

# Lehmann Family of ROC Curves

**Mithat Gönen**

gonenm@mskcc.org

**Glenn Heller**

hellerg@mskcc.org

Department of Epidemiology and Biostatistics

Memorial Sloan-Kettering Cancer Center

307 East 63 Street, New York, NY 10021

## Summary

Receiver operating characteristic (ROC) curves evaluate the discriminatory power of a continuous marker to predict a binary outcome. The most popular parametric model for an ROC curve is the binormal model, which assumes that the marker, after an unspecified monotone transformation, is normally distributed conditional on the outcome. Here we present an alternative to the binormal model based on the Lehmann family, also known as the proportional hazards specification. The resulting ROC curve and its functionals (such as the area under the curve) have simple analytic forms. Closed-form expressions for the functional estimates and their corresponding asymptotic variances are derived. This family accommodates the comparison of multiple markers, covariate adjustments and clustered data through a regression formulation. Evaluation of the underlying assumptions, model fitting and model selection can be performed using any off the shelf proportional hazards statistical software

package.

*Key words:* Regression, accuracy, concordance, proportional hazards

## 1 Introduction

ROC curves have become the standard tool for evaluating the discriminatory power of medical diagnostic tests and they are commonly used in assessing the predictive ability of binary regression models. In a typical setting one has a binary indicator and a set of predictions or marker values. The goal is to see how well the marker values predict the binary indicator. The principal idea is to dichotomize the marker at various thresholds and compute the resulting sensitivity and specificity. A plot of sensitivity (true positive fraction or TPF) versus one minus specificity (false positive fraction or FPF) is the ROC curve. It provides a complete picture of various levels of sensitivity and specificity that can be achieved using the marker. When dealing with predictions from a regression model instead of a diagnostic marker, the same principle applies so we will use the term “marker” generically from this point on to refer to the variable for which an ROC curve is desired.

An empirical ROC curve may be obtained by connecting the observed (TPF, FPF) pairs. The area under the empirical ROC curve is a one-to-one function of the two-sample Wilcoxon statistic and Somers’ D (Pratt and Gibbons, 1981). The empirical curve is attractive because it makes minimal assumptions, but it does not generalize easily to allow covariate adjustments or clustered data. When such generalizations are needed, most analysts work with the binormal model. The binormal model assumes that the marker values follow a normal distribution, possibly after a monotone

transformation (Dorfman and Alf, 1996; Hanley, 1996). The normalizing transformation can be pre-specified or estimated from the data. In the latter case the Box-Cox transformation has been widely used in practice (Zou and Hall, 2000; Faraggi and Resier 2002). More recent work has extended the use of Box-Cox transformation in the case of covariate adjustment as well (Faraggi, 2003; Schisterman, 2004).

One notable exception is the work of Metz, Herman and Shen (1998) which uses the concept of “truth state runs” and uses a latent variable binormal model where each segment of the model is defined by a single truth state run. While this method does not require a transformation, truth state runs need to be of sufficient length to estimate the corresponding segment of the latent variable model. This can be problematic, especially in the case of covariate adjustments and clustered data.

The literature is replete with regression analyses of ROC curves, a framework which provides adjustments for covariates and clustering. A recent survey of this literature is given by Pepe (2003). The binormal model, after specifying the transformation, can be formulated as a regression model, with the marker value as the dependent variable and the disease status as the independent variable. This can be easily extended by adding covariates and covariate-disease status interactions to the right hand side of the model. The binormal model has the advantage of using familiar methods based on the normal distribution, but assumes a normalizing transformation can be determined.

Pepe (1998) classified ROC regression procedures under three headings: Modeling the marker values, modeling summary measures of accuracy and direct modeling of ROC curves. As she noted, modeling summary measures of accuracy does not allow for continuous covariates, hence it is not a regression model in the conventional sense.

The direct modeling of ROC curves, while making fewer assumptions, requires large sample sizes, is difficult to implement (Pepe, 2000 and Alonzo and Pepe, 2002), and lacks goodness of fit diagnostics. Modeling the marker values, however, has many practical advantages, including ease of implementation and the availability of model checking methods.

In this article we present a semiparametric approach, based on the proportional hazards specification, to modeling the marker values. The proposed method enables model fitting, inference, and diagnostics for model specification, using standard statistical software. The proportional hazards framework for the ROC analysis is presented in Section 2. Section 3 covers covariate adjustments, comparison of markers, and the incorporation of clustered data. Section 4 presents the analysis of the utility of chemical shift magnetic resonance imaging in differentiating normal and benign vertebral marrow processes using the proposed model. Section 5 contains a discussion and provides our conclusions.

