Memorial Sloan-Kettering Cancer Center
Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology
& Biostatistics Working Paper Series

Year 2006 *Paper* 11

Lehmann Family of ROC Curves

Mithat Gonen* glenn heller†

*Memorial Sloan-Kettering Cancer Center, gonenm@mskcc.org

†Memorial Sloan-Kettering Cancer Center, hellerg@mskcc.org

http://biostats.bepress.com/mskccbiostat/paper11

Lehmann Family of ROC Curves

Mithat Gonen and glenn heller

Abstract

Receiver operating characteristic (ROC) curves are useful in evaluating the ability of a continuous marker in discriminating between the two states of a binary outcome such as diseased/not diseased. The most popular parametric model for an ROC curve is the binormal model which assumes that the marker 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. We derive closed-form expressions for the asymptotic variances of the estimators for various quantities of interest. This family easily accommodates comparison of multiple markers, covariate adjustments and clustered data through a regression formulation. Evaluation of the underlying assumptions, model fitting and model selection can all be performed using any off the shelf proportional hazards statistical software package.

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 are useful in evaluating the ability of a continuous marker in discriminating between the two states of a binary outcome such as diseased/not diseased. The most popular parametric model for an ROC curve is the binormal model which assumes that the marker 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. We derive closed-form expressions for the asymptotic variances of the estimators for various quantities of interest. This family easily accommodates comparison of multiple markers, covariate adjustments and clustered data through a regression formulation. Evaluation of the underlying assumptions, model fitting and model selection can all 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 a model, assuming that an explicit monotone transformation of the marker values follow a normal distribution. This gives rise to

the binormal model (Dorfman and Alf, 1996; Hanley, 1996).

The literature is replete with regression analyses of ROC curves, a framework which provides adjustments for covariates and clustering. A survey of this literature is found in 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 regression parameters can be estimated using least squares or maximum likelihood. The binormal model has the advantage of using familiar methods based on the normal distribution, but requires the stringent assumption that the normalizing transformation is specified.

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. She further notes that direct modeling of ROC curves, while making fewer assumptions than the other two, requires large sample sizes, is difficult to implement, and lacks methods for model checking. Finally, modeling the marker values has many practical advantages, including ease of implementation and availability of model checking methods. The chief disadvantage is that it requires full specification of a parametric model for the conditional distribution of marker values.

In this article we present an alternative semiparametric approach to modeling ROC curves that incorporates covariates. Under Pepe's classification it would be considered an example of modeling the marker values and carries the advantages of

model fitting, inference, and diagnostics for model specification, using most popular statistical software packages. The model is semiparametric and is based on the proportional hazards specification. 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 family. 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

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

where the underlying survival distributions (S_1, S_0) are left unspecified, but their relationship is governed by a single parameter θ. The parameter θ 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. The survival functions for the two groups are oriented by assuming $\theta \geq 1$. 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.

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

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

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

$$y = x^{\theta}. \tag{3}$$

We note that (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

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

the Lehmann specification in (1) may be rewritten as

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

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 θ.

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

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

and $\theta = e^{\beta}$. One can estimate β, and consequently, θ, 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

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

The area under the ROC curve is estimated as

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

and its variance is estimated by

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

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

$$\widehat{pAUC}(x_0) = \int_0^{x_0} x^{\theta}\, dx = (\hat{\theta} + 1)^{-1} x_0^{\hat{\theta}+1} \tag{8}$$

with variance estimate

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

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 θ, 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,

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

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

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

which yields the covariate-adjusted ROC curve

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

where

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

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^{β_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 any number of covariates using (10).

3.2 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.

If interest is in comparing markers, one can use (10) with U as an indicator variable denoting the two markers to be compared. This single model gives rise to an ROC curve for each marker under consideration: x_1^{θ} and $x^{\theta_1+\theta_3}$, where $\theta_i = e^{\beta_i}$. In addition, this approach provides a direct way to test whether the two ROC curves are different using the null hypothesis $H_0 : \beta_3 = 0$. A Wald, score, or likelihood ratio statistic can be constructed from the partial likelihood and the maximum partial likelihood estimates to test this null hypothesis. It is likely that the software used to estimate the model parameters will provide one or more of these tests by default. This approach can be extended to multiple markers using a matrix U of dummy variables within the

regression framework.

