-1}(x)\right) dx \\ &= \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}}\right) = \Phi\left(\frac{(\mu_1 - \mu_0)/\sigma_1}{\sqrt{1 + \sigma_0^2/\sigma_1^2}}\right) \\ &= \Phi\left(\frac{\alpha}{\sqrt{1+\beta^2}}\right) \end{aligned} \tag{2.2}$$

This shows that the area under the binormal ROC curve is an increasing function of α and a decreasing function of β. For non-informative biomarker $\alpha = 0$ ($\mu_0 = \mu_1$), $\Phi\left(\frac{\alpha}{\sqrt{1+\beta^2}}\right) = 0.5$ and the plot of ROC curve in (2.1) is a diagonal line.

Estimation of the binormal ROC curve involves estimating the binormal parameters α and β. Many methods and algorithms above have been proposed for estimating these parameters [2, 7]. In this paper, we used the maximum likelihood (ML) approach to estimate α and β. Let $y_1^0, y_2^0, \ldots, y_m^0$ and $y_1^1, y_2^1, \ldots, y_n^1$ be random samples from the non-diseased and diseased populations, respectively. Then the ML estimators of the mean (μ) and variance (σ) vectors are: $\hat{\mu_0} = \frac{\sum_{i=1}^m y_i^0}{m} = \bar{y^0}$; $\hat{\mu_1} = \frac{\sum_{i=1}^n y_i^1}{n} = \bar{y^1}$; $\hat{\sigma_0^2} = \frac{\sum_{i=1}^m (y_i^0 - \bar{y^0})^2}{m}$; and $\hat{\sigma_1^2} = \frac{\sum_{i=1}^n (y_i^1 - \bar{y^1})^2}{n}$. From these, we obtain the ML estimators of the binormal parameters: $\hat{\alpha} = \frac{\hat{\mu_1} - \hat{\mu_0}}{\hat{\sigma_1}}$ and $\hat{\beta} = \frac{\hat{\sigma_0}}{\hat{\sigma_1}}$. Plugging these in (2.1) and (2.3), we estimate the ROC curve and the corresponding AUC under the binormal model: $\widehat{\text{ROC}}(x) = \Phi\left(\hat{\alpha} + \hat{\beta}\Phi^{-1}(x)\right)$ and $\widehat{\text{AUC}} = \Phi\left(\frac{\hat{\alpha}}{\sqrt{1+\hat{\beta}^2}}\right)$. Note that these estimates can be biased when the normality assumption is violated.

2.2 Semi-parametric Method

In this section, we present an alternative to the bi-normal ROC model based on the most popular Cox-PH model. In other words, we consider the Lehman model for the estimation of the ROC curve [5, 6] . Unlike the bi-normal model, the Lehmann model does not need to make any assumptions about the underlying distribution of the continuous biomarker; and like the binormal model, it involves estimation of model parameters. Thus, the Lehmann approach can be described as a semi-parametric method. The Lehmann model states that log(sensitivity) is proportional to log(1-specificity). That is, the association between sensitivity and 1-specificity can be expressed by:

$$S_1(t) = S_0(t)^\theta, \tag{2.3}$$

where, $0 < \theta \leq 1$ is measure of biomarker diagnostic performance with values close to zero indicating higher performance. Equation (2.3) is similar to the proportional hazards (PH) model [8, 9] assumption widely used in survival analysis. If we denote $S_0(t)$ by $x \in (0, 1)$, $S_1(S_0^{-1}(x)) = S_1(t)$. Then using the proportionality assumption, the Lehmann ROC curve can be written as

$$\begin{aligned} \text{ROC}(x) &= S_1(S_0^{-1}(x)) \\ &= x^\theta. \end{aligned} \tag{2.4}$$

It is easy to see from (2.4) that as θ varied along its range, a family of Lehmann ROC curves is generated. As in the case of binormal model, we obtain a closed for expression for the area under the Lehmann ROC curve:

$$\begin{aligned} \text{AUC}(x) &= \int_0^1 x^\theta dx \\ &= \frac{1}{1+\theta}. \end{aligned} \tag{2.5}$$

Estimation of the Lehmann ROC curve (2.4) requires estimating the parameter θ. The

connection between the Lehmann and Cox PH model enable us to estimate θ. Based on the Cox PH model assumption, the covariate D have proportional effect on the hazard function of Y. That is, $h_1(y|D) = h_0(y)e^{\beta_1 D}$ and $\frac{h_1(y,D=1)}{h_0(y,D=0)} = e^{\beta_1}$. Then using the Cox and Lehmann model assumptions, the biomarker diagnostic performance measure $\theta = e^{\beta_1}$. The parameter β_1 is estimated by $\hat{\beta}$ which can be obtained using Cox's [8, 9] partial maximum likelihood estimation procedure available in many statistical software packages. The parameter θ can be estimated by applying the invariance property of maximum likelihood estimators: $\hat{\theta} = e^{\hat{\beta}_1}$. Using the Delta method, an approximate variance for $\hat{\theta}$ can be expressed as $\hat{\theta}^2\sigma^2_{\hat{\beta}_1}$, where $\sigma^2_{\hat{\beta}_1}$ is variance of $\hat{\beta}_1$. Similarly, the variance for the area under Lehmann ROC curve (2.5) can be estimated by $\hat{\theta}^2\sigma^2_{\hat{\beta}_1}(1+\hat{\theta})^{-4}$.

