Relative Risk Forests for Exercise Heart Rate Recovery as a Predictor of Mortality

Hemant ISHWARAN, Eugene H. BLACKSTONE, Claire E. POTHIER, and Michael S. LAUER

Recent studies have confirmed heart rate fall after treadmill exercise testing, or *heart rate recovery*, as a powerful predictor of mortality from heart disease. Heart rate recovery depends on central reactivation of vagal tone and decreased vagal activity is a risk factor for death. If heart rate recovery is defined as the fall in heart rate after 1 minute following peak exercise, then a heart rate recovery value of 12 beats per minute (bpm) or lower has been shown to be a good prognostic threshold for identifying patients at high risk. Although this finding establishes a simple, useful relationship between heart recovery and mortality, a working understanding of how heart rate recovery interacts with other characteristics of a patient in determining risk of death is still largely unexplored. Such knowledge, addressed in this article, could improve the prognostic value of the exercise test. Our analysis is based on over 23,000 patients who underwent exercise testing. A rich assortment of data was collected on these patients, including clinical and physiological information, heart rate recovery, and other exercise test performance measures. Our approach was to grow *relative risk forests*, a novel method that combines random forest methodology with survival trees grown using Poisson likelihoods. Our analysis reveals a complex relationship between peak heart rate, age, level of fitness, heart rate recovery, and risk of death.

KEY WORDS: Cox regression; MARS; Proportional hazards; Random forests; Relative risk trees; Stochastic variable selection.

1. INTRODUCTION

Exercise stress testing is commonly used to assess patients with known or suspected coronary artery disease (Lauer 2001). During exercise testing, exercise capacity, heart rate changes, and changes on the electrocardiogram are recorded. Although exercise testing has classically been considered as a diagnostic test to identify patients likely to have important coronary artery disease, recent work has focused on its powerful prognostic value (Gibbons et al. 1997, 2002a,b; Lauer 2001; Williams, Fihn, and Gibbons 2001).

During exercise, the heart rate rises due to regulatory effects of the autonomic nervous system. Autonomic nervous system function is determined by the balance of activity of its sympathetic and parasympathetic components: Sympathetic function results in increased heart rate and blood pressure; parasympathetic function results in their decrease. Heart rate rise during exercise is largely due to rapid withdrawal of parasympathetic tone (also known as vagal tone) as well as increased sympathetic tone (Hammond and Froelicher 1985; Arai et al. 1989). Failure of heart rate to rise appropriately during exercise, known as chronotropic incompetence (Lauer, Okin, Larson, Evans and Levy 1996; Lauer et al. 1999), is a predictor of all-cause mortality and coronary heart disease events (Lauer et al. 1996; 1999; Lauer 2001).

Recently we have focused on fall in heart rate immediately after exercise, or heart rate recovery (Imai et al. 1994), as a potential predictor of morality (Cole, Blackstone, Pashkow, Snader and Lauer 1999; and Cole, Foody, Blackstone, and Lauer 2000; Nishime, Cole, Blackstone, Pashkow, and Lauer 2000; Shetler et al. 2001). Decrease in heart rate during the first minute after exercise is largely a function of reactivation of parasympathetic function (Imai et al. 1994). Depressed parasympathetic function is associated with increased risk of death in a wide spectrum of patients (Schwartz, La Rovere, and Vanoli 1992; La Rovere, Bigger, Marcus, Mortara and Schwartz 1998; La Rovere et al. 2001). Thus we hypothesize that attenuated heart rate recovery immediately after exercise is associated with increased risk of death (Cole et al. 1999).

1.1 Heart Rate Recovery as a Predictor of Death

The hypothesis that heart rate recovery is an independent predictor of mortality has been tested and validated in a number of cohorts. Our initial effort involved over 2,000 patients (Cole et al. 1999) for which heart rate recovery was defined as the heart rate at peak exercise minus the heart rate measured 1 minute later. We defined a cut-off value for an abnormal heart rate recovery based on maximization of a log-rank statistic. Based on this, patients with heart rate recoveries less than or equal to 12 beats per minute (bpm) were found to be at substantially increased risk of death compared to those who had a normal heart rate recovery (greater than 12 bpm).

Figure 1(a) provides confirmation of these findings in context to the dataset analyzed here. The figure records the score test statistic from a Cox proportional hazards model using different heart rate recovery threshold values. For each threshold value, the Cox model included a 0/1 dichotomized covariate for heart rate recovery (equal to 1 if heart rate recovery was less than or equal to the threshold value, otherwise 0). Included in the Cox model were additive terms for additional covariates identified as being important (see Remark 1 for details). The score test statistic has a flat maximum between 8 and 12 bpm and then drops off rapidly. The maximum seen at 12 bpm agrees with the findings found in Cole et al. (1999). Our later analysis also confirms 12 bpm as an optimal threshold for identifying patients at high risk.

1.2 Patient Cohort

Not only is heart rate recovery an independent predictor of mortality, it is also predictive of mortality after adjusting for several factors (Snader et al. 1997; Cole et al. 1999, 2000; Diaz,

Hemant Ishwaran is Associate Staff (E-mail: *ishwaran@bio.ri.ccf.org*), Department of Biostatistics and Epidemiology Wb4, Eugene Blackstone is Full Staff (E-mail: *blackse@ccf.org*), Department of Thoracic and Cardiac Surgery F25, Claire Pothier is Clinical Exercise Physiologist and Research Assistant, Department of Cardiovascular Medicine (E-mail: *pothiec@ccf.org*), and Michael Lauer is Full Staff (E-mail: *lauerm@ccf.org*), Department of Cardiovascular Medicine F25, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland OH 44195. The authors thank the editor, Tom Louis, the two referees, and the associate editor for their time and helpful comments. This work was supported by grant HL-66004-01 from the National Heart, Lung, and Blood Institute.

Journal of the American Statistical Association
September 2004, Vol. 99, No. 467, Applications and Case Studies
DOI 10.1198/016214504000000638

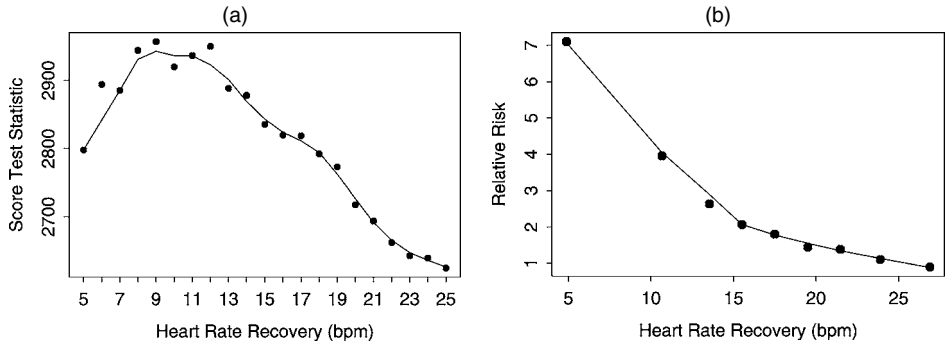

Figure 1. Score Text Statistic and Relative Risk. (a) Score test statistic as a function of heart rate recovery in a Cox regression with additive terms (superimposed line is loess estimate). Heart rate recovery is dichotomized. Note the large score value at 12 bpm. (b) Relative risk versus heart rate recovery from a Cox regression in which heart rate recovery was treated as a factor with 10 levels based on its deciles (no other covariates were included in the model). Relative risks computed at each decile of heart rate recovery with last decile used as the baseline.