## 2 Model

Let  $V$  be the marker,  $D = 0, 1$  be the binary indicator and let  $S_0 = S_{D=0}$  and  $S_1 = S_{D=1}$  denote the survival functions (one minus the cumulative distribution function) of the marker for the two different values of the binary indicator. A semiparametric relationship is proposed

$$S_1(v) = [S_0(v)]^\theta, \tag{1}$$

where the underlying survival distributions  $(S_1, S_0)$  are left unspecified, but their relationship is governed by a single parameter  $\theta$ . This semiparametric relationship

between survival distributions was originally proposed by Lehmann (1953). We will call (1) the Lehmann assumption and the resulting ROC curves, the Lehmann family of ROC curves. If subjects with and without disease are labeled  $D = 1$  and  $D = 0$  respectively, and subjects with disease are more likely to have higher marker values, then the survival functions for the two groups are oriented by the specification  $0 < \theta \leq 1$ . The parameter  $\theta^{-1}$  represents the odds that a subject belonging to the  $D = 1$  group has a higher marker value relative to a subject belonging to the  $D = 0$  group.

We will use  $x$  to denote the false positive fraction and  $y$  to denote the corresponding true positive fraction so that the  $(x, y)$  pairs form the ROC curve. The relationship between the false positive fraction and the true positive fraction, can be represented as

$$y = S_1(S_0^{-1}(x)), \quad x \in [0, 1]. \quad (2)$$

Using (1) in (2) yields the general form of the Lehmann family of ROC curves:

$$y = x^\theta. \quad (3)$$

We note that if  $0 < \theta < 1$ , then (3) is concave everywhere on the unit interval, a desirable property for ROC curves, since it implies a monotone increasing curve that lies above the 45-degree line. Figure 1 shows a spectrum of ROC curves belonging to this family.

An alternative form for the Lehmann relationship between two groups is based on the hazard function. Defining the hazard function as

$$h(v) = \lim_{\Delta v \rightarrow 0} \frac{\Pr(v \leq V < v + \Delta v | V \geq v)}{\Delta v}$$

the Lehmann specification in (1) may be rewritten as

$$\frac{h(v)}{\tilde{h}(v)} = \theta. \quad (4)$$

Note that in this case  $h = h_{D=1}$  and  $\tilde{h} = h_{D=0}$ , but the general notation will be helpful in subsequent sections. The identity (4) is the reason the Lehmann relationship is referred to as the proportional hazards specification (Cox, 1972, 1975). This connection to proportional hazards model provides an extensive body of literature and software for the estimation and inference of the odds parameter  $\theta$ .

Cox regression modules in statistical software can be used for this purpose using  $V$  as the outcome and  $D$  as the independent variable. Formally, we set

$$h_1(v, D) = h_0(v) \exp\{\beta D\}$$

and  $\theta = e^\beta$ . One can estimate  $\beta$ , and consequently,  $\theta$ , using the Cox partial likelihood.

We will use  $\hat{\beta}$  for the partial likelihood estimate and

$$V(\hat{\theta}) = \exp\{2\hat{\beta}\}V(\hat{\beta})$$

for its estimated variance, where  $V(\hat{\beta})$  is computed as the inverse of the information matrix from the partial likelihood.

Estimation and inference of the ROC curve and continuous measures of the curve, are derived from the proportional hazards framework. For example, the pointwise variance estimate of the smooth ROC curve is, using the delta method, given by

$$V(y(x)) = \left[ x^{\hat{\theta}} \log x \right]^2 V(\hat{\theta}). \quad (5)$$

The area under the ROC curve is estimated as

$$\widehat{AUC} = \int_0^1 x^{\hat{\theta}} dx = (\hat{\theta} + 1)^{-1} \quad (6)$$

and its variance is estimated by

$$V(\widehat{AUC}) = (\hat{\theta} + 1)^{-4}V(\hat{\theta}). \quad (7)$$

Finally, the partial area under the curve up to  $x_0$ ,  $pAUC(x_0)$ , can be estimated using

$$\widehat{pAUC}(x_0) = \int_0^{x_0} = (\hat{\theta} + 1)^{-1}x_0^{\hat{\theta}+1} \quad (8)$$

with variance estimate

$$V(\widehat{pAUC}(x_0)) = \left( \frac{x^{\hat{\theta}+1}}{\hat{\theta} + 1} \right)^2 \left[ \frac{[x^{\hat{\theta}} \log x]^2 V(\hat{\theta})}{(x^{\hat{\theta}+1})^2} + \frac{V(\hat{\theta})}{(\hat{\theta} + 1)^2} - \frac{2x^{\hat{\theta}+1} \log x V(\hat{\theta})}{x^{\hat{\theta}+1}(\hat{\theta} + 1)} \right]. \quad (9)$$

Although the ROC curve is generically represented as a function of survival functions, the Lehmann specification of the ROC curve, given by (3), depends only on the odds parameter  $\theta$ , and does not require the estimation of survival functions explicitly. In addition, there are several methods developed and implemented for model diagnostics (Lin, 1991; Grambsch and Therneau, 1994) that can assist the analyst in determining if the proportional hazards assumption is warranted for the specific ROC analysis. A graphical approach for checking the proportional hazards specification, based on the partial sums of the residuals, is demonstrated in our data example in Section 4.

### 3 Further Applications of Regression

The Lehmann specification of the ROC curve lends itself to extensions in several important contexts: covariate adjustment, comparison of ROC curves for several markers, and clustered data. All of these can be represented in a proportional hazards regression framework, as discussed in this section.