3 Simulation Study

3.1 Data Generation

This section presents some simulation results conducted to evaluate the performance of the Lehman ROC curve in comparison to the bi-normal and empirical ROC curves. Two sets of simulation studies were considered: data were generated from normal and Weibull distributions. The normal distribution violates the proportional hazard assumption. However, proportional hazard assumption holds for the Weibul distribution. In each case, we generated $5,000$ independent data sets, each of which included m and n subjects generated from the non-diseased and diseased populations, respectively. Various values of m and n and different degrees of overlap between the diseased and non-diseased population distributions were considered.

3.1.1 Normal Distribution

In the first part of our simulation, we generated data from normal distributions. More specifically, we assumed that the biomarker follows a normal distribution in both non-diseased and diseased subject: $Y_0 \sim N(\mu_0, \sigma_0)$ and $Y_1 \sim N(\mu_1, \sigma_1)$. Various values of the location (μ_0 & μ_1) and spread (σ_0 & σ_1) parameters were considered. As described in section 2,

these parameters are related to area under the ROC by th AUC=$\Phi\left(\frac{(\mu_1-\mu_0)/\sigma_1}{\sqrt{1+\sigma_0^2/\sigma_1^2}}\right)$. Thus, for a given true AUC, μ_0, σ_0 and σ_1 parameter values, μ_1 is estimated by $\mu_0+\Phi^{-1}(\text{AUC})\sqrt{\sigma_0^2+\sigma_1^2}$. Three scenarios for the area under ROC curve were considered: AUC=0.55 (weak classifier); AUC=0.70 (moderate classifier) and AUC=0.90 = (strong classifier).

3.1.2 Weibull Distribution

In the second part of our simulation study, we assumed the biomarker for the non-diseased and diseased subjects follow two parameter Weibull distribution [10, 11]. with same shape parameter and different scale parameters: $Y_0 \sim W(\delta, \theta_0)$ and $Y_1 \sim W(\delta, \theta_1)$. Then the density functions for the non-diseased and diseased population take the form $f_0(y_0, \delta, \theta_0) = \frac{\delta}{\theta_0} y_0^{\delta-1} e^{-\frac{y_0^\delta}{\theta_0}}$ and $f_1(y_1, \delta, \theta_1) = \frac{\delta}{\theta_1} y_1^{\delta-1} e^{-\frac{y_1^\delta}{\theta_1}}$ where y_i, θ_i, , $\delta > 0$ for $i = 0, 1$. Thus, the sensitivity and 1-specificity are, respectively, given by $S_1(t) = e^{-\frac{t^\delta}{\theta_1}}$ and $S_0(t) = e^{-\frac{t^\delta}{\theta_0}}$. If we denote $S_0(t) = x \in (0, 1)$ then $t = -[\theta_0 \ln x]^{1/\delta}$, ROC(t)=$S_1(t) = S_1(-[\theta_0 \ln x]^{1/\delta}) = x^{\frac{\theta_0}{\theta_1}}$ and AUC $= \int_0^1 \text{ROC}(x)dx = \frac{\theta_1}{\theta_0+\theta_1}$. For given AUC and θ_0 values, θ_1 is estimated by $\theta_0\left[\frac{\text{AUC}}{1-\text{AUC}}\right]$. As in the case of the normal distribution, we considered AUC=0.55, AUC=0.70 and AUC=0.90.

3.2 Simulation Results

We generated generated 5,000 independent data sets from each distribution. The two ROC models were applied to the generated 5,000 independent data sets, and the means of the 5,000 different AUC parameter estimates together with their corresponding biases and mean squared errors (MSEs) were calculated. Overall, the results of our simulation studies show that the two ROC models yield comparable estimates of the AUC. The discrepancy between the biases of the two models increases as the sample size or the degree of separation between the distributions of diseased and non-diseased population increase. Moreover, as the sample size increases, the stand errors of the AUC estimates decrease. The results presented here are representative of the many simulations performed.

Table 1 shows the simulation results when the biomarker follows a normal distribution in both diseased and non-diseased subjects. Our simulation work of the normal data indicates

that the binormal ROC model performs better compared to the Lehman-Cox ROC model. The superiority of the binormal ROC model is more noticeable as sample size increases or when the degree of separation between the two distributions is high, i.e., for higher AUC. The MSEs of both models decrease as the sample size decrease. The binormal ROC model yielded smaller bias which becomes negligible as the sample size increases. This is true regardless the closeness or separation of the two distributions. However, the Lehman-Cox model yielded biased AUC estimates. The bias of the Lehman-Cox model was larger for larger sample sizes or higher degree of separation between the distributions of the diseased and non-diseased subjects.