Brunken, Blackstone, Snader and Lauer 2001; Watanabe et al. 2001). Nonetheless, there are a number of important unanswered questions (Shetler et al. 2001). How does heart rate recovery interact with other clinical and physiological variables? What is its prognostic behavior when considered as a continuous variable? Standard methods provide some insight, but not detailed understanding. Consider, for example, Figure 1(b). Relative risk values are shown from a Cox regression analysis of heart rate recovery treated as a factor with 10 levels based on its deciles. The figure shows that relative risk decreases as heart rate recovery increases. However, it is not easy to extend such an analysis to adjust for patient clinical variables and potentially complex interactions. For example, standard methods such as stepwise regression, which are often used to select models within a Cox regression framework, are often not only unreliable, but may simply be too adhoc to discover the complex interactions.

To address these issues, we analyze a large database collected from 23,701 patients referred to the Cleveland Clinic between 1990 and 1999 for symptom-limited exercise testing. Each patient underwent an upright cool-down period for the first 2 minutes after recovery. All patients underwent a structured chart review and interview prior to testing (Snader et al. 1997; Cole et al. 1999; Lauer et al. 1999). Detailed data regarding reason for testing, symptoms, cardiac risk factors, other medical diagnoses, prior cardiac and noncardiac procedures, medications, resting electrocardiogram, resting heart rate, and blood pressure were recorded prospectively prior to testing. During each stage of exercise, and during the first 5 minutes of recovery, data were recorded regarding heart rate, blood pressure, ST-segment changes, symptoms, and arrhythmias. Mean follow-up among survivors was 5.7 years (range .75 to 10.1 years) during which 1,617 patients died.

Remark 1. Over 90 variables were available in the database for analysis. We choose 34 of these as follows. A Cox regression model was fit for each candidate variable which included additionally an additive effect for heart rate recovery and an interaction effect between the two variables. Only variables whose interaction with heart rate recovery were significant at a .05 level (two-sided test) were chosen. The purpose of this was to select only covariates with a known involvement with heart rate recovery. This fits with our general plan to understand how heart rate recovery interacts with other covariates.

1.3 Organization of Paper

Our analysis is based on a new method we refer to as *relative risk forests*. This method combines the use of relative risk trees (LeBlanc and Crowley 1992) with random forest methodology (Breiman 2001) as a way to estimate reliably relative risk values. Relative risk forests circumvent some key difficulties in implementing relative risk trees. For example, the problems of when to stop growing a tree and how to optimally prune a tree are avoided because we grow trees to full size. Risk values, or what we refer to as ensemble risk values, are computed by aggregating across different risk trees. This reduces variance and avoids the instability of working with a single relative risk tree. On the other hand, relative risk forests take advantage of key features of a risk tree. They exploit their rich structure and retain the same simple relative risk interpretation. An important feature of random forests is that they can be applied with large numbers of variables (Amit and Geman 1997) and are especially useful in settings like ours where variables are heavily correlated. Sections 2 and 3 lay out the details of the method. Most of the analysis is given in Section 4. The article concludes with a summary and discussion. Some alternative methods for flexible inference in Cox regression models, such as adaptive regression splines used by LeBlanc and Crowley (1999), will be reviewed there.

2. POISSON TREE LIKELIHOODS

Random forests were grown using relative risk trees. A relative risk tree is a nonparametric technique introduced by LeBlanc and Crowley (1992) for estimating risk in proportional hazard survival analysis problems. A key feature in this approach is that a relative risk tree can be grown using the classification and regression tree (CART) methodology of Breiman, Friedman, Olshen, and Stone (1984). This rests on an equivalence between survival tree and Poisson tree likelihoods. In this way, splits, and hence trees, are formed using well-known principles of generalized linear models harnessed within the power of CART.

The idea behind relative risk trees is as follows. Let t_i denote the survival time for patient i and $\mathbf{x}_i$ the patients' covariate information. The data are $\{(t_i, \delta_i, \mathbf{x}_i) : i = 1, \ldots, n\}$, where

δ_i indicates right-censoring information: $\delta_i = 1$ if t_i is an observed time of death; otherwise $\delta_i = 0$ if i is censored. We assume that censoring and survival times are independent. The full likelihood L_F for a tree T is

$$L_F = \prod_{h \in T_*} \prod_{i \in I_h} \lambda_h(t_i)^{\delta_i} \exp\bigl(-\Lambda_h(t_i)\bigr),$$

where T_* are the terminal nodes for T and I_h are the observation labels $\{i : \mathbf{x}_i \in S_h\}$ for observations in the region S_h corresponding to node h. Here $\lambda_h(t)$ and $\Lambda_h(t)$ are the hazard and cumulative hazard functions for h. Under the Cox proportional hazards assumption (Cox 1972),

$$\lambda_h(t) = \lambda_0(t) \exp(\beta_h),$$

where $\lambda_0(t)$ is the baseline hazard function and β_h is the regression constant for node h. The likelihood L_R for a relative risk tree is (LeBlanc and Crowley 1992)

$$L_R = \prod_{h \in T_*} \prod_{i \in I_h} \bigl(\lambda_0(t_i)\theta_h\bigr)^{\delta_i} \exp\bigl(-\Lambda_0(t_i)\theta_h\bigr),$$

where $\theta_h = \exp(\beta_h)$ and $\Lambda_0(t)$ is the baseline cumulative hazard function.

Notice, however, that if Λ_0 (and λ_0) is known, then L_R is equivalent up to a multiplicative constant to the likelihood

$$L_P = \prod_{h \in T_*} \prod_{i \in I_h} \bigl(\Lambda_0(t_i)\theta_h\bigr)^{\delta_i} \exp\bigl(-\Lambda_0(t_i)\theta_h\bigr),$$

which is the likelihood from a tree derived under a Poisson model with response δ_i and mean $\mu_i = \Lambda_0(t_i)\theta_h$ for $i \in I_h$.

2.1 Tree Splits

The equivalence between survival tree and Poisson tree likelihoods shows that if $\Lambda_0(t)$ is known then the relative risk tree can be grown using a Poisson model. LeBlanc and Crowley (1992) proposed using the Nelson–Aalen (1978) estimator for $\Lambda_0(t)$. This is the Breslow (1972) estimator in a Cox model without covariates. LeBlanc and Crowley (1992) referred to this as their "one-step" estimator, which we shall denote by $\widehat{\Lambda}_0^1(t)$. In the following section we will explain the meaning behind this terminology. Using the one-step estimator $\widehat{\Lambda}_0^1(t)$, LeBlanc and Crowley (1992) proposed growing a relative risk tree by growing a Poisson tree with likelihood

$$\widehat{L}_P = \prod_{h \in T_*} \prod_{i \in I_h} \bigl(\widehat{\Lambda}_0^1(t_i)\theta_h\bigr)^{\delta_i} \exp\bigl(-\widehat{\Lambda}_0^1(t_i)\theta_h\bigr).$$

In growing the tree, splits are made using the deviance residual

$$d_i = 2\left[\delta_i \log\left(\frac{\delta_i}{\widehat{\Lambda}_0^1(t_i)\widehat{\theta}_h^1}\right) - \bigl(\delta_i - \widehat{\Lambda}_0^1(t_i)\widehat{\theta}_h^1\bigr)\right],$$

where $\widehat{\theta}_h^1$ is the estimate for a node h under the proposed split (note that $0 \times \log 0 = 0$). This is simply the maximum likelihood estimate,

$$\widehat{\theta}_h^1 = \frac{\sum_{i \in I_h} \delta_i}{\sum_{i \in I_h} \widehat{\Lambda}_0^1(t_i)}.$$

The deviance for a node h is $D(h) = \sum_{i \in I_h} d_i$, which is the log-likelihood ratio test statistic when the null is the saturated model at h. Improvement in the deviance for split s at a node h into left and right daughter nodes $l_s(h)$ and $r_s(h)$ is

$$\Delta D(s, h) = D(h) - \bigl[D\bigl(l_s(h)\bigr) + D\bigl(r_s(h)\bigr)\bigr].$$

The tree is split by the variable at s whose split s leads to the smallest value $\Delta D(s, h)$. Typically a tree is grown until some criterion is met, then pruned. In our random forests implementation, we grow a tree to its full size without pruning (the only restriction is that a node contain no fewer than N observations; where N is typically some small number). A nice feature of a random forests approach is it avoids two important problems with trees: (a) when to stop growing a tree and (b) how to optimally prune a tree.