### 3.1 Covariate Adjustments

Covariate adjustment is important in ROC analysis when the marker threshold for group membership is a function of a concomitant covariate. For example, the Prostate Specific Antigen (PSA) level is a validated marker for prostate cancer. PSA, however, increases as men age. Thus, an adjustment for age would improve an ROC analysis using PSA as a marker for prostate cancer.

Tosteson and Begg (1988) showed that a regression model with an interaction term can be used to estimate a covariate-adjusted ROC curve. In the context of the Lehmann family this amounts to a proportional hazards regression model,

$$h(V|D, U) = \tilde{h}(v) \exp\{\beta_1 D + \beta_2 U + \beta_3 DU\} \quad (10)$$

with  $U$  as the concomitant covariate. The ratio of the two hazard models with group membership  $D = 1$  and  $D = 0$  results in

$$\frac{h(V|D = 1, U)}{h(V|D = 0, U)} = e^{\beta_1 + \beta_3 U}, \quad (11)$$

which yields the covariate-adjusted ROC curve

$$y(u, x) = x^{\theta(u)} \quad (12)$$

where

$$\theta(u) = \exp\{\beta_1 + \beta_3 U\} \quad (13)$$

The interaction between  $D$  and  $U$  in the model enables the hazard ratio to reflect the effect of the covariate  $U$ , otherwise the right hand side of (11) would simply be  $e^{\beta_1}$ .

The use of the interaction term in the ROC analysis is not specific to the proportional hazards model and can be observed in all regression models following the Tosteson-Begg approach. Note that expressions (5-9) still hold when  $\hat{\theta}$  is replaced by  $\hat{\theta}(u)$ ,

which itself is a contrast that can be estimated from the underlying regression model along with its standard error. Covariate adjustment can be extended to multiple covariates using (10).

### 3.2 Clustered Data

Clustered data arise naturally in many radiologic imaging studies. As technology advances, so-called full-body scans render multiple evaluations possible for each patient. For example, for a cancer patient one may evaluate the primary tumor, local lymph nodes and distant metastases all on the same scan leading to clustered data. It is possible to obtain ROC curves within the Lehmann family taking into account the clustering. Let  $k = 1, \dots, K$  index the observations on the same patient and consider the following marginal model

$$h_k(V|D) = \tilde{h}_k(v) \exp\{\beta_k D\} \quad (14)$$

which gives rise to an ROC curve for each  $k$

$$y_k(x) = x^{\theta_k} \quad (15)$$

where, again,  $\theta_k = \exp\{\beta_k\}$ . Assuming the Lehmann specification holds, the estimates  $(\hat{\beta}_1, \dots, \hat{\beta}_K)$ , derived from the marginal partial likelihood score functions, are consistent. The covariance between  $(\hat{\beta}_k, \hat{\beta}_l)$  can be consistently estimated by  $\hat{v}_{kl} = \hat{a}_k^{-1} \hat{w}_{kl} \hat{a}_l^{-1}$ , where  $\hat{a}_k$  is the negative second derivative from the  $k$ th partial likelihood and  $\hat{w}_{kl}$  is the estimated covariance from the  $k$ th and  $l$ th partial likelihood score functions. Due to its form, the estimated variance matrix  $\hat{V}$ , which is composed of its elements  $\hat{v}_{kl}$ , is sometimes called a sandwich estimator.

It is possible to use covariates on the right hand side of (14) in exactly the same way as in (10). If  $U$  denotes the covariate of interest then the following marginal model will provide covariate-adjusted ROC curves under clustering:

$$h_k(V|D, U) = \tilde{h}_k(v) \exp\{\beta_{1k}D + \beta_{2k}U + \beta_{3k}DU\}. \quad (16)$$

This model can be fit for each  $k$  using the partial likelihood for the point estimates and the sandwich estimator for the variance estimate, where  $\hat{A}_k$  and  $\hat{W}_{kl}$  are now matrices.

### 3.3 Comparing the ROC Curves of Several Markers

The comparison of two markers is an important case of covariate adjustment. In radiology, a new imaging technique (such as positron emission tomography) may be in competition with standard of care (such as computed tomography) in detecting disease. In the field of biomarkers it may be of interest to compare several ways of evaluating a marker. An example from the field of prostate cancer surveillance is whether a baseline PSA measurement or the change in PSA over time, summarized by an index such as PSA velocity, is a better predictor of disease recurrence. In prediction modeling, there may be competing models. For example, using the same data one may use different statistical techniques to make predictions such as logistic regression, classification trees, or neural networks. Another possibility is that one might have an emerging predictor variable such as a genetic variant, and it would be of interest to see if a prediction model using the new predictor variable along with the traditional variables is better than one that uses traditional variables only.