Tables 2 to 4 show the simulation results when the biomarker follows a Weibul distribution in both diseased and non-diseased subjects. Weibul distribution is very flexible. Depending on the values of the shape parameter δ, it can be used to model a variety of distributions including exponential, Rayleigh, normal and extreme-value. In particular, for $\delta = 1$ and $\delta = 2$ it is equivalent to exponential and Rayleigh distributions, respectively. For δ between 3 and 4, the distribution becomes like the normal distribution. For large δ values, the Weibul distribution becomes extreme value distribution. Our simulation work of the Weibul data indicates that the Lehman-Cox ROC model performs better compared to the binormal ROC model. When the shape parameter is small (Table 2) or large (Table 4), the Lehman-Cox ROC model out performs the binormal ROC model. However, when the shape parameter is moderate (Table 3), the two models yielded comparable results. This is to be expected for moderate values, such as $\delta = 3.5$, the Weibul distribution is bell shabed and symmetric like the normal distribution. As it was the case with the normal data, the discrepancy between the two models were more noticeable for larger sample sizes or larger degrees of separation between the diseased and non-diseased distributions.

4 Data Analyses

In this section, we present analyses of three real data sets related to HIV. The datasets include weak, moderate and strong classifying biomarkers. Moreover, the data sets include biomarkers who fulfill and violate the normality and Cox-PH assumptions. For each dataset,

Table 1: Mean, bias and root mean squared error (RMSE) of the AUC estimates

	AUC=0.55		AUC=0.70		AUC=0.90	
	B	L	B	L	B	L
$m = n = 5$						
Mean	0.5481	0.5443	0.6794	0.6982	0.8852	0.8751
Bias	-0.0019	-0.0057	-0.0073	-0.0206	-0.0148	-0.0249
MSE	0.0325	0.0340	0.0250	0.0294	0.0092	0.0142
$m = n = 15$						
Mean	0.5492	0.5426	0.6971	0.6711	0.8953	0.8611
Bias	-0.0008	-0.0074	-0.0029	-0.0289	-0.0047	-0.0389
MSE	0.0106	0.0096	0.0088	0.0096	0.0029	0.0060
$m = n = 30$						
Mean	0.5499	0.5416	0.6997	0.6693	0.8982	0.8570
Bias	-0.0001	-0.0084	-0.0003	-0.0307	-0.0018	-0.0430
MSE	0.0052	0.0046	0.0043	0.0051	0.0015	0.0042
$m = n = 60$						
Mean	0.5504	0.5414	0.6994	0.6659	0.8990	0.8526
Bias	0.0004	-0.0086	-0.0006	-0.0341	-0.0010	-0.0474
MSE	0.0027	0.0023	0.0021	0.0032	0.0007	0.0034
$m = n = 100$						
Mean	0.5503	0.5408	0.6993	0.6644	0.8986	0.8497
Bias	0.0003	-0.0092	-0.0007	-0.0356	-0.0014	-0.0503
MSE	0.0016	0.0014	0.0013	0.0025	0.0004	0.0033
$m = n = 200$						
Mean	0.5503	0.5407	0.7000	0.6639	0.8996	0.8487
Bias	0.0003	-0.0093	0.0000	-0.0361	-0.0004	-0.0513
MSE	0.0008	0.0007	0.0006	0.0019	0.0002	0.0030

Normal distributions; $\sigma_0 = \sigma_1 = 3$, $\mu_0 = 5$

Table 2: Mean, bias and root mean squared error (RMSE) of the AUC estimates

	AUC=0.55		AUC=0.70		AUC=0.90	
	B	L	B	L	B	L
$m = n = 5$						
Mean	0.5528	0.5520	0.6949	0.6990	0.8443	0.8995
Bias	0.0028	0.0020	-0.0051	-0.0011	-0.0556	-0.0005
MSE	0.0314	0.0333	0.0215	0.0267	0.0095	0.0092
$m = n = 10$						
Mean	0.5541	0.5510	0.70195	0.70140	0.8325	0.9002
Bias	0.0041	0.0010	0.0020	0.0014	-0.0675	0.0002
MSE	0.0154	0.0148	0.0094	0.0116	0.0081	0.0041
$m = n = 15$						
Mean	0.5538	0.55016	0.7009	0.6997	0.828	0.9008
Bias	0.0038	0.0002	0.0009	-0.0003	-0.0717	0.0008
MSE	0.0102	0.0094	0.0060	0.0077	0.0077	0.0026
$m = n = 30$						
Mean	0.5534	0.5490	0.7018	0.7006	0.8216	0.9002
Bias	0.0034	-0.0010	0.0018	0.0006	-0.0784	0.0002
MSE	0.0051	0.0045	0.0028	0.0037	0.0076	0.0012
$m = n = 60$						
Mean	0.5562	0.5508	0.7025	0.7018	0.8164	0.9005
Bias	0.0062	0.0008	0.0025	0.0018	-0.0836	0.0005
MSE	0.0025	0.0021	0.0013	0.0017	0.0078	0.0006
$m = n = 100$						
Mean	0.5562	0.5506	0.7015	0.7007	0.8146	0.8998
Bias	0.0062	0.0006	0.0015	0.0007	-0.0854	-0.0002
MSE	0.0015	0.0013	0.0008	0.0010	0.0078	0.0004
$m = n = 200$						
Mean	0.5559	0.5502	0.7010	0.7007	0.8128	0.9000
Bias	0.0059	0.0002	0.0010	0.0007	-0.0872	0.0000
MSE	0.0008	0.0006	0.0004	0.0005	0.0079	0.0002