2.2 Relative Risk Values

By fixing $\Lambda_0(t)$ at $\widehat{\Lambda}_0^1(t)$, a relative risk tree can be grown by recursive partitioning applied to a Poisson model. The relevant likelihood $\widehat{L}_P$ corresponds to the likelihood from a Poisson model with a log-link, where each i has an offset value of $\log(\widehat{\Lambda}_0^1(t_i))$. That is, if $i \in I_h$,

$$\log(\mu_i) = \log\bigl(\widehat{\Lambda}_0^1(t_i)\bigr) + \beta_h,$$

where $\log(\widehat{\Lambda}_0^1(t_i))$ is an offset value and β_h is an unknown parameter (what we later define as the log relative risk).

The use of the offset can be applied in available software to build relative risk trees and to derive relative risk values for patients. Our analysis was implemented using the RPART algorithm for Splus devised by Therneau and Atkinson (1997), but can be implemented in any recursive partitioning software that fits Poisson trees.

The estimate $\widehat{\theta}_h^1$ computed using this method should be interpreted with caution though, because it is only an approximation. This is because $\widehat{\theta}_h^1$ is derived by estimating $\Lambda_0(t)$ by $\widehat{\Lambda}_0^1(t)$, but $\widehat{\Lambda}_0^1(t)$ is only appropriate in a model without covariates. A more appropriate estimate for $\Lambda_0(t)$ should reflect the values for θ_h in the relative risk tree L_R; that is, it should take into account the underlying covariates used in building the tree. LeBlanc and Crowley (1992) discussed this issue. They suggested an iterative procedure for more accurate estimation of θ_h that is applied while growing the tree. However, implementing this method is computationally impractical for a large dataset. Instead, we apply this iterative procedure *after* the tree is grown.

The method works as follows. Given the one-step estimator $\widehat{\theta}_h^1$, an updated Breslow estimator $\widehat{\Lambda}_0^2(t)$ for the cumulative hazard is computed from this:

$$\widehat{\Lambda}_0^2(t) = \sum_{\{i\,:\,t_i \le t\}} \frac{\delta_i}{\sum_{h \in T^*} n_h(t_i)\widehat{\theta}_h^1},$$

where $n_h(t_i) = \#\{j : j \in I_h, t_j \ge t_i\}$ is the number of individuals in node h who are at risk at time t_i. Then, given $\widehat{\Lambda}_0^2(t)$, an updated estimator $\widehat{\theta}_h^2$ is computed,

$$\widehat{\theta}_h^2 = \frac{\sum_{i \in I_h} \delta_i}{\sum_{i \in I_h} \widehat{\Lambda}_0^2(t_i)}.$$

These two steps are repeated iteratively giving a sequence of estimators $(\widehat{\Lambda}_0^j(t_i), \widehat{\theta}_h^j)$ for $j = 2, \ldots, J$. A similar idea was used earlier by Aitkin and Clayton (1980) and Clayton and Cuzick

(1985) as a method for fitting Cox regression models using generalized linear models.

The number of steps J used in the iterative scheme can be fixed in advance, or a convergence criterion can be used. We stopped when the log relative risk $\widehat{\beta}_h^j = \log(\widehat{\theta}_h^j)$ stabilized. Specifically, the iterative scheme terminated at the first $j \geq 2$ if

$$\sum_{h \in T^*} \frac{n_h}{n} \left| \widehat{\beta}_h^j - \widehat{\beta}_h^{j-1} \right| \leq \epsilon, \tag{1}$$

where n_h equals the number of individuals in a node h. We used a value of $\epsilon = .05$. This reflects the internal accuracy of our log relative risk values: They are accurate up to 5%. We found that three steps were usually enough to satisfy our stopping criterion (1). At termination, the final estimator for θ_h and $\Lambda_0(t)$ is $\widehat{\theta}_h^j$ and $\widehat{\Lambda}_0^j(t)$ for some $j \geq 2$. For notational convenience we drop the use of a superscript and write these final estimators as $\widehat{\theta}_h$ and $\widehat{\Lambda}_0(t)$, respectively.

To summarize, the method to grow a relative risk tree is as follows:

1. Compute the Nelson–Aalen estimator $\widehat{\Lambda}_0^1(t)$.
2. Grow a tree using recursive partitioning applied to a Poisson model with a log-link where δ_i is the response, $\log(\widehat{\Lambda}_0^1(t_i))$ is an offset value, and $\mathbf{x}_i$ are the covariates. Note that by default RPART uses $\Delta D(s, h)$ as a splitting rule.
3. The fitted tree yields one-step estimates $\widehat{\theta}_h^1$ for each terminal node h. Use the iterative scheme to compute a new value $\widehat{\theta}_h$ using the stopping criterion (1). Note that the iterative scheme provides an updated estimate $\widehat{\Lambda}_0(t)$ which can be useful for inference (see Sec. 3.2).
4. To determine the relative risk for patient i, drop their covariate $\mathbf{x}_i$ down the tree. If $\mathbf{x}_i$ lands in S_h, define

$$R_i = \frac{\widehat{\lambda}_0(t_i)\widehat{\theta}_h}{\widehat{\lambda}_0(t_i)} = \widehat{\theta}_h.$$

This is the relative risk for i when compared to the mean unit in the study. We call R_i the relative risk for i, and we call $\widehat{\beta}_h$ the log relative risk for i. Note that the mean unit is automatically selected by the procedure.

3. RELATIVE RISK FORESTS

A relative risk tree produces a relative risk value R_i for each patient i. More generally, for each covariate $\mathbf{x}$ in the covariate space $\mathcal{X}$ there is some function $H : \mathcal{X} \to \Re^+$ for the tree that gives the relative risk value for $\mathbf{x}$ (simply drop $\mathbf{x}$ down the tree). We grow many relative risk trees from our data by adapting the random forests approach of Breiman (2001). This produces different H functions which we aggregate to form an overall estimate for the risk of a patient. The advantage in growing many trees and using an aggregated risk estimate is that it is a way to reduce variance (Breiman 2001). Prediction based on individual trees can be poor because trees are considered to be unstable procedures (Breiman 1996). Growing many trees and aggregating is a way to fix this. It also leads to classifiers and predictors that are drawn from a richer class of models.

The idea is to grow a tree by injecting two types of randomness into the process. To grow a tree:

1. Bootstrap the data. That is, select n values with replacement from $\{1, \ldots, n\}$. The bootstrapped data are $\{(t_{i_j}, \delta_{i_j}, \mathbf{x}_{i_j}) : j = 1, \ldots, n\}$.
2. In addition to heart rate recovery, randomly select F additional covariates from the full input set. This gives a total of $F + 1$ covariates.
3. Grow a relative risk tree to full size using the bootstrapped data points and the covariates selected in step 2. To ensure that nodes are not too small, only allow nodes with at least N observations (we took $N = 20$).