One practical aspect where marker comparison differs from other covariate adjustments is study design. Most marker comparison studies are paired in nature because it is usually feasible to evaluate the competing markers within patient. As a result, the covariance between multiple marker measurements needs to be taken into account. One approach is to use the marginal proportional hazards model. In this case, the estimates derived from (10) are obtained by solving an estimating equation rather than maximizing the partial likelihood. This method is explained in more detail in the next subsection.

3.3 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 metastatases all on the same scan leading to clustered data. For $k = 1, \ldots, K$ markers, the marginal proportional hazards regression models

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

are applied to produce marker specific covariate adjusted ROC curves

$$y_k(u,x) = x^{\theta_k(u)}.$$

Assuming the Lehmann specification is correct, the partial likelihood estimates are consistent, but $(\hat{\theta}_1, \ldots, \hat{\theta}_K)$ are correlated, requiring an adjustment to the variance calculation. The asymptotic variance is estimated by $\hat{V}\hat{W}\hat{V}$, where $\hat{V}$ is the variance

matrix from the marginal partial likelihoods and the matrix $\hat{W}$ provides the between marker correlation information (Wei et al. 1989). This marginal approach is applicable when the data are paired, or for any data set where each patient contributes multiple observations.

4 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 which represents the ROC curve of a coin flip. Before we can use the Lehmann curve, we need to confirm that proportional hazards assumption is not violated. 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 assumption of proportional hazards (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 partial likelihood we find that $\hat{\beta} = -0.355$ with a standard error of 0.088, which corresponds to $\hat{\theta} = 0.701$. 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 contribued for each patient. The coefficient β 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.

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 as age measured in years. Then we dichotomize

age at the observed median of 58 years and re-fit the model with age as a binary variable. The resulting parameter estimates and standard errors (in parentheses) are given in Table 1, and Figure 5 provides adjusted ROC curves when age is dichotomized. 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. This is also evident from Figure 5.

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)
Age (> 58)	Ignored	-0.663 (0.139)	-0.508 (0.159)	0.430 (0.180)
Age (> 58)	Adjusted	-0.663 (0.268)	-0.508 (0.204)	0.430 (0.301)

5 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 full parametric specification of the distribution of the marker value for the two reference populations, hence it is less restrictive than the fully-parametric models used for the same purpose. A popular alternative approach is the binormal model, which assumes that a transformation is available or can be derived to produce normally distributed marker values. Since these two assumptions do not intersect 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, unlike the binormal model which necessitates that the normalizing transformation be known or derived. Conversely, the Lehmann family of ROC curves is indexed by a single parameter and affords less flexibility than the binormal model which has two parameters.

The proposed model has two major advantages for the practicing statistician. Both of these advantages stem from the regression representation. The first advantage is operational. The proportional hazards model has become the primary vehicle for the analysis of censored data, and all mainstream statistical packages provide estimates, inferences, and model diagnostics for this model and the resulting ROC analysis. The second advantage is conceptual. It is possible to formulate most practical problems of interest in the regression framework. For example, simultaneous modeling and comparison of two or more markers can be seen as a regression problem with dummy variables. Covariate adjustments, which are sometimes necessary because a covariate is thought to influence the accuracy of the marker, are naturally modeled as regression problems. All of these regression problems, with multiple independent variables, can be easily fit using the available software as mentioned above. In addition, clustered data can also be handled within this framework. Marginal models have become standard in recent years and it is possible with most statistical software to obtain a robust estimate of the variance, which can be used when individual subjects contribute multiple observations to the analyses.

References

Cox, D.R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187-220.

Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269-76.

Dorfman D.D., Alf E. (1968). Maximum likelihood estimation of parameters of signal detection theory–a direct solution. *Psychometrika* 33:117-24.

Grambsch, P.M., 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. *Med Dec 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 *J Amer Stat Assoc*, 73:167-170.

Lin DY. (1991). Goodness-of-Fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators. *J Amer Stat Assoc* 86:725–728.

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 (2003) *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford Statistical Science Series, Oxford University Press.

Pratt, J.W. & Gibbons, J.D. (1981). *Concepts of Nonparametric Theory.* Springer-

Verlag, New York.

Somers, R.H. (1962). A similarity between Goodman and Kruskal's tau and Kendall's Tau, with a partial interpretation of the latter. *J. Am. Statist. Assoc.* 57:804-12.