Weibul distributions; $\delta = 1$, $\theta_0 = 1$

Table 3: Mean, bias and mean squared error (MSE) of the AUC estimates

	AUC=0.55		AUC=0.70		AUC=0.90	
	B	L	B	L	B	L
$m = n = 5$						
Mean	0.5481	0.5516	0.6922	0.7001	0.8917	0.9022
Bias	-0.0019	0.0016	-0.0078	0.0001	-0.0083	0.0022
MSE	0.0323	0.0338	0.0259	0.0264	0.0083	0.0088
$m = n = 15$						
Mean	0.5483	0.5489	0.6994	0.7003	0.8987	0.9004
Bias	-0.0017	-0.0011	-0.0006	0.0003	-0.0013	0.0004
MSE	0.0108	0.0096	0.0085	0.0074	0.0029	0.0027
$m = n = 30$						
Mean	0.5527	0.5525	0.7035	0.7019	0.9006	0.8999
Bias	0.0027	0.0025	0.0035	0.0019	0.0006	-0.0001
MSE	0.0051	0.0044	0.0042	0.0036	0.0014	0.0012
$m = n = 60$						
Mean	0.5507	0.5507	0.7030	0.7006	0.9024	0.9007
Bias	0.0007	0.0007	0.0030	0.0006	0.0024	0.0007
MSE	0.0026	0.0021	0.0021	0.0017	0.0007	0.0006
$m = n = 100$						
Mean	0.5500	0.5493	0.7021	0.6997	0.9021	0.8996
Bias	0.0000	-0.0007	0.0021	-0.0003	0.0021	-0.0004
MSE	0.0015	0.0013	0.0013	0.0010	0.0004	0.0003
$m = n = 200$						
Mean	0.5510	0.5503	0.7031	0.7001	0.9029	0.8999
Bias	0.0010	0.0003	0.0031	0.0001	0.0029	-0.0001
MSE	0.0007	0.0006	0.0007	0.0005	0.0002	0.0002

Weibul distributions; $\delta = 3.5$, $\theta_0 = 1$

Table 4: Mean, bias and mean squared error (MSE) of the AUC estimates

	AUC=0.55		AUC=0.70		AUC=0.90	
	B	L	B	L	B	L
$m = n = 5$						
Mean	0.5509	0.5551	0.6845	0.7008	0.8872	0.9031
Bias	0.0009	0.0051	-0.0155	0.0008	-0.0128	0.0031
MSE	0.0302	0.0318	0.0262	0.0253	0.0102	0.0089
$m = n = 15$						
Mean	0.5438	0.5485	0.6848	0.7008	0.8912	0.9000
Bias	-0.0062	-0.0015	-0.0152	0.0008	-0.0088	0.0000
MSE	0.0105	0.0095	0.0091	0.0075	0.0037	0.0025
$m = n = 30$						
Mean	0.5447	0.5504	0.6865	0.7009	0.8926	0.9004
Bias	-0.0053	0.0004	-0.0135	0.0009	-0.0074	0.0004
MSE	0.0054	0.0045	0.0048	0.0037	0.0020	0.0012
$m = n = 60$						
Mean	0.5463	0.5510	0.6853	0.7001	0.8941	0.9002
Bias	-0.0037	0.0010	-0.0147	0.0001	-0.0059	0.0002
MSE	0.0026	0.0021	0.0025	0.0018	0.0010	0.0006
$m = n = 100$						
Mean	0.5458	0.5503	0.6866	0.7011	0.8948	0.9005
Bias	-0.0042	0.0003	-0.0134	0.0011	-0.0052	0.0005
MSE	0.0016	0.0013	0.0016	0.0011	0.0006	0.0004
$m = n = 200$						
Mean	0.5457	0.5501	0.6854	0.7002	0.8941	0.9000
Bias	-0.0043	0.0001	-0.0146	0.0002	-0.0059	0.0000
MSE	0.0008	0.0006	0.0009	0.0005	0.0003	0.0002

Weibul distributions; $\delta = 20$, $\theta_0 = 1$

we computed both the binormal and Lehman ROC curves and their corresponding AUCs. Comparison of the AUCs from the two ROC models based on the asymptotic or approximate confidence intervals (CIs) can be misleading. Thus, we used CIs based on percentile bootstrap method [12] with 5k iterations.

4.1 Age at first sex and HIV

Early age of sexual onset is known to be associated with higher levels of HIV, HSV-2 other sexually transmitted infection (STIs)[13]. The association between early age at first sex and increased risk of STIs can be due several factors including behavioral and biological. Women who start their sexual debut at a young age tended to be more sexually active and are more likely to engage in subsequent risky behaviors, such as unprotected sex, and multiple partners. Moreover, the physiologic and immunological immaturity of the female genital tract may increase susceptibility to trauma during sexual intercourse thereby facilitating entry of infectious organisms. We aim here to assess the diagnostic performance of age at first sex as a biomarker of HIV acquisition. We used age of sexual onset data from 151 HIV positive and $1,234$ HIV negative women who participated in the Moshi Infertility survey. The survey was conducted from November 2002 to March 2003 in the Moshi urban district of Tanzania. The rationale, methods, and recruitment of the participants have been described in detail previously [13, 14].