Steps 1 and 2 introduce randomness. They produce a set of instructions which can be encoded as a random vector, Θ, which is used to grow the relative risk tree. The estimated value for the relative risk for a covariate $\mathbf{x}$ is $H(\mathbf{x}, \Theta)$. The random forest estimate for risk is obtained by implementing the procedure independently B times to get B iid values $\Theta_1, \ldots, \Theta_B$ and B corresponding trees. The relative risk forest is $\{H(\mathbf{x}, \Theta_b) : b = 1, \ldots, B\}$. The *ensemble relative risk* for a covariate $\mathbf{x}$ is

$$R_e(\mathbf{x}) = \frac{1}{B} \sum_{b=1}^{B} H(\mathbf{x}, \Theta_b).$$

In a similar way one can define many other kinds of ensemble values, such as ensemble log relative risks, which we denote by $\ell R_e(\mathbf{x})$. While the ensemble relative risk is preferred for its simple interpretation, the log relative risk is a useful measurement for prediction purposes. For example, in Section 4 we discuss a nonparametric regression of $\ell R_e(\mathbf{x})$ on covariates as a way for predicting log risk.

By construction, R_e will likely have some functional relationship to heart rate recovery. This is because step 2 ensures that heart rate recovery is always one of the variables considered for a split when growing a tree, and thus is likely to be involved in the risk values produced by a tree. This can be thought of as a form of *weak supervision*. Because one of the primary goals of the analysis is to study how relative risk varies with respect to heart rate recovery, it is important to ensure that each tree can potentially be grown using its value. However, weak supervision does not guarantee that heart recovery will always be used for growing a tree, nor does it guarantee that ensemble values will be directly related to it. Each of our other 34 covariates are known to be predictors for risk in combination with heart rate recovery and offer some competition. If heart rate recovery is a key variable, then trees will naturally split on it. On the other hand, trees will likely split on other variables, thus leading to ensemble values that can depend on heart rate recovery, but also possibly other patient characteristics if they are significant in predicting mortality.

3.1 Selecting the Number of Features, F

To ensure that random forests have good prediction properties, it is important to check that the correct amount of randomization has been introduced for Θ. This means that we need to determine an appropriate number of randomly selected covariates, F, to be used in step 2 of the procedure. If we select too few covariates, the trees might be too sparse, and our ensemble estimator R_e will have suboptimal properties. Too many covariates, and the trees will be highly correlated, which can

also degrade performance. As discussed in Breiman (2001), one method for assessing the accuracy of a forest is through its generalization error. This value depends on the correlation between trees and the strength of a tree. For regression trees (Breiman 2001, thm. 11.2), if Y is the response, then the tree correlation is measured by the correlation between the residuals $Y - H(\mathbf{X}, \Theta)$ and $Y - H(\mathbf{X}, \Theta')$, where Θ and Θ' are independent, whereas the strength of a tree is measured by its generalization error $E_\Theta E_{\mathbf{X},Y}(Y - H(\mathbf{X}, \Theta))^2$ (the smaller this value, the stronger a tree). The generalization error for a forest involves a tradeoff between these two values. As F increases, the strength of a tree increases, which contributes to a lower forest generalization error; at the same time, however, the correlation between residuals increases, which increases error.

Our setup is, of course, different. Not only are we working with a Poisson regression model, but more important we do not have an *observed* Y response value. We do not know what the relative risk is (our "Y value"), and, in fact, our goal is to estimate its value by R_e. However, given the large size of the dataset, there is a way to honestly estimate the generalizability of R_e. To do so, we split the data randomly into two parts of equal size. Trees and forests are then grown separately and independently for each of the datasets, giving two sets of estimates for $R_e(\mathbf{x})$ for each $\mathbf{x}$ in the combined dataset. By comparing these two sets of ensemble estimates, we can study how generalizable R_e is. Also, by varying the number of randomly selected covariates, we can decide what the right amount of randomization is.

Remark 2. Even with the relatively large number of deaths (1,617), we still needed to be careful to ensure that there was an even distribution of deaths when randomly splitting the data. Therefore, when splitting the data, we used a two-stage procedure by first evenly splitting censored cases and then splitting cases due to death.

Figure 2 is the plot of the ensemble log relative risk, ℓR_e, from forests grown using different choices for F (ranging from 1 to 10). Values for the horizontal and vertical axes are the values for ℓR_e computed from forests grown from the two halves of the randomly split data. Forests were grown using the strategy outlined earlier in this section. In each case, $B = 35$ trees were grown to full size with nodes restricted so they had no fewer than $N = 20$ values. We experimented with larger choices for B, even considering forests with up to $B = 100$ trees, but found that ensemble values stabilized rapidly with usually 20–30 trees being more than enough to get accurate values. Further evidence of this stabilization will be provided shortly.

Embedded in each plot of Figure 2 is the R-squared value from a linear regression obtained by regressing the vertical value of ℓR_e on its horizontal value. The best fit in terms of R-squared (.929) occurs for $F = 1$, which are forests grown using heart rate recovery and one additionally randomly selected covariate. However, while the overall fit measured by R-squared is good, observe how the plot appears more granular compared to the others, and that the range of ℓR_e values is tighter. The next best fit occurs for the value $F = 2$ (R-squared of .926), but notice how the granularity has improved. For $F = 3$, granularity has all but disappeared. This

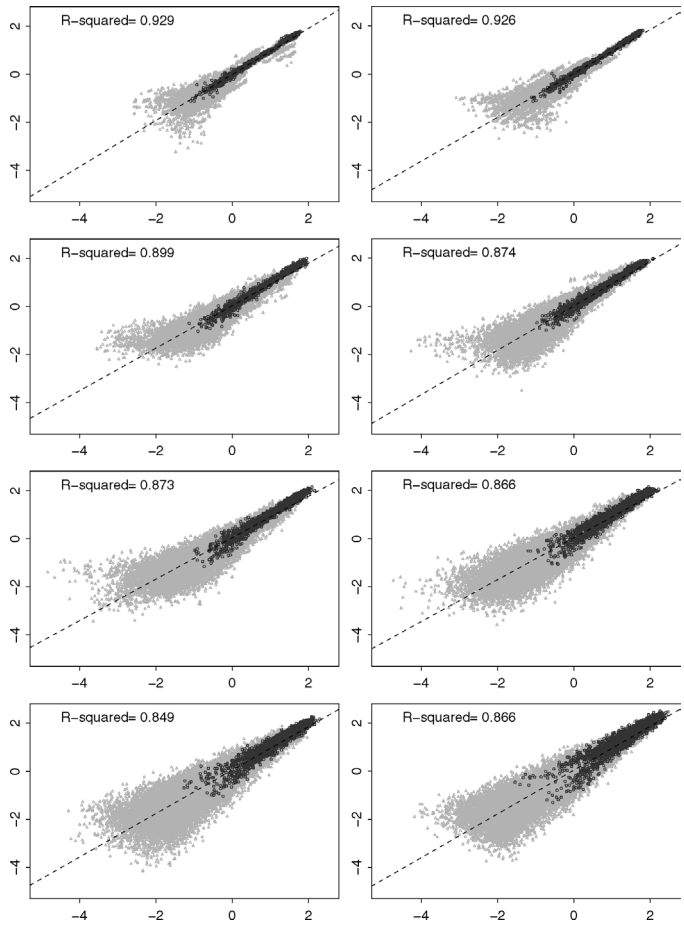

Figure 2. Horizontal and Vertical Axes of Scatterplots Correspond to Ensemble Log Relative Risk Values Estimated From Randomly Split Datasets (triangular symbols in gray indicate censored values; circles in black are deaths; thin dashed line is fitted regression line). Total number of covariates are $F + 1$, where $F = 1, 2, 3, 4, 5, 6, 7, 10$ (left to right, and top to bottom, respectively). Note how ensemble values are more reliably estimated for deaths; however, as F increases the random forests overtrain on these values.

pattern persists for higher values of F; the R-squared stabilizes, and the range of values for ℓR_e no longer increases, suggesting that there is no change in performance as F increases and that the forests have stabilized. However, a careful inspection of the points corresponding to deaths in the plots indicates a different story. Looking carefully at the band of death cases in the center of the plots (circles highlighted in black), we see these values begin to diverge from the regression line as F increases (especially noticeable when $F = 10$). Clearly the forests begin to overtrain when F becomes too large. To avoid this, it seems wise to choose a smaller value for F. The value $F = 5$ appears the most appropriate. At this value we see no evidence of overtraining and R-squared has stabilized. This is the value we chose in our analysis.