One practical aspect where marker comparison differs from other covariate adjust-

ments is study design. Most marker comparison studies are paired in nature because it is usually feasible to evaluate the competing markers within patient. The marginal model approach for clustered data, discussed in the previous section, can be applied to this study design. For example, the comparison of two markers ( $k = 1, 2$ ) is accomplished using the contrast  $c^T\beta$ , where  $c^T = (1, -1)$  and  $\beta^T = (\beta_1, \beta_2)$  are the coefficients from the marginal model(14). A Wald test for the equality of two ROC curves is constructed by  $\hat{C}/\sqrt{\text{Var}(\hat{C})}$ , where  $\hat{C}$  represents the marginal partial likelihood contrast estimate  $\hat{\beta}_1 - \hat{\beta}_2$  and  $\text{Var}(\hat{C})$  is obtained using the sandwich estimator described in section 3.2.

## 4 Simulation Study

We conducted a simulation study to evaluate the effects of model misspecification on the Lehmann family and the binormal model. Data were generated from a Weibull distribution to satisfy the Lehmann family specification and a normal distribution that met the binormal model assumptions. The results of the Weibull and normal ROC simulations, based on the estimated AUC, the standard error of the AUC estimate, and the simulation standard error of the AUC estimate, are summarized in Tables 1 and 2. One thousand replicates were run for each simulation. For the results of the Weibull simulations in Table 1, 100 marker values were generated for diseased and 100 marker values for non-diseased individuals. The shape parameter in both groups was set to 1. The scale parameter was 1 for the non-diseased group, and was varied in the diseased group to obtain a range of AUC values. For the normal data simulations presented in Table 2, the 100 marker values in the non-diseased group

were generated from a standard normal and the 100 marker values from the diseased group were generated from a normal with mean  $\mu$  and variance 1, where  $\mu$  dictated the AUC value.

For the Weibull simulations in Table 1, both methods provide unbiased estimates of the AUC. For the binormal model, the estimates were computed after applying a Box-Cox transformation (Zou and Hall, 2000), and the estimated standard error was derived ignoring the additional variability caused by the estimation of the Box-Cox parameter (Wieand et al., 1989). Without the Box-Cox transformation, there was substantial bias in the binormal model (data not shown). The asymptotic standard error of the Lehmann family estimate was very close to the simulation standard error, confirming the validity of the Lehmann approach. In contrast, the binormal standard error overestimated the true variability when the signal was weak and underestimated the standard error when the signal was strong. We conclude that for the Weibull simulations, the Box-Cox transformation removes the bias in the binormal AUC estimates, but the estimated asymptotic standard error for the AUC estimate is unreliable.

Table 2 shows that the binormal model performed well when the model was correctly specified, displaying no bias and minor deviations from the simulation standard error. The Lehmann model, however, was biased, although the variability estimate of the estimated AUC was good even though the model was incorrectly specified.

Overall, these simulations support the premise that Lehmann family is preferred if the data follow the proportional hazards model and similarly the binormal model should be retained if the normal distribution is valid. The results underscore the importance of model checking before proceeding with the analysis.

## 5 Example

Zajick et al (2005) report a study on the utility of chemical shift magnetic resonance in differentiating normal, benign and malignant vertebral marrow processes. The marker of interest is the percent difference between the in-phase and out-phase signal intensities. Their focus was on establishing a range of values for signal intensity change in normal vertebral marrow. Here we use their data for an objective that has not been pursued in their article: evaluating the ability of signal intensity change in discriminating between normal and benign vertebral marrow processes.

A total of 569 normal vertebrae were evaluated on 75 patients, as compared with 215 benign lesions in 92 patients. Figure 2 presents the histograms of the signal intensity change for normal and benign vertebrae separately. The two distributions have some overlap suggesting that the marker may not have the ability to discriminate the two classes. The empirical ROC points, represented with open circles in Figure 3, verifies this suspicion since it is only slightly better than the diagonal line. Prior to the employment of the Lehmann based ROC curve, it is prudent to confirm the proportional hazards assumption. The thick line in Figure 4 is the observed score process and the dotted lines are 100 sample paths generated from the score process under the proportional hazards assumption (Lin, 1993). Since the observed process is typical of the sample paths obtained under the model, there is no evidence against proportional hazards between normal and benign patients, validating the assumptions underlying the ROC curves in Figure 3.

In our first analysis we ignore the fact that patients contribute multiple vertebrae to the analysis and assume that the signal intensity change is independent across

vertebrae, conditional on the gold standard (normal/benign). Using the partial likelihood,  $\hat{\beta} = -0.355$  ( $\hat{\theta} = 0.701$ ) with a standard error of 0.088. The resulting member of the Lehmann family of ROC curves is plotted with a solid line and the dotted lines around it represent the asymptotic pointwise 95% confidence intervals. We then obtained  $\hat{\beta}$  using estimating equations to adjust for the clustering due to multiple observations contributed for each patient. The coefficient  $\beta$  is again estimated to be  $-0.355$  but the standard error is now 0.144. The wider set of dotted lines in Figure 3 represent the confidence intervals obtained using the marginal model.

