Lehmann Family of ROC Curves

Mithat Gönen, Ph. D.

gonenm@mskcc.org

Memorial Sloan-Kettering Cancer Center

Glenn Heller, Ph. D.

Memorial Sloan-Kettering Cancer Center

hellerg@mskcc.org

Contact Author and Reprint Requests:

Mithat Gönen

Department of Epidemiology and Biostatistics

Memorial Sloan-Kettering Cancer Center

307 East 63 Street, New York, NY 10021

Word Count: 3327

Abstract

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 a 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 and the sensitivity at a given level of specificity) 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, clustered data, 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 to the case of covariate adjustment (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 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. Some examples from regression methods for ROC curves are Zou and Hall (2000), Alonzo and Pepe (2002) and Janes and Pepe (2005). 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. An example of a binormal regression model is presented by Hunink et al (1990).

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, has not yet been widely adopted, possibly due to computational challenges (Pepe, 2000 and Alonzo and Pepe, 2002). Another disadvantage for direct modeling is the lack of goodness of fit diagnostics, although recent work by Cai and Zheng (2007) may remedy this.

In summary, we find that modeling the marker values has many practical advantages, including ease of implementation and the availability of model checking methods. In this article, we present a semiparametric model for the marker values based on the proportional hazards specification. 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 an 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. We present in the appendix `SAS`, `R` and `STATA` code to obtain the ROC curves and diagnostics corresponding to the Lehmann family.

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 θ. 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 θ^{-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. As shown by Le (1997), 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 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

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

the Lehmann specification in (1) may be rewritten as

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

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 the proportional hazards model provides an extensive body of literature and software for the estimation and inference of the odds parameter θ.

Proportional hazards 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 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 the estimated variance of $\hat{\theta}$, 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

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

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 estimation of the survival functions. In addition, there are several methods developed and implemented for model diagnostics (Lin et al., 1993; 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 5.

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 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 metastatases all on the same scan leading to clustered data. It is possible to obtain ROC curves within the Lehmann family to take into account the clustering. Let $k = 1, \ldots, K_i$ index the observations on the i^{th} patient and consider the following model

$$h(V|D_{ki}) = \tilde{h}_0(v) \exp\{\beta D_{ki}\} \tag{14}$$

which gives rise to the same ROC curve as before

$$y(x) = x^{\theta} \tag{15}$$

where $\theta = e^{\beta}$. Assuming the Lehmann specification holds, the estimate of β obtained from the partial likelihood score function are consistent, even in presence of clustering. Estimation of standard error requires the use of a robust covariance estimate to account for the intracluster dependence as in Lee, Wei, and Amato (1992). The variance of $\hat{\beta}$ can be consistently estimated by $\hat{a}^{-1}\hat{w}\hat{a}^{-1}$, where $\hat{a}$ is the negative second derivative from the partial likelihood and $\hat{w}$ is the estimated covariance from the partial likelihood score function. Due to its form, this estimator of the variance is sometimes called a sandwich estimator. These estimates are routinely produced with proportional hazards regression software (see Appendix).

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 Cox model

will provide covariate-adjusted ROC curves under clustering:

$$h(V|D,U) = \tilde{h}(v)\exp\{\beta_1 D + \beta_2 U + \beta_3 DU\}. \tag{16}$$

as long as the standard errors are obtained through the sandwich estimator.

3.3 Comparing the ROC Curves of Several Markers

The comparison of two markers is an important application of ROC curves. 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 total PSA or free PSA 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.

From our regression standpoint, comparison of two markers is a special case of covariate adjustment. Define U as an indicator variable for marker A vs marker B:

$$U = \begin{cases} 1 & \text{if marker A} \\ 0 & \text{if marker B} \end{cases} \tag{17}$$

and consider the model (10):

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

The two markers have identical ROC curves when $\beta_3 = 0$ so a test for the significance of this coefficient serves as a comparison for the two markers.

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. The robust estimation of standard error for clustered data, discussed in the previous section, can be applied to this situation. A Wald test for the equality of two ROC curves in a paired design can constructed by $\hat{\beta}_3/\sqrt{\text{Var}(\hat{\beta}_3)}$, where $\hat{\beta}_3$ is the usual partial likelihood estimate and $\text{Var}(\hat{\beta}_3)$ is the sandwich estimate of the standard error that takes into account the paired 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.

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 the sample paths generated from a simulated score process obtained by replacing the local martingale residuals with their observed counting process multiplied by independent standard normal random variables (Lin et al., 1993). The score process is the derivative of the partial loglikelihood function, used to estimate the regression coefficient β in the proportional hazards model. The term process is used to indicate that the derivative is indexed by the marker variable. This resampling procedure is repeated 100 times to generate the 100 sample paths depicted in Figure 4. The simulated paths are generated under the null hypothesis that the proportional hazards model is the correct fit. This figure was produced using SAS code that is provided in the appendix. Since the observed process is typical of the sample paths obtained under the simulated proportional hazards model, there is no evidence against proportional hazards between normal and benign patients, validating the assumptions underlying the ROC curves in Figure 3. For comparison we also provide a binormal fit to the ROC curve (dotted line in Figure 5).

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 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 0.021 (ignoring clustering) or 0.035 (adjusted for clustering). The corresponding confidence limits are (0.546, 0.630) and (0.518, 0.658) confirming the difficulty of distinguishing between the 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 and 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 1. The coefficient estimate $\hat{\beta}_3$ is positive indicating a decrease in accuracy with increasing age, however it is not significantly different from 0 when clustering is taken into account. Figure 6 displays the AUC as a function of age, where the decreasing discriminatory power of the percent difference in signal intensities in older patients can be visualized.