Wei, L. J.; Lin, D. Y.; Weissfeld, L. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *J Amer Stat Assoc* 84:1065-1073, 1989.

Tosteson A.N., Begg C.B. (1988). A general regression methodology for ROC curve estimation. *Med Decis Making.* 8:204-15.

Zajick DC, Morrison WB, Schweitzer ME, Parellada JA, Carrino JA. (2005). Benign and malignant processes: normal values and differentiation with chemical shift MR imaging in vertebral marrow. *Radiology*, 237:590-6.

Zhou XH, McClish DK Obuchowski, NA. (2003). *Statistical Methods in Diagnostic Medicine.* Wiley, New York.

Figures

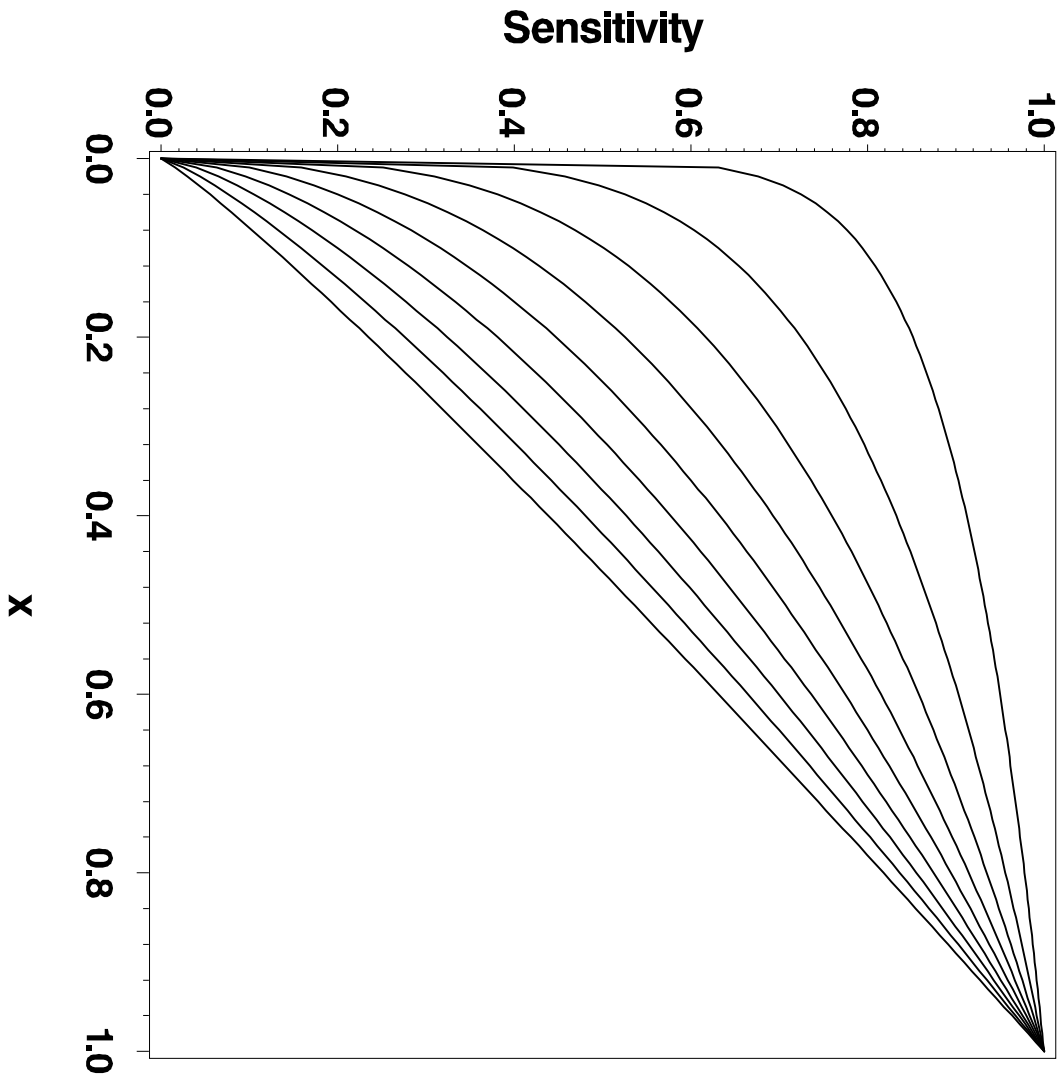
Figure 1: Members of the Lehmann family with parameters ranging from 0.1 (closest to 45-degree line) to 0.9.