The area under the curve is 0.588 with a standard error of 0.030 (ignoring clustering) or 0.050 (adjusted for clustering). The corresponding confidence limits are (0.529, 0.647) and (0.490, 0.686) confirming the difficulty of distinguishing normal and benign processes. In contrast, the area under the empirical curve is 0.597 with a standard error of 0.025, which is very close to the estimates obtained above ignoring clustering. We note that the data in this example is consistent with the binormal model as well and the AUC estimated from a binormal model (ignoring clustering) is 0.605 with a standard error of 0.025.

Finally, the ROC analysis is adjusted for age. Typically, vertebral marrow processes are more difficult to image in older patients, due to the effects of aging on the vertebrae confounding disease-related abnormalities. We first fit the proportional hazards regression model (10) with  $U$  representing age measured in years. The resulting parameter estimates and standard errors (in parentheses) are given in Table 3. While there is some evidence that the ROC curve is a function of age,  $\hat{\beta}_3$  is not significantly different from 0, especially when clustering is taken into account. Figure 5 displays the AUC as a function of age, making clear the decreasing discriminatory

power of the percent difference in signal intensities in older patients.

## 6 Discussion

In this article we presented a model based method to obtain smooth ROC curves. The model is based on the Lehmann (or proportional hazards) assumption and can accommodate a variety of research questions such as covariate adjustments and clustered data. All the analyses can be performed with the built-in functionality of off-the-shelf software. The approach does not require a full parametric specification of the distribution of the marker values for the two reference populations. The price for this flexibility is a loss of efficiency relative to an analysis based on a correctly specified parametric model (Oakes, 1977).

A popular alternative approach is the binormal model, which assumes that the marker values are normally distributed, possibly after being subjected to a monotone transformation. Since these two assumptions do not overlap one can consider the proportional hazards and the binormal model to be complementary.

The Lehmann assumption is equivalent to assuming the existence of a monotone transformation producing marker values with an extreme value distribution (Kalbfleisch, 1978), but does not require that the transformation is specified or even estimated. Thus, the Lehmann specification is more robust than the binormal model. Conversely, the Lehmann family of ROC curves is indexed by a single parameter and affords less flexibility than the two parameter binormal model.

The proposed model has two major advantages for the practicing statistician. Both of these advantages stem from the regression representation. The first advan-

tage is operational. The proportional hazards model has become the primary vehicle for the analysis of survival data, and all mainstream statistical packages provide estimates, inferences, and model diagnostics for this model and hence the resulting ROC analysis. The second advantage is conceptual. It is possible to formulate most practical ROC problems using a regression model. For example, simultaneous modeling and comparison of two or more markers can be seen as a regression problem with dummy variables. Covariate adjustment, which is sometimes necessary because a covariate is thought to influence the accuracy of the marker, is naturally modeled through a regression framework. Clustered data, with individuals contributing multiple marker data, can be analyzed using marginal regression models that enable a robust variance estimate. Each of these ROC analyses can be performed using the available proportional hazards software.

## References

- Alonzo, T.A. and Pepe, M.S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics* **3**, 421-432.
- Cox, D.R. (1972). Regression models and life tables (with Discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187-220.
- Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269-76.
- Dorfman, D.D. and Alf, E. (1968). Maximum likelihood estimation of parameters of signal detection theory—a direct solution. *Psychometrika* **33**, 117-24.
- Faraggi, D. (2003). Adjusting ROC Curves and Related Indices for Covariates. *The Statistician*. **52**, 179-192.

- Faraggi, D. and Reiser, B. (2002). Estimation of the area under the ROC curve. *Statistics in Medicine* **30**, 3093-3106.
- Grambsch, P. M. and Therneau, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*, **81**, 515-526.
- Hanley, J. A. (1969). Confidence intervals-rating method data. [Binormal ROC curve-ordinal data]. *Journal of Mathematical Psychology*, **6**, 487-496.
- Hanley, J. A. (1988). The robustness of the binormal model used to fit ROC curves. *Medical Decision Making* **8**, 197-203.
- Hanley, J. A. (1996) The use of the 'binormal' model for parametric roc analysis of quantitative diagnostic tests. *Statistics in Medicine*, **15**, 1575-1585.
- Kalbfleisch, J. D. (1978). Likelihood Methods and Nonparametric Tests *Journal of the American Statistical Association*, **73**, 167-170.
- Lehmann, E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics* **24**, 23-43.
- Lin, D. Y. (1991). Goodness-of-Fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators. *Journal of the American Statistical Association* **86**, 725-728.
- Metz, C. E., Herman, B.A., Shen, J-H. (1998). Maximum-likelihood estimation of ROC curves from continuously-distributed data. *Statistics in Medicine* **17**, 1033-1053.
- Oakes, D. (1977). The Asymptotic Information in Censored Survival Data. *Biometrika* **64**, 441-448.
- Pepe, M. S. (1998). Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. *Biometrics* **54**, 124-135.