Remark 3. Another way to proceed is through an out-of-bootstrap analysis. As before, randomly split the data into two parts of equal size and grow relative risk trees for each of the two datasets. Use each tree to estimate the log relative risk over the combined out-of-bootstrap sample. This way each out-of-bootstrap value has two estimated values; call them ℓR_1 and ℓR_2. Create B trees independently using the same

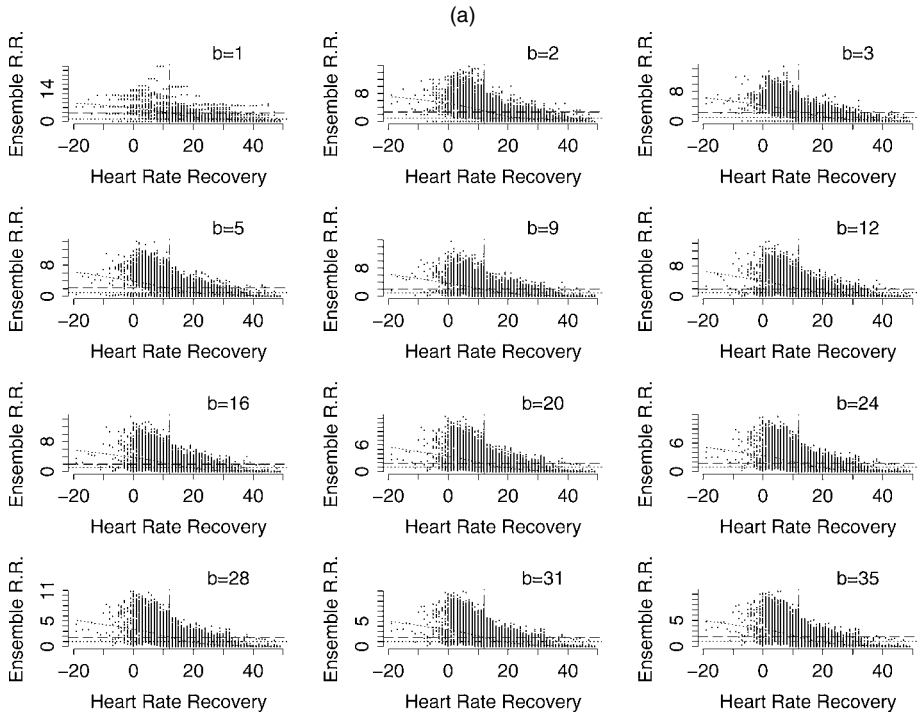

Figure 3(a). Ensemble Relative Risk. Ensemble relative risk versus heart rate recovery as a function of number of trees b (thick dashed horizontal line is the mean relative risk value when heart rate recovery equals 12 bpm; thin dashed horizontal line is relative risk of 1.0; smoothed curve is loess estimate). Plot is based on F = 5 randomly selected features in addition to heart rate recovery.

process. If B is sufficiently large, every data point will have been included in a combined out-of-bootstrap sample several times. Let $\ell R_{e,1}$ and $\ell R_{e,2}$ be the ensemble values obtained from ℓR_1 and ℓR_2. Plot $\ell R_{e,1}$ and $\ell R_{e,2}$ for every data point. Choose F as before on the basis of the scatterplot.

3.2 Heart Rate Recovery versus Ensemble Values

Figures 3 and 4 illustrate forests grown from the complete dataset using $F = 5$. We grew $B = 35$ trees using the strategy outlined earlier in this section. Figure 3(a) shows how quickly R_e stabilizes. Even after growing only $b = 5$ trees, R_e starts to stabilize, and after about 20–30 trees there seems to be no change. Figure 3(b) plots the final estimate for R_e for each patient versus the heart rate recovery value. Figure 4 plots the patient averaged ℓR_e versus heart rate recovery [compare the detail in Figure 4 to Figure 1(b), based on a conventional Cox analysis]. Notice in Figure 4(a) that the log relative risk drops to a low value for heart recovery values of 25 bpm or higher. This has prognostic implications because it identifies a group of patients that can safely be classified as being at low risk.

Figure 4(b), a close-up of Figure 4(a), contains two sets of estimates for ℓR_e. The first (thick line) is a loess curve estimated from the mean values of Figure 4(a). The second estimate is the averaged predicted values from a quadratic linear regression model (thin dashed line). The quadratic model uses our top seven variables as main effects (one of these being heart rate recovery) and all pairwise interactions between these seven variables (see Sec. 4 for details). Both estimates identify 13 bpm as the point where the log relative risk is approximately 0, therefore identifying 12 bpm as the first heart recovery value that puts patients at high risk relative to the baseline value. Note that we prefer the use of the loess estimate and the predicted values from our quadratic model over the patient-averaged mean values for ℓR_e. An analysis based solely on the mean value will be less reliable than a smoothed version due to variability (note the jumps seen at 12 and 13 bpm).

To investigate risk behavior more closely, we considered survival probabilities, comparing patients at 12 bpm to those with higher values. Let $R_{i,b}$ be the relative risk for patient i from a tree b and $\widehat{\Lambda}_{0,b}(t)$ be the estimated baseline cumulative hazard from the same tree. If T_i is survival time, the ensemble proba-

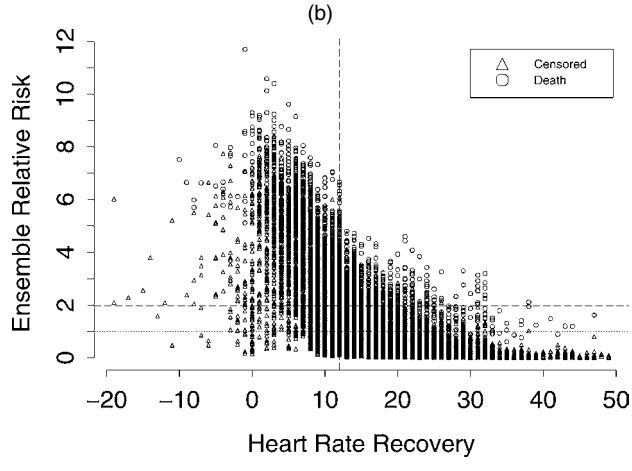

Figure 3(b). Final Estimate for Ensemble Relative Risk [horizontal lines defined as in (a); vertical line is heart rate recovery of 12 bpm].

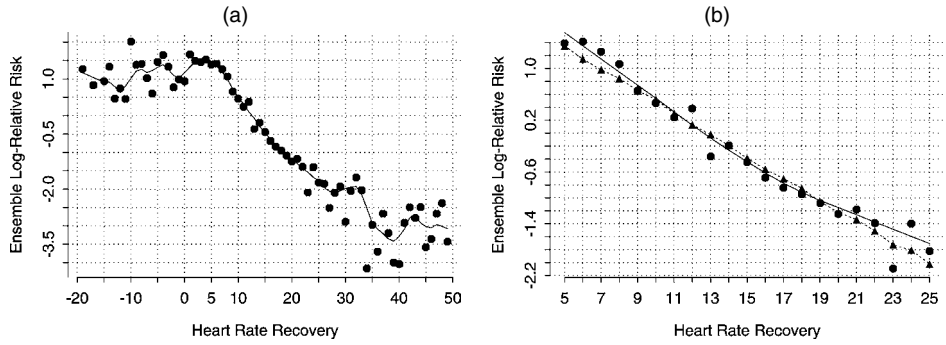

Figure 4. Ensemble Log Relative Risk. Mean ensemble log relative risk versus heart rate recovery obtained from Figure 3(b). Smoothed curve is a loess estimate.Close-up plot. Thin dashed line is predicted value from a quadratic linear regression model. Thick line is loess estimate.

bility that i survives longer than time t is

$$\Pr\{T_i \geq t\} = \frac{1}{B}\sum_{b=1}^{B} \exp\bigl(-R_{i,b}\,\widehat{\Lambda}_{0,b}(t)\bigr).$$

For concreteness, we used a value of $t = 4.5$ years, corresponding to roughly the 60th percentile for the event times of all deaths. Figure 5(a) depicts the density for the probability of surviving 4.5 years stratified by the value of heart rate recovery (12 through 16 bpm). This shows a clear difference in survival behavior for patients whose heart recovery is 12 bpm and, on the upper end, for patients with heart recoveries of 16 bpm. Patients with heart recoveries of 13 and 14 bpm, and to a lesser extent 15 bpm, exhibit roughly similar behavior, and are in between the two extremes. This matches up with the Kaplan–Meier survival curves presented in Figure 5(b). Combined with our analysis of Figure 4(b), this further establishes 12 bpm as an optimal value for putting a patient at high risk.