Figure 1 (a) displays distribution of age at fist sex by HIV status. Visual inspection of figure 1 (a),diagnostic plots (such as Q-Q) and statistical tests (such as Kolmogorov-Smirnov and Shapiro-Wilks) indicate that the distribution of age at firs sex is right-skewed in both groups of women, i.e., minimal to modest deviations from normality. Moreover, graphical diagnostics and statistical tests indicate violations of the proportional hazards model assumption. Due the robustness of binormal ROC curve, these deviations from normality assumptions have little or negligible consequences on the binormal AUC estimates. The dotted lines in the figure indicate median ages at first sex for HIV positive and negative participants. The median ages at first sex was significantly higher for HIV negative (median=19 years; IQR=19-21) compared to HIV positive (median=17 years; IQR=16-19.5) women (p-value$<$0.001). Despite the significant differences in median ages, the two distri-

butions overlap indicating that the degree of separation is low. That is, age at first sex has low performance in discriminating HIV positive from HIV negative women. This can be clearly seen from the binormal and Lehman ROC curves displayed in figure 1 (b). The two ROC curves are overlaying thereby yielding similar areas under the ROC curves. The area under binormal and Lehman ROC curves together with their corresponding 95% CI based on percentile bootstrap with 5k iterations are 0.60 (95% CI=0.56-0.65) and 0.59 (95% CI=0.55-0.64), respectively. The distributions of age at first sex have the same shape in both groups of women, i.e., comparable standard deviations. Thus, as shown in the results of the simulation studies of the normal data 3.2, the two ROC models yielded similar results. Both models show poor performance of age at first sex as a predictive biomarker of HIV status.

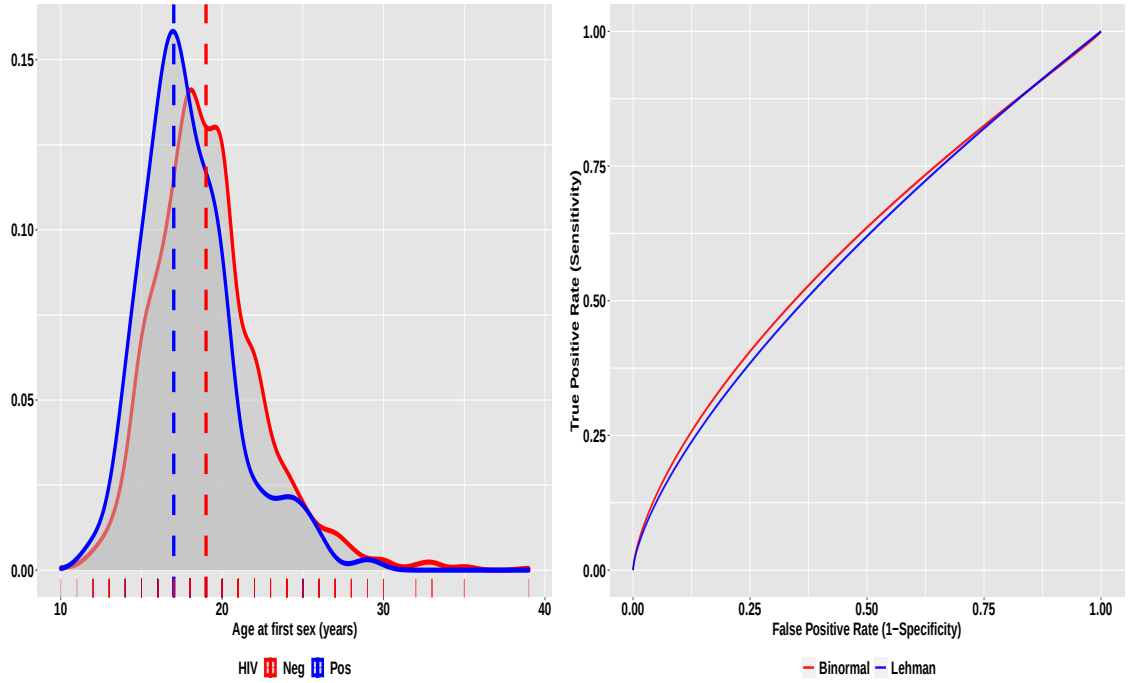

Figure 1: (a). Distribution of age at first sex by HIV status, (b). ROC curves for the binormal and Lehman models

4.2 Baseline CD4+ T cell count and immune restoration

The Human Immunodeficiency Virus (HIV) damages essential immune cells called CD4+ T cells. These cells help the body to fight infection and disease. The decline in CD4 T cells lead to increasing immunodeficiency thereby resulting in advanced stage of HIV infection