- Pepe, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56**, 352-359.
- Pepe, M. S. (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford Statistical Science Series, Oxford: Oxford University Press.
- Pratt, J. W. and Gibbons, J. D. (1981). *Concepts of Nonparametric Theory*. New York: Springer Verlag.
- Schisterman, E. F., Faraggi, D., and Reiser, B. (2004). Adjusting the generalized ROC curve for covariates *Statistics in Medicine*. **23**, 3319-3331.
- Somers, R. H. (1962). A similarity between Goodman and Kruskal's tau and Kendall's Tau, with a partial interpretation of the latter. *Journal of the American Statistical Association* **57**, 804-12.
- Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84**, 1065-1073.
- Tosteson, A. N. and Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **8**, 204-15.
- Wieand, S., Gail, M. H., James, B. R., and James, K. L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* **76**, 585-592.
- Zajick, D. C., Morrison, W. B., Schweitzer, M. E., Parellada, J. A., and Carrino, J. A. (2005). Benign and malignant processes: normal values and differentiation with chemical shift MR imaging in vertebral marrow. *Radiology*, **237**, 590-596.
- Zou, K. H. and Hall, W. J. (2000). Two transformation models for estimating an ROC curve derived from continuous data *Journal of Applied Statistics* **27**, 621-631.

**Table 1:** Simulation results for the Weibull case.

True AUC	Lehmann Family			Binormal Model		
	Estimate	SE	Simulation SE	Estimate	SE	Simulation SE
0.50	0.500	0.0355	0.0368	0.533	0.0396	0.0242
0.55	0.549	0.0353	0.0363	0.552	0.0394	0.0336
0.60	0.600	0.0347	0.0347	0.598	0.0384	0.0389
0.65	0.652	0.0336	0.0338	0.654	0.0366	0.0379
0.70	0.701	0.0321	0.0316	0.704	0.0343	0.0350
0.75	0.751	0.0299	0.0290	0.750	0.0316	0.0323
0.80	0.800	0.0270	0.0274	0.800	0.0278	0.0312
0.85	0.851	0.0233	0.0229	0.853	0.0229	0.0264
0.90	0.900	0.0187	0.0178	0.899	0.0175	0.0220
0.95	0.950	0.0124	0.0121	0.955	0.0094	0.0140

**Table 2:** Simulation results for the normal case.

True AUC	Lehmann Family			Binormal Model		
	Estimate	SE	Simulation SE	Estimate	SE	Simulation SE
0.960	0.947	0.0123	0.0139	0.959	0.0086	0.0120
0.910	0.911	0.0173	0.0174	0.910	0.0161	0.0202
0.856	0.807	0.0250	0.0302	0.856	0.0224	0.0259
0.793	0.828	0.0261	0.0255	0.793	0.0284	0.0319
0.749	0.785	0.0290	0.0278	0.750	0.0315	0.0341
0.691	0.724	0.0319	0.0328	0.691	0.0350	0.0377
0.647	0.724	0.0332	0.0307	0.648	0.0369	0.0361
0.571	0.642	0.0355	0.0339	0.571	0.0391	0.0365
0.536	0.616	0.0361	0.0336	0.543	0.0395	0.0298
0.500	0.500	0.0355	0.0368	0.534	0.0396	0.0253

**Table 3:** Age-adjusted parameter estimates and standard errors

Model	Clustering	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
Age (years)	Ignored	-1.288 (0.446)	-0.017 (0.006)	0.014 (0.007)
Age (years)	Adjusted	-1.288 (0.894)	-0.017 (0.009)	0.014 (0.013)