4. VARIABLE SELECTION AND MULTIVARIATE ADAPTIVE REGRESSION SPLINES

To gain a better understanding of the underlying factors involved in patient mortality, we studied how the ensemble log relative risk depended on covariates other than heart recovery. Because our forests were grown using a wide range of variables, it is quite likely that the estimated log relative risk will have a complex relationship with many of the covariates. For example, reconsider Figure 3(b). There we see patients with heart rate recovery values higher than 12 bpm, but whose relative risk is up to two times higher than the mean unit. We also find patients with heart recovery values near 10 bpm with half the relative risk. It is clear that these kinds of behavior cannot be explained solely on the basis of heart rate recovery. To study how the log relative risk varies with respect to covariate information, we relied on the multivariate adaptive regression splines (MARS) method described in Friedman (1991).

Some thought, however, is first needed to decide what variables should be included in a MARS analysis. Because each of our 34 covariates, including heart rate recovery, is known to be potentially informative in predicting mortality, a naive application of MARS to predict ℓR_e using the full set of covariates will yield very large models that will be hard to translate into simple clinical terms. Instead, our approach was to first reduce the set of covariates to a smaller subset by performing a linear regression with ℓR_e as the response. We used the stochastic variable selection (SVS) method developed in Ishwaran and Rao (2000, 2003). Variables were ranked and their absolute posterior means computed. Posterior means were then translated into a relative importance value from 0 to 100 by dividing by the largest value; the value from the most informative covariate. The analysis was restricted to data for which the heart recovery was in the range of [5, 25] bpm. Restricting values to be no smaller than 5 makes sense as these are most meaningful from a clinical perspective. The 5-bpm category represents roughly the 5th percentile from the data. Also patients with values larger than 25 bpm are at low risk and are of lesser interest (recall Fig. 4).

Not surprisingly, our SVS analysis identified heart rate recovery as the most important variable. Following this, vari-

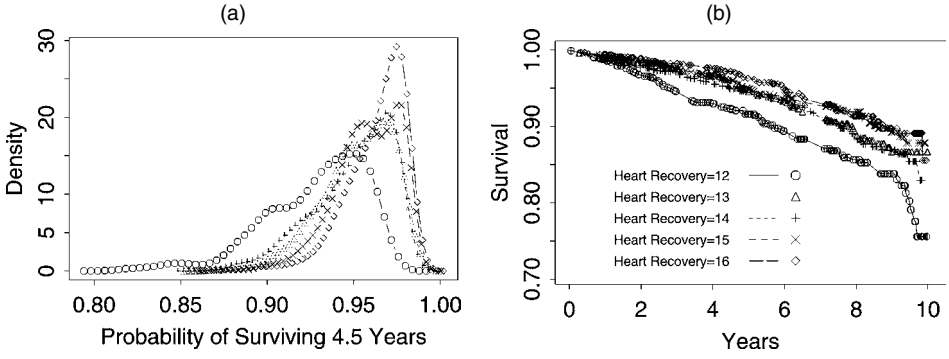

Figure 5. (a) Probability of Surviving 4.5 Years Stratified by Heart Recovery Value (12–16 bpm). Depicted is the density for the value of this probability for all patients. (b) Kaplan–Meier survival curves stratified by heart recovery.

ables with a relative importance of more than 15 were, in order: (2) PEAKHR, peak exercise heart rate; (3) AGE, age of a patient; (4) IMAGE, whether the patient was referred for image testing; (5) FITNESS, overall fitness level of a patient (encoded as 1–5, with 1 being best); (6) PRIORCAD, whether the patient had a history of cardiopulmonary disease; and (7) LOWCRI, whether the patient was chronotropic incompetent.

For the MARS analysis we used heart rate recovery and these six covariates. This gave a total of seven covariates representing a parsimonious, but informative, list made up of key exercise test measurements, patient demographic information, and cardiac history. These were also the seven variables used in our quadratic model of Figure 4(b). Just as in the SVS analysis, we used ℓR_e for the Y value. We also restricted attention to data for which heart recovery was in the range of $[5, 25]$. Models were built using MARS's forward stepwise procedure with the total number of product truncated spline basis functions restricted to 35 with no more than three terms allowed for any basis function. All potential knots for each of the seven covariates were investigated in growing the model. Lack of fit for models were computed using the GCV (generalized cross-validation) criterion of Friedman (1991, Sec. 3.6). This is an adjusted mean square error, with a penalty adjustment for degrees of freedom using a cost complexity function. The cost complexity was estimated using 10-fold cross-validation. The model obtained from the forward stepwise procedure was then trimmed back using MARS's one-at-a-time backward stepwise procedure. This gave several potential models. The one with the smallest lack of fit (again computed using the GCV with the cross-validated complexity function) was selected as our final model.

4.1 Estimated Model

The final model contained 10 spline basis functions and 21 product spline basis functions representing an assortment of main effects and interaction terms. For convenience, these are listed below (mirrored basis functions are listed on the same line):

```
BF1 = max(0,HRRECOV - 18.000)          BF2 = max(0,18.000 - HRRECOV)
BF3 = max(0,PEAKHR - 129.000)          BF4 = max(0,129.000 - PEAKHR)
BF5 = max(0,AGE - 45.000)              BF6 = max(0,45.000 - AGE)
BF7 = max(0,FITNESS - 2.000)           BF8 = max(0,2.000 - FITNESS)
BF9 = max(0,PRIORCAD + .322569E-07)
BF10 = max(0,HRRECOV - 12.000)*BF5     BF11 = max(0,12.000 - HRRECOV)*BF5
BF12 = max(0,IMAGE + .875696E-07)
BF13 = max(0,HRRECOV - 20.000)*BF9     BF14 = max(0,20.000 - HRRECOV)*BF9
BF15 = max(0,FITNESS - 4.000)*BF10     BF16 = max(0,4.000 - FITNESS)*BF10
BF17 = max(0,FITNESS - 3.000)*BF10
BF19 = max(0,PEAKHR - 135.000)*BF5     BF20 = max(0,135.000 - PEAKHR)*BF5
BF21 = max(0,HRRECOV - 19.000)*BF19    BF22 = max(0,19.000 - HRRECOV)*BF19
BF23 = max(0,HRRECOV - 7.000)*BF6      BF24 = max(0,7.000 - HRRECOV)*BF6
BF25 = max(0,HRRECOV - 9.000)*BF3      BF26 = max(0,9.000 - HRRECOV)*BF3
BF27 = max(0,HRRECOV - 13.000)*BF5
BF29 = max(0,HRRECOV - 14.000)*BF5
BF31 = max(0,HRRECOV - 11.000)*BF5
BF33 = max(0,HRRECOV - 7.000)*BF7      BF34 = max(0,7.000 - HRRECOV)*BF7
BF35 = max(0,FITNESS - 1.000)*BF23
```

Many of the spline functions are related to heart rate recovery; in fact, of the 21 product spline interactions, 19 involve heart recovery. These product terms show that MARS is using different knot values for heart recovery in tandem with other variables to estimate log relative risk. The variables primarily involved are peak heart rate, age, and fitness. It is tricky, though, to decipher from the basis functions exactly what the underlying relationship is.