(i.e., AIDS), and finally death. Treatment with highly active antiretroviral therapy (HAART) is used to restore CD4+ T cells in HIV-infected patients. Patients with advanced HIV/AIDS prior to HART initiation are less likely to achieve immune recovery faster compared to HIV patients with less advanced HIV/AIDS. We aim here to assess the performance of pre-HAART CD4+ T-lymphocyte counts as predictive biomarkers of post-HAART immune recovery in HIV-infected children without TB co-infection. We included 113 TB-negative HIV-infected children who were enrolled through a retrospective study in Accra, Ghana. The children were between 0 and 13 years of age and received HAART treatment between June 2004 and December 2009. The children were followed during the study period until they achieved immune recovery or were censored on the last day of contact. Immune recovery was defined as achieving and maintaining a target CD4+ T cell percentage of 25% following the initiation of HAART. The study methods, and participants have been described in detail previously[15].

Out of the 113 children included in our analyses, 77% (n=87) achieved immune recovery and 23% (n=26) failed to achieve immune recovery. The distribution of pre-HAART CD4+ T cells counts are summarized in figure 2 (a). Children who achieved immune recovery had significantly higher pre-HAART CD4+ T cells counts (<0.001). Median pre-HAART CD4+ T cells counts was 445 cells/mm^3 (IQR=298-758) and 139 cells/mm^3 (IQR=51-495) in patients who achieved and failed to achieve recovery, respectively. Visual inspection of figure 2 (a), diagnostic plots and statistical tests indicate the normal distribution does not fit reasonable well to the data. However, statistical tests and graphical diagnostics based on the Schoenfeld residuals [16] indicate no evidence of violation of the proportional hazards model assumption. That is, Weibul distribution fits well to the CD4+ T cells counts data from both groups. The binormal and Lehman ROC curves are displayed in figure 2 (b). The area under the binormal and Lehman ROC curves were 0.72 (95% CI=0.64-0.81)and 0.71 (95% CI=62-81), respectively. Despite the departure from normality, the two models yielded comparable results thereby showing the robustness of the binormal ROC curve. Both models show that pre-HAART CD4+ T-lymphocyte counts are adequate indicators of pediatric immune recovery. In both models, pre-HAART CD4+ T-lymphocyte counts had more than a 70% probability of correctly distinguishing a recovered from non-recovered

patient.

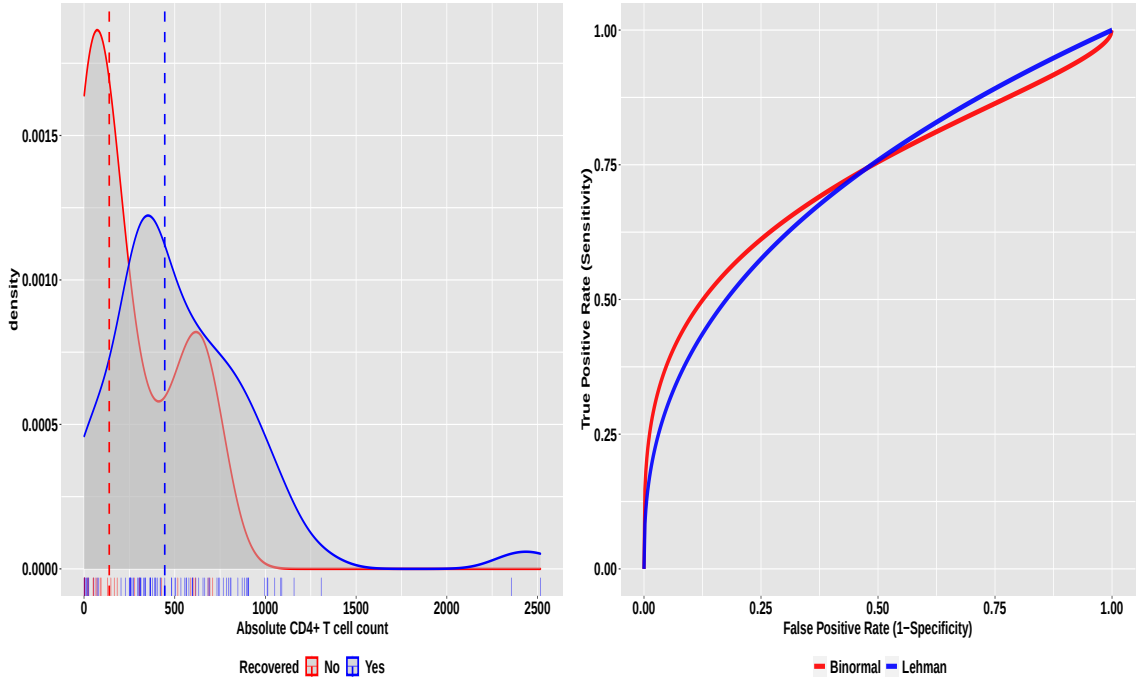

Figure 2: (a). Distribution of baseline CD4 T cell count by immune recovery status; (b). ROC curves for the binormal and Lehman models

4.3 Cerebrospinal fluid glucose and TB

HIV and Tuberculosis (TB) are highly prevalent in most regions experiencing severe HIV epidemics. In sub-Saharan Africa, where HIV prevalence is highest, substantially higher TB rates are seen. We aim here to assess the diagnostic performance of cerebrospinal fluid (CSF) glucose for discriminating subjects with tuberculous meningitis (TBM) from those without TBM. TBM is the most common form of central nervous system tuberculosis. We used CSF glucose data from 472 HIV positive patients who participated in a prospective cohort study of TBM. The study was conducted between April 2014 and August 2017 at the University Teaching Hospital (UTH) in Lusaka, Zambia. The study design, study population, laboratory and diagnostic methods used in the study are described previously in [17].