## Figures

**Figure 1:** Members of the Lehmann family with parameters ranging from 0.1 (closest to 45-degree line) to 0.9.

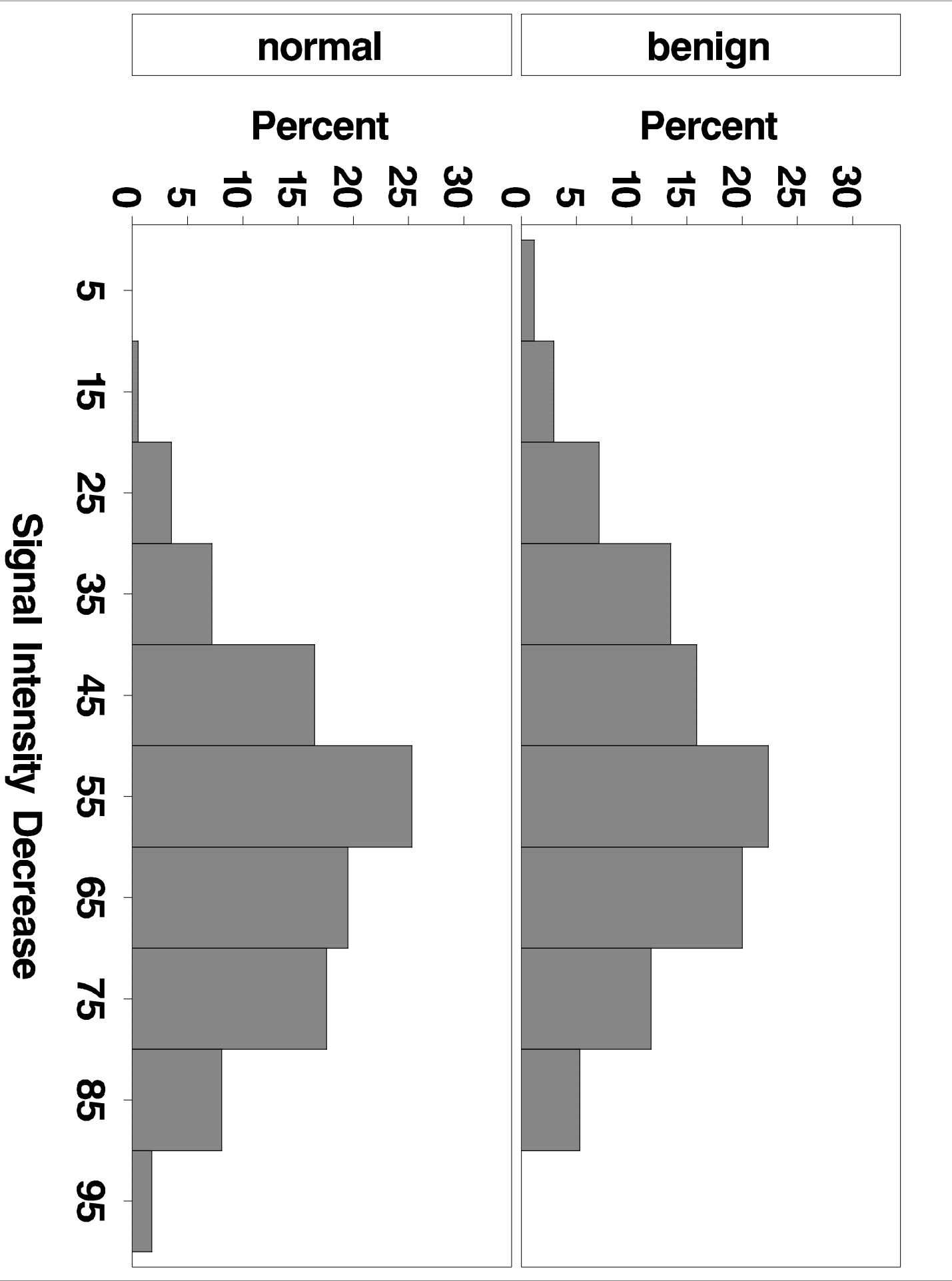
**Figure 2:** Histogram of percent difference between the in-phase and out-phase signal intensities for normal and benign vertebrae

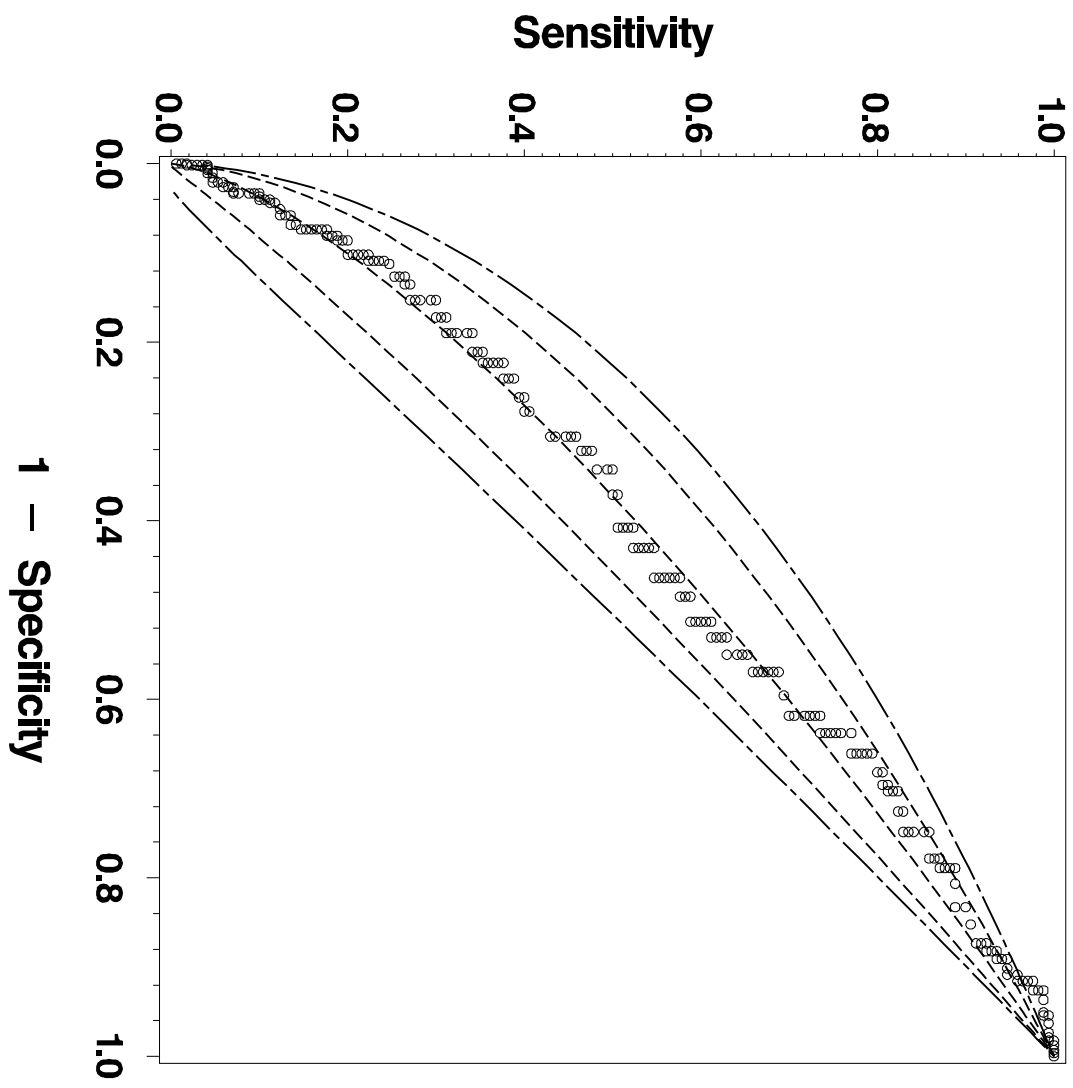
**Figure 3:** Empirical ROC points (open circles), smooth ROC curve (solid line) and 95% pointwise confidence limits using the partial likelihood (narrower dotted lines) and marginal model (wider dotted lines)