We used a Trellis conditional plot to gain better insight into how ℓR_e depends on covariates. Values for ℓR_e were color coded and displayed in a multipanel format. See Figure 6 in gray scale. Each panel shows how ℓR_e varies as a function of heart rate recovery and peak heart rate, conditioned on the value of age and fitness. Age was broken into two groups corresponding to patients less than or equal to 45 years and greater than 45

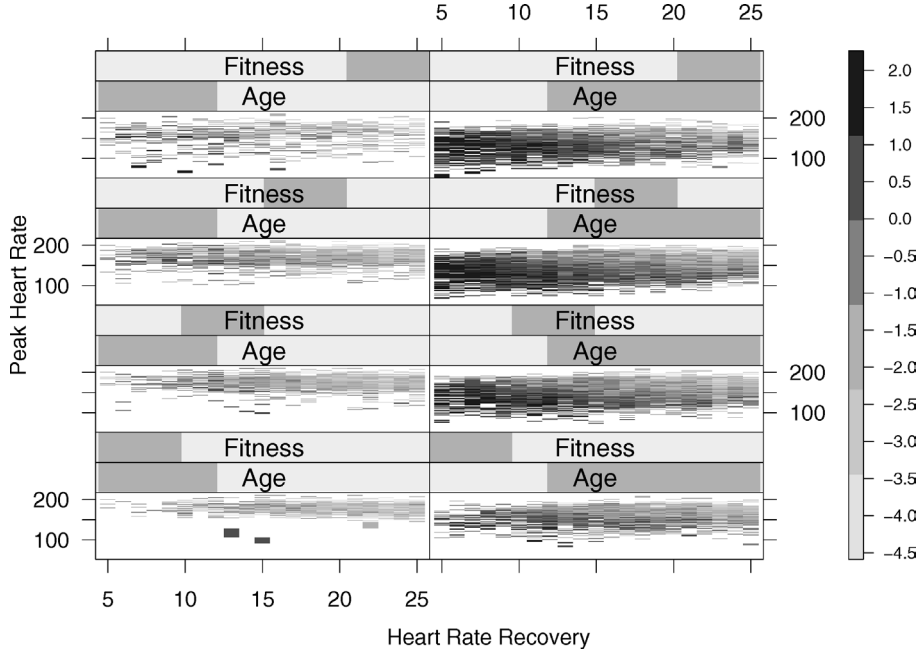

Figure 6. Multipanel Conditional Plot of Ensemble Log Relative Risk as a Function of Heart Rate Recovery, Peak Heart Rate, Age, and Fitness. Age is conditioned into two groups corresponding to patients less than or equal to 45 years and greater than 45 years of age. Levels of fitness are 1–2, 2–3, 3–4, and 4–5. The gray scale key on right side indicates the ensemble log relative risk value.

years of age. This choice was suggested by the basis functions BF5 and BF6 and their product interactions with heart rate recovery BF23, BF24, BF27, BF29, and BF31. Levels of fitness were conditioned into four overlapping groups: 1–2, 2–3, 3–4, and 4–5.

Looking at the left- and right-hand plots, we see an age effect. Plots on the right, corresponding to older patients, exhibit a tendency toward higher log relative risks. A fitness effect is also discernible. From top to bottom there is an overall decrease in logrelative risk, showing that fitness improves risk outlook (a value of 1 indicates best overall fitness). Observe how this trend is more noticeable on the right-hand plots, suggesting that there is an age interaction effect at play as well (the functions BF15, BF16, and BF17 from our MARS analysis also confirm this). Also of special interest is the interaction between peak heart rate and heart rate recovery. Peak heart rate is often mistakenly thought of as a surrogate of heart recovery; however, Figure 6 dispels this belief. This can be seen most clearly on the right-hand plots. As heart rate recovery increases, log-relative risk decreases; however, the rate of decrease has a complex relationship with peak heart rate. Notice that this effect is slightly different across fitness levels. Thus fitness may also play a role in this interaction.

5. SUMMARY AND DISCUSSION

Relative risk forests, a synthesis of random forests and relative risk trees, were introduced as a simple, but powerful, method for estimating relative risk values. Our motivation for introducing this method was to gain a better understanding of the complex dynamics between mortality, heart rate recovery, a known powerful predictor of mortality, and other clinical and physiological characteristics of a patient collected during stress testing. Among the more interesting findings from our analysis was a fairly elaborate functional relationship between log relative risk and heart rate recovery. From this relationship we were able to determine a heart rate recovery threshold rule for identifying patients with abnormal heart behavior which confirms previous work. Postprocessing estimated log relative risk values showed that peak heart rate has a complex relationship with heart rate recovery and mortality, dispelling a popularly held belief that peak heart rate is a surrogate of heart recovery. We also discovered an important interaction between level of fitness, heart rate recovery, and age.

Methodologically, relative risk forests provide a simple, automated way to develop nonparametric survival regression models and should be especially useful in settings where complex interactions may exist among variables. Relative risk forests give interpretable risk assignments to a patient made relative to an automatically selected mean unit in the study. Such a risk value is meaningful and correct under the assumption of proportional hazards, an implicit assumption of the method. In examples where the proportional hazards assumption is known to be too simplistic, the method may not be appropriate, and other nonparametric techniques should be used instead. For example, the HARE method developed by Kooperberg, Stone, and Truong (1995) applies to general hazard functions.

Outside the restriction to proportional hazards, relative risk forests can be used generally. The method is easily implemented in standard software and requires little tuning. In fact, the only tuning parameter in the procedure involves selecting F, the number of randomly selected covariates used in growing a tree. As we have shown, the value for F can be chosen using a simple cross-validation method. We also note that although our methods in this paper focused on growing supervised trees, which always included heart rate recovery, the method can be used in a completely unsupervised fashion, or for that matter, it could be adapted to include more elaborate forms of supervision (we are currently experimenting with the idea of selecting covariates according to an automated weighting scheme, with covariates considered more interesting given higher weights of being included in the tree building process). Such modifications are straightforward.

Relative risk forests can also be used to predict relative risk values. As was done in Section 4, ensemble log relative risk values can be used as the response in a nonparametric regression and the resulting equation used for prediction; a similar approach was also discussed in LeBlanc and Crowley (1999) who fit adjusted dependent variables and martingale residuals from Cox models using adaptive regression splines. The use of a regression model has the advantage of a closed-form expression for the log relative risk in terms of predictors. However, interpreting models in clinical terms can be problematic because of the possibility of higher-order interactions. Graphical tools such as multipanel conditional Trellis plots are a simple and effective method that can be used. Another graphical approach that could be tried are the partial dependence plots discussed in Hastie, Tibshirani, and Friedman (2001, chap. 10.13).

[Received January 2003. Revised January 2004.]

REFERENCES

Aalen, O. O. (1978), "Nonparametric Inference for a Family of Counting Processes," *The Annals of Statistics*, 6, 701–726.

Aitkin, M., and Clayton, D. (1980), "The Fitting of Exponential, Weibull and Extreme Value Distributions to Complex Censored Survival Data Using GLIM," *Applied Statistics*, 29, 156–163.

Amit, Y., and Geman, D. (1997), "Shape Quantization and Recognition With Randomized Trees," *Neural Computation*, 9, 1545–1588.