Among the 472 HIV positive patients included in our analyses, 92 (19%) and 380 (81%) were, respectively, classified as MTB positive and negative patients using the gold standard CSF culture. Figure 3 (a) presents the distribution of CSF glucose by MTB status. HIV

patients with MTB has significantly lower CSF glucose levels compared to HIV patients without MTB (P-value<0.001). The median CSF scores in TMB negative and positive subjects were 3.35 mg/ml (IQR=2.61-3.98) and 1.47 mg/dl (IQR=0.96-2.20), respectively. Statistical tests of normality, visual inspection of figure 3 and other diagnostic plots show departure from the normality assumption. Moreover, diagnostic plots (Schoenfeld residuals) indicate violations of the proportional hazards model assumptions. The overlap in the distribution of CSF glucose between MTB positive and negative participants is small, therefore we expect CSF glucose to be a a good biomarker of MTB. The binormal and Lehman ROC curves are displayed in figure 3 (b). The area under the binormal and Lehman ROC curves were 0.80 (95%=0.75-0.86) and 0.76 (95%=0.69-0.84), respectively. In both models, CSF glucose had more than a 76% probability of correctly discriminating subjects with tuberculous meningitis (TBM) from those without TBM.

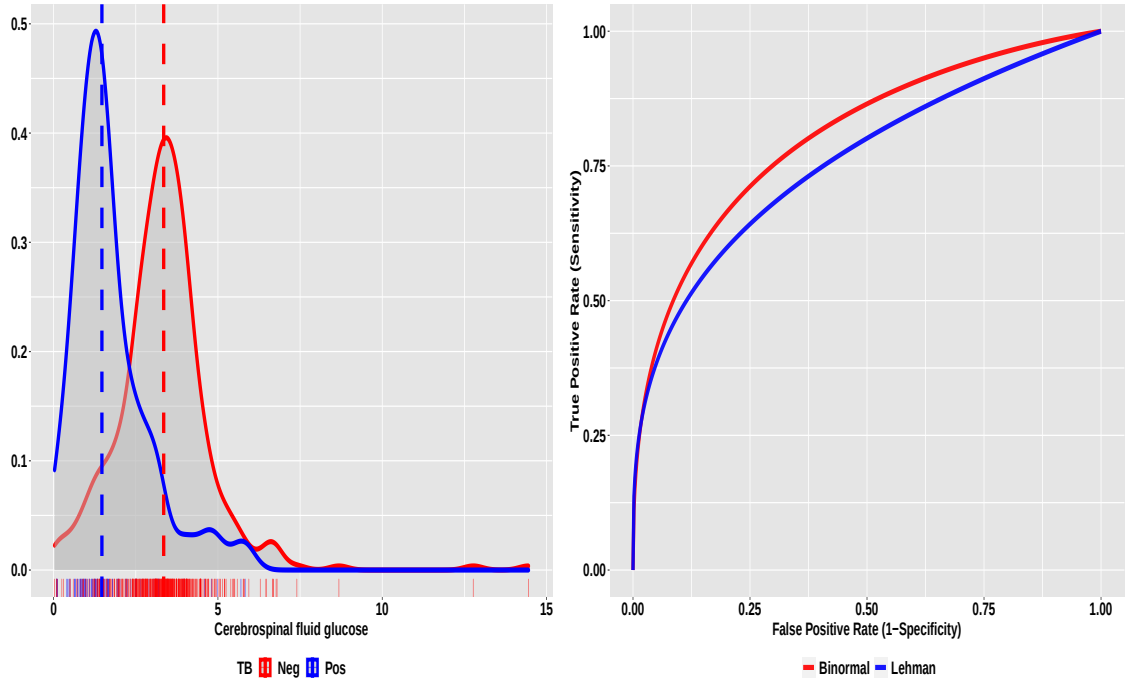

Figure 3: (a). Distribution of cerebrospinal fluid glucose by TB status; (b). ROC curves for the binormal and Lehman models

5 Conclusion

In this work, we investigate the performances of the binormal and Lehman ROC curves using simulation studies. Further, real data sets from HIV/AIDS research were analyzed for illustrative purposes. Two sets of simulation studies were considered. In the first part of our simulation study, we generated data from normal distributions. That is, we assumed that the biomarker follows a normal distribution in both non-diseased and diseased subject. In the second part of our simulation study, we considered non-normal data. More specifically, we generated data assuming the biomarker for the non-diseased and diseased subjects follow two parameter Weibull distribution with same shape parameter and different scale parameters. Weibul distribution is very useful because manipulation of the shape parameter allows it to be equivalent to a variety of distributions including exponential, Rayleigh, normal and extreme-value. Simulation results suggest that the binormal ROC model performs better for normal data. The superiority of binormal model is more apparent for larger sample sizes or large degree of separation between the distributions of the diseased and non-diseased subjects. However, for non-normal data, the Lehman-Cox model performance was better or comparable to the binormal model. This is to be expected for the proportional hazard assumption holds with the data generated from Weibull distributions.