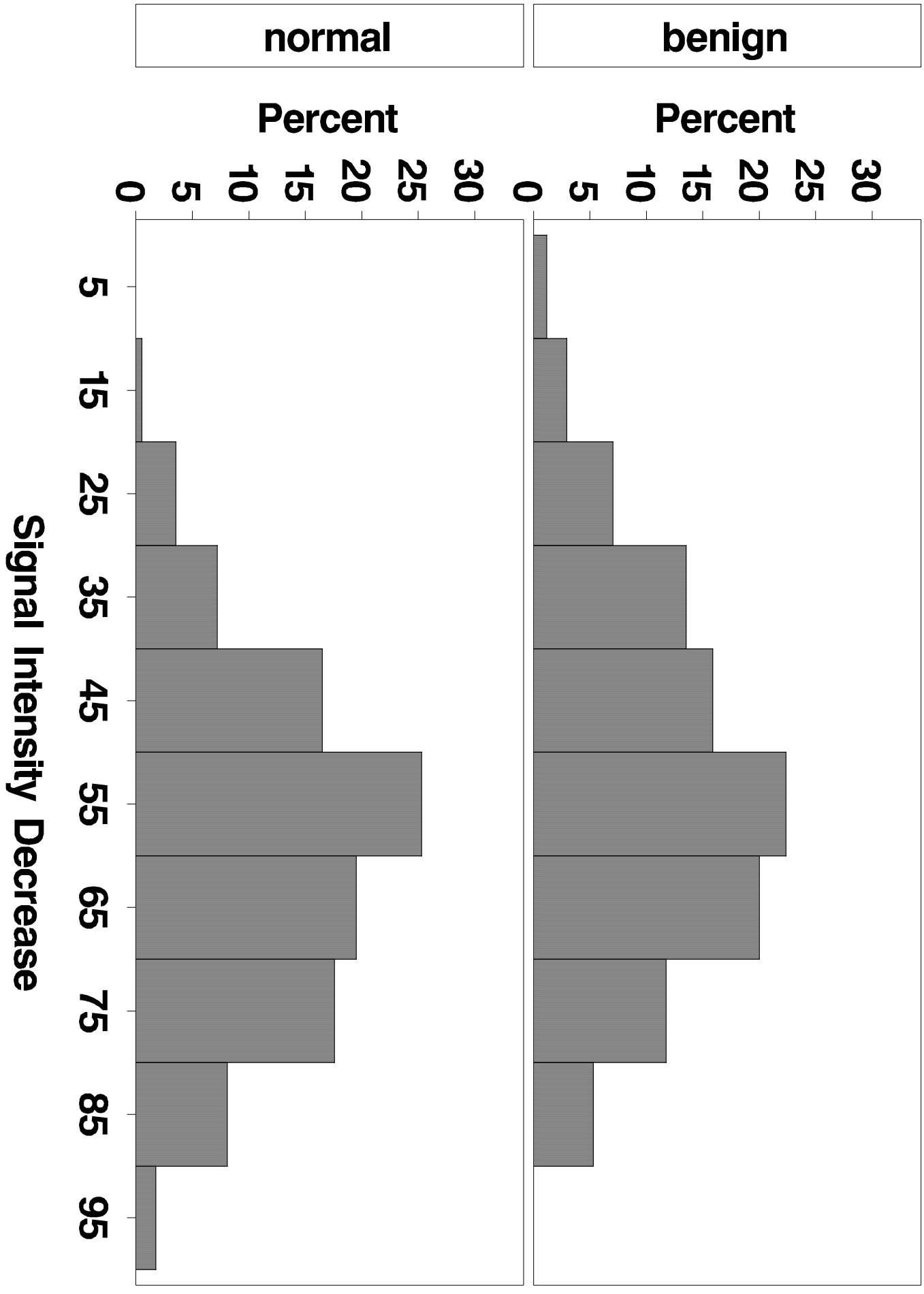
Figure 2: Histogram of percent difference between the in-phase and out-phase signal intensities for normal and benign vertebrae