Arai, Y., Saul, J. P., Albrecht, P. et al. (1989), "Modulation of Cardiac Autonomic Activity During and Immediately After Exercise," *American Journal of Physiology*, 256, 132–141.

Breiman, L. (1996), "Heuristics of Instability and Stabilization in Model Selection," *The Annals of Statistics*, 24, 2350–2383.

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45, 5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984), *Classification and Regression Trees*, Belmont, CA: Wadsworth.

Breslow, N. (1972), "Contribution to the Discussion of Regression Models and Life Tables by D. R. Cox," *Journal of the Royal Statistical Society*, Ser. B, 34, 216–217.

Clayton, D., and Cuzick, J. (1985), "The EM Algorithm for Cox's Regression Model Using GLIM," *Applied Statistics*, 34, 148–156.

Cole, C. R., Blackstone, E. H., Pashkow, F. J., Snader, C. E., and Lauer, M. S. (1999), "Heart-Rate Recovery Immediately After Exercise as a Predictor of Mortality," *New England Journal of Medicine*, 341, 1351–1357.

Cole, C. R., Foody, J. M., Blackstone, E. H, and Lauer, M. S. (2000), "Heart Rate Recovery After Submaximal Exercise Testing as a Predictor of Mortality in a Cardiovascularly Healthy Cohort," *Annals of Internal Medicine*, 132, 552–555.

Cox, D. R. (1972), "Regression Models and Life Tables," *Journal of the Royal Statistical Society*, Ser. B, 34, 187–200.

Diaz, L. A., Brunken, R. C., Blackstone, E. H., Snader, C.E, and Lauer, M. S. (2001), "Independent Contribution of Myocardial Perfusion Defects to Exercise Capacity and Heart Rate Recovery for Prediction of All-Cause Mortality in Patients With Known or Suspected Coronary Heart Disease," *Journal of the American College of Cardiology*, 37, 1558–1564.

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines" (with discussion), *The Annals of Statistics*, 19, 1–141.

Gibbons, R. J., Balady, G. J, Beasley J. W. et al. (1997), "ACC/AHA Guidelines for Exercise Testing," A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Exercise Testing), *Journal of the American College of Cardiology*, 30, 260–311.

Gibbons, R. J., Balady, G. J., Bricker, J. T. et al. (2002a), "ACC/AHA 2002 Guideline Update for Exercise Testing: Summary Article," A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines), *Journal of the American College of Cardiology*, 40, 1531–1540.

Gibbons, R. J., Balady, G. J., Bricker, J. T. et al. (2002b), "ACC/AHA 2002 Guideline Update for Exercise Testing: Summary Article," A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Update the 1997 Exercise Testing Guidelines), *Circulation*, 106, 1883–1892.

Hammond, H. K., and Froelicher, V. F. (1985), "Normal and Abnormal Heart Rate Responses to Exercise," *Progress in Cardiovascular Diseases*, 27, 271–296.

Hastie, T., Tibshirani, R., and Friedman, J. (2001), *The Elements of Statistical Learning*, New York, Springer-Verlag.

Imai, K., Sato, H., Hori, M. et al. (1994), "Vagally Mediated Heart Rate Recovery After Exercise Is Accelerated in Athletes but Blunted in Patients With Chronic Heart Failure," *Journal of the American College of Cardiology*, 24, 1529–1535.

Ishwaran, H., and Rao, J. S. (2000), "Bayesian Nonparametric MCMC for Large Variable Selection Problems," unpublished manuscript.

Ishwaran, H., and Rao, J. S. (2003), "Detecting Differentially Expressed Genes in Microarrays Using Bayesian Model Selection," *Journal of the American Statical Association*, 98, 438–455.

Kooperberg, C., Stone, C. J., and Truong, Y. K. (1995), "Hazard Regression," *Journal of the American Statical Association*, 90, 78–94.

La Rovere, M. T., Bigger, J. T., Marcus, F. I, Mortara, A., and Schwartz, P. J. (1998), "Baroreflex Sensitivity and Heart-Rate Variability in Prediction of Total Cardiac Mortality After Myocardial Infarction: ATRAMI (Autonomic Tone and Reflexes After Myocardial Infarction) Investigators," *Lancet*, 351, 478–484.

La Rovere, M. T., Pinna, G. D., Hohnloser, S. H. et al. (2001), "Baroreflex Sensitivity and Heart Rate Variability in the Identification of Patients at Risk for Life-Threatening Arrhythmias: Implications for Clinical Trials," *Circulation*, 103, 2072–2077.

Lauer, M. S. (2001), "Exercise Electrocardiogram Testing and Prognosis: Novel Markers and Predictive Instruments," *Cardiology Clinician*, 19, 401–414.

Lauer, M. S., Alexe, S., Pothier, C. E., Blackstone, E. H., Ishwaran, H. and Hammer, P. L. (2002), "Use of the 'Logical Analysis of Data' Method for Assessing Long-Term Mortality Risk After Exercise Electrocardiography," *Circulation*, 106, 685–690.

Lauer, M. S., Francis, G. S., Okin, P. M., Pashkow, F. J., Snader, C. E., and Marwick, T. H. (1999), "Impaired Chronotropic Response to Exercise Stress Testing as a Predictor of Mortality," *Journal of the American Medical Association*, 281, 524–529.

Lauer, M. S., Okin, P. M., Larson, M. G., Evans, J. C., and Levy, D. (1996), "Impaired Heart Rate Response to Graded Exercise: Prognostic Implications of Chronotropic Incompetence in the Framingham Heart Study," *Circulation*, 93, 1520–1526.

LeBlanc, M., and Crowley, J. (1992), "Relative Risk Trees for Censored Survival Data," *Biometrics*, 48, 411–425.

LeBlanc, M., and Crowley, J. (1999), "Adaptive Regression Splines in the Cox Model," *Biometrics*, 55, 204–213.

Nishime, E. O., Cole, C. R., Blackstone, E. H., Pashkow F. J., and Lauer, M. S. (2000), "Heart Rate Recovery and Treadmill Exercise Score as Predictors of Mortality in Patients Referred for Exercise ECG," *Journal of the American Medical Association*, 284, 1392–1398.

Schwartz, P. J., La Rovere, M. T. and Vanoli, E. (1992), "Autonomic Nervous System and Sudden Cardiac Death: Experimental Basis and Clinical Observations for Post-Myocardial Infarction Risk Stratification," *Circulation*, 85, 77–91.

Shetler, K., Marcus, R., Froelicher, V. F. et al. (2001), "Heart Rate Recovery: Validation and Methodologic Issues," *Journal of the American College of Cardiology*, 38, 1980–1987.

Snader, C. E., Marwick, T. H., Pashkow, F. J., Harvey, S. A., Thomas, J. D., and Lauer M. S. (1997), "Importance of Estimated Functional Capacity as a Predictor of All-Cause Mortality Among Patients Referred for Exercise Thallium Single-Photon Emission Computed Tomography: Report of 3,400 Patients From a Single Center," *Journal of the American College of Cardiology*, 30, 641–648.

Therneau, T. M., and Atkinson, E. J. (1997), "An Introduction to Recursive Partitioning Using the RPART Routines," Technical Report Series No. 61, Department of Health Science Research, Mayo Clinic, Rochester, MN.

Watanabe, J., Thamilarasan, M., Blackstone, E. H., Thomas, J. D., and Lauer, M. S. (2001), "Heart Rate Recovery Immediately After Treadmill Exercise and Left Ventricular Systolic Dysfunction as Predictors of Mortality: The Case of Stress Echocardiography," *Circulation*, 104, 1911–1916.

Williams, S. V., Fihn, S. D., and Gibbons, R. J. (2001), "Guidelines for the Management of Patients With Chronic Stable Angina: Diagnosis and Risk Stratification," *Annals of Internal Medicine*, 135, 530–547.