The binormal model is the most frequently used method. The model yields smooth estimates of ROC curves and a closed form expression for the area under the ROC curve. However, the model requires that the biomarker or a transformation of the biomarker follows normal distribution in both diseased and non-diseased subjects. This distributional assumption is restrictive as it does not always hold. Like the binormal model, the Lehman model provides smooth estimates of ROC curves and a closed form expression for the area under the ROC curve. However, unlike the binormal model, the Lehman model is easy to calculate and does not make any assumption about the underlying distribution of the biomarker. Further, the Lehmann model can be used to obtain covariate-adjusted ROC curves, and accommodates inference of correlated and longitudinal biomarkers. Thus, the Lehman model should be considered as an alternative when the normality assumption of the binormal model is questionable.

Acknowledgments

The authors are grateful for the financial support from NIH/NAIDS 2P30 AI060354-11. The authors also thanks the study participants and the principal investigator of the study for sharing the data with us.

References

[1] Hanley, JA ; Mcneil, BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology April 1982, Vol.143(1), pp.29-36*

[2] Margaret S. Pepe The Statistical Evaluation of Medical Tests for Classification and Prediction. New York, Oxford University Press 2003.

[3] L. Gonçalves, A. Subtil, and P. Bermudez, "ROC Curve Estimation: An Overview," REVSTAT – Statistical Journal*, vol. Vol. 12, no. No. 1, pp. 1–20, 2014.*

[4] Xiao-Hua Zhou, Nancy A. Obuchowski, Donna K. McClish. Statistical Methods in Diagnostic Medicine, 2nd ed, Jon Wiley & Sons 2011.

[5] M. Ghebremichael, H. Michael, J. Tubbs, and E. Paintsil, "Comparing the Diagnostics Accuracy of CD4+ T-Lymphocyte Count and Percent as a Surrogate Markers of Pediatric HIV Disease," Journal of Mathematics and Statistics*, vol. 15, pp. 55–64, Jan. 2019.*

[6] M. Gönen and G. Heller, "Lehmann Family of ROC Curves," Medical decision making : an international journal of the Society for Medical Decision Making*, vol. 30, no. 4, pp. 509–517, 2010.*

[7] Tianxi Cai and Chaya Moskowitz. Semi-parametric estimation of the binormal RO Ccurve for a continuous diagnostic test. Biostatistics 2004,5(4), 573586

[8] Cox DR 1972. Regression models and life tables (with discussion). J R Statist Soc B 34: 187220

[9] Cox DR 1975. Partial likelihood. Biometrika 62: 269-276.

[10] S. Pundir and R. Amala, "Evaluation of Area under the Constant Shape Bi-Weibull ROC Curve," Journal of Modern Applied Statistical Methods*, vol. 13, pp. 304–328, May 2014.*

[11] Debasis Kundu and Rameshwar D. Gupta. Estimation of P (Y< X) for Weibull distribution, IEEE Transactions on Reliability. July 2006, 55(2).

[12] Bradley Efron (1987) Better Bootstrap Confidence Intervals, Journal of the American Statistical Association, 82:397, 171-185.

[13] M. Ghebremichael, U. Larsen, and E. Paintsil, "Association of Age at First Sex With HIV-1, HSV-2, and Other Sexual Transmitted Infections Among Women in Northern Tanzania:," Sexually Transmitted Diseases, *vol. 36, pp. 570–576, Sept. 2009.*

[14] Ulla Larsen, Joseph Mlay, Said Aboud, Ronald Ballard, Noel E Sam, John F Shao, Saidi H Kapiga. Design of a community-based study of sexually transmitted infections/HIV and infertility in an urban area of northern Tanzania. Sex Transm Dis. 2007;34:2024.

[15] Ghebremichael M, Habtemicael S. Effect of tuberculosis on immune restoration among HIV-infected patients receiving antiretroviral therapy. J Appl Stat. 2018;45:23572364.

[16] P. Grambsch and T. Therneau (1994), Proportional hazards tests and diagnostics based on weighted residuals. Biometrika, 81, 515-26.

[17] Omar K. Siddiqi, Gretchen L. Birbeck, Musie Ghebremichael, Eugene Mubanga, Shawn Love, Clayton Buback, Barry Kosloff, Helen Ayles, Masharip Atadzhanov, Keertan Dheda, and Igor J. Koralnik. Prospective Cohort Study on Performance of Cerebrospinal Fluid (CSF) Xpert MTB/RIF, CSF Lipoarabinomannan (LAM) Lateral Flow Assay (LFA), and Urine LAM LFA for Diagnosis of Tuberculous Meningitis in Zambia. J Clin Microbiol. 2019 Aug; 57(8)