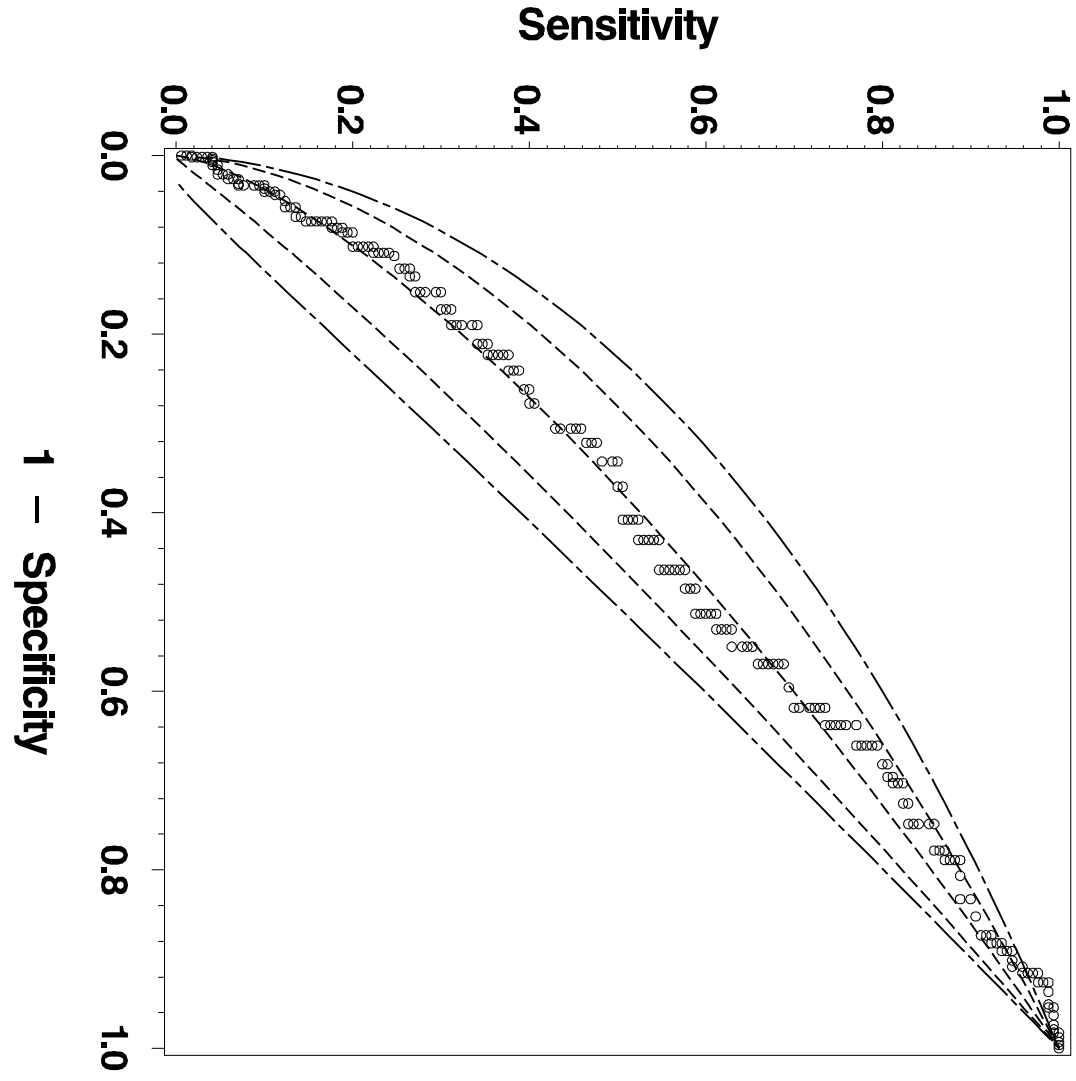
Figure 3: Empirical ROC points (open circles), smooth ROC curve (solid line) and 95% pointwise confidence limits using the partial likelihood (narrower dotted lines) and marginal model (wider dotted lines)

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

Figure 5: Age adjusted ROC curves, dotted line is for age$\leq$ 58 and the solid line is for age$>$ 58.


Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
x
0.0
0.2
0.4
0.6
0.8
1.0




Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 − Specificity
0.0
0.2
0.4
0.6
0.8
1.0