The increase in standard errors and the widths of the confidence intervals in Table 1 when adjusted for clustering is notable. This underscores the importance of adjusting for clustering, especially in data sets like this when the marker values exhibit substantial intra-patient correlation.

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 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 in ROC analysis is the binormal model, which assumes that the marker values are normally distributed, possibly after being subjected to a monotone transformation. The binormal model specifies the ROC curve with two parameters. In contrast, the Lehmann assumption is equivalent to assuming the existence of a monotone transformation producing marker values with an extreme value distribution, the logarithm of a Weibull distribution (Kalbfleisch, 1978), but does not require that the transformation is specified or even estimated. The Lehmann family of ROC curves is indexed by a single parameter. Thus, the binormal model is less robust in its normal parametric specification, but is more flexible in the number of parameters used to specify this distribution.

A natural question of pratical importance is choosing between binormal and Lehmann models. Goodness-of-fit tests using the score process is helpful in deciding whether a Lehman model is a good fit. Standard methods, such as Q-Q plots or the Shapiro-Wilk test, for checking the normality of marker values separately in gold

standard negative and positive groups can be used for deciding whether the binormal model fits the data reasonably well. More sophisticaed methods are available for the binormal model as well (Cai and Zheng, 2007). If only one of the two models is a good fit then it would be the prudent choice. If both models provide good fits then the choice may depend on non-statistical concerns, such as interpretation and communication of the results to the clinicians or software preferences of the statistician analyzing the data.

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 survival data, and all mainstream statistical packages provide estimates, inferences, and model diagnostics for this model. This machinery can be readily used with no additional effort for 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.

Cai T., Zheng Y. (2007). Model Checking for ROC Regression Analysis. *Biometrics* **63**, 152-163.

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.

Hunink MGM, Richardson DK, Doubilet PM, Begg CB (1990) Testing for Fetal Pul-

monary Maturity: ROC Analysis Involving Covariates, Verification Bias, and Combination Testing. Medical Decision Making, Aug 1990; vol. 10: pp. 201 - 211.

Kalbfleisch, J. D. (1978). Likelihood Methods and Nonparametric Tests *Journal of the American Statistical Association*, **73**, 167–170.

Le CT (1997). Evaluation of Confounding Effects in ROC Studies. *Biometrics*, Vol. 53, No. 3 (Sep., 1997), pp. 998-1007

Lehmann, E. L. (1953). The power of rank tests. *Annals of Mathematical Statistics* **24**, 23–43.

Lin, D. Y., Wei, L. J., Ying, Z. (1993). Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals. *Biometrika* **80**, 557–572.

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.

Appendix

In this appendix we provide simple statements to fit the univariate or marginal regression models described in Sections 2, 3 and 4. The code here will produce partial likelihood estimates of the regression parameter β and their attendant variance estimates. The desired ROC curve can then be produced using the techniques explained in Section 2. In addition, code used to assess the proportional hazards assumption, as described in Section 4, is presented. We assume that the working data set is called `ph` with the marker called `signal` and gold standard called `group`. In the case of clustered data, the clustering variable is called `name`.

5.1 SAS Code

The following code fits the proportional hazards model. The option `descending` is helpful when higher values of the gold standard variable is indicative of disease, which is the most common coding in pratice.

```
proc phreg data=ph;
class group / descending;
model signal=group ;
assess ph / npath=100; run;
```

The marginal model of Section 3.2 can be fit by the following modification:

```
proc phreg data=ph covsandwich(aggregate);
class group / descending;
model signal=group ;
id name;
```

```
run;
```

In this code `id` statement identifies the clustering variable and the `covsandwich` option requires the use of variance estimates based on the sandwich method.

The graphical method of model checking described in Section 4 is performed using the `assess` statement as shown below. The option `npath` specifies the number of sample paths to be used. It is not necessary to fit the marginal to perform the model check; `assess` statememt can be used with the basic model as well.

```
proc phreg data=ph covsandwich(aggregate);
class group / descending;
model signal=group ;
assess ph / npath=100;
id name;
run;
```

5.2 R Code

Proportional hazards fitting in R is primarily achieved by the function `coxph` library `survival`. The following command fits the basic model without clustering

```
coxph(Surv(signal) ~ factor(group),data=ph)
```

and the following fits the marginal model using the robust variance estimate

```
coxph(Surv(signal)~factor(group)+cluster(name),data=ph)
```

The graphical method of model checking is not available in this library, but an alternate method due to Grambsch and Therneau (1994), which produces a p-value for the null hypothesis that the data follow proportional hazards is available:

```
cox.zph(coxph(Surv(signal)~factor(group),data=ph))
```

5.3 STATA Code

In STATA, one first needs to set the stage for propotional hazards analysis by `stset signal` before fitting the basic model using `stcox i.group` or the marginal with the robust estimate using `stcox i.group, robust`. Graphical model checking is not available but the test that is performed by `cox.zph` in `R` can be conducted by the following sequence of statements:

```
stcox group, schoenfeld(temp)
estat phtest
```

Table 1: 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 parameter θ in equation (1) 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: Lehmann (solid line) and binormal (dotted line) ROC curves overlaid

Figure 6: AUC as a function of age.