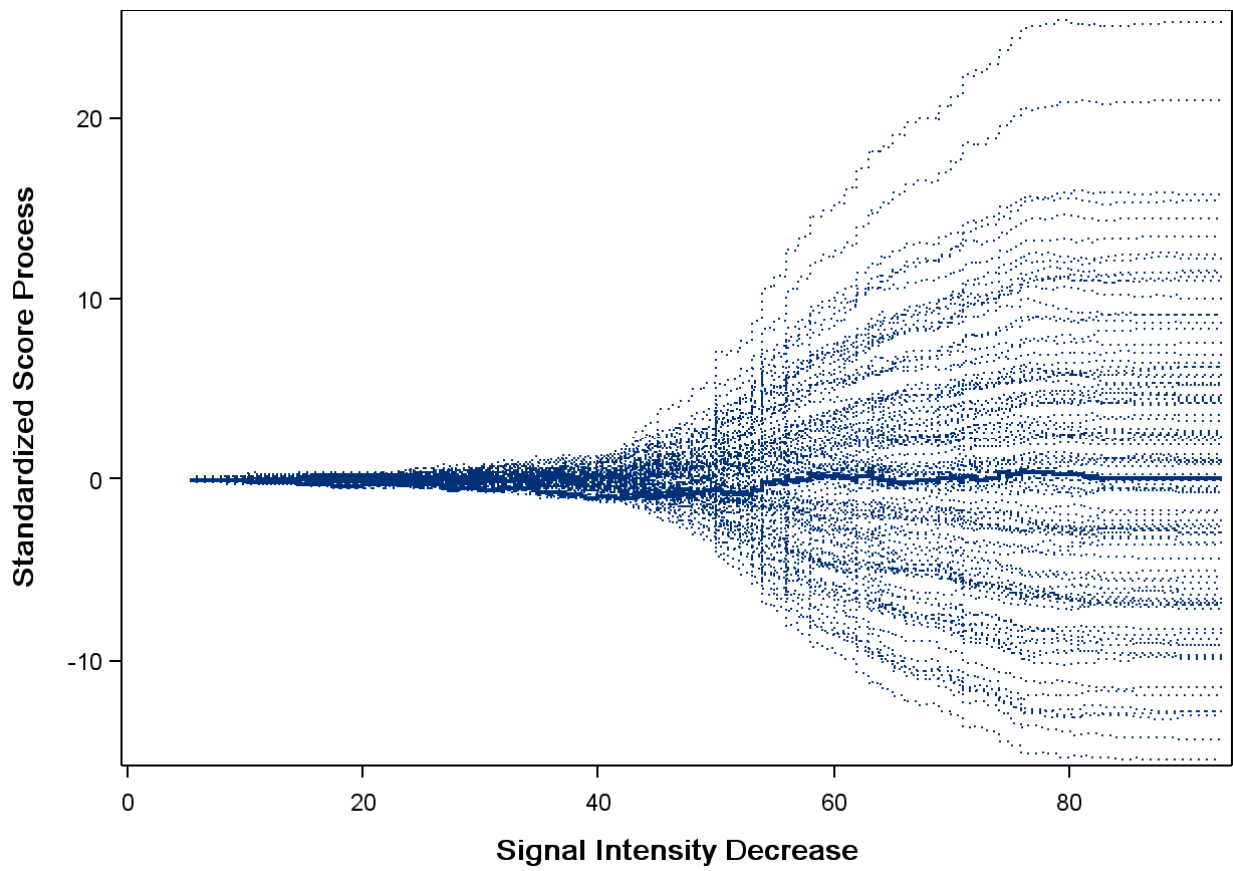
**Figure 4:** Score process for checking the assumption of proportional hazards.

**Figure 5:** AUC as a function of age.









# Area Under the Lehmann ROC Curve

