A Multilevel Approach to Individual Tree Survival Prediction

Charles E. Rose Jr., Daniel B. Hall, Barry D. Shiver, Michael L. Clutter, and Bruce Borders

Abstract: Traditionally, modeling of permanent plot individual tree survival has not considered the multiple sources of heterogeneity and correlation that may arise due to the multilevel data structure inherent in the design (e.g., clustering of trees within a plot). Permanent plots are sampled periodically; therefore, data are interval-censored because it is only known that a tree died between two successive measurement occasions. Here, we adopt the complementary log–log (CLL) function for modeling permanent plot interval-censored individual tree survival data. The CLL function is derived directly from the likelihood function of a fully specified statistical model that accounts for interval censoring. In addition, our CLL model considers silvicultural treatment effects on survival. Our data are from young permanent plot loblolly pine plantations that have been measured annually beginning at age 1 year. Each plot was randomly assigned one of four cultural treatments: control (C), herbicide (H), fertilization (F), and herbicide and fertilization (HF). Here we extend the individual tree survival CLL model to allow for trees within a plot and plot level random effects. We demonstrate individual tree survival predictions with and without the inclusion of random effects. FOR. SCI. 52(1):31–43.

Key Words: Hazard function, interval-censored, random effects, complementary log–log function, proportional hazards.

MORTALITY PREDICTION is an important component of individual tree forest growth and yield systems. Modeling individual tree mortality, or alternatively survival (hereafter the modeling of survival or mortality is used interchangeably) began in the 1970s with a study by Hamilton (1974) in which he used the logistic equation. The logistic equation has become the individual tree survival model of choice (e.g., Monserud 1976, Wykoff et al. 1982, Vanclay 1991, Shen et al. 2001), which is perhaps due to ease of parameter interpretation. The logistic equation has been used to model individual tree mortality for a variety of species and stand conditions (e.g., Hamilton 1986, Monserud and Sterba 1999, Huebschmann et al. 2000, Yao et al. 2001). Although the logistic cumulative distribution function is the most widely used for modeling individual tree survival, other models have been applied for the same purpose, such as the Richards function, and Weibull, gamma, and negative binomial distributions (e.g., Buford and Hafley 1985, Kobe and Coates 1997). In addition, nontraditional individual tree survival methods have been developed such as the binary classification tree (CART) (Dobbertin and Biging 1998) and an artificial neural network (Guan and Gertner 1991). Other recent individual tree-survival analysis studies have used the Cox proportional hazards model (Volney 1998) and the log-normal distribution (Preisler and Slaughter 1997) to allow for more flexibility in the hazard function. However, individual tree survival models have generally ignored the heterogeneity that may occur due to the clustering of trees within a plot.

Multiple sources of heterogeneity occur naturally for most permanent plot forestry studies because of the multilevel data structure inherent in the design, i.e., measurement occasions are nested within trees (e.g., repeated measurements) and trees are nested within plots. In addition, forestry survival data are usually interval-censored because it is usually only known that a tree died within an interval (e.g., between successive survey dates). Recently, multilevel mixed effects models have become more common in forestry (e.g., Gregoire et al. 1995, Hall and Bailey 2001, Fang and Bailey 2001). However, these multilevel models typically have a continuous response and the lowest level variation is assumed normally distributed. This assumption is not appropriate when modeling binary or proportion data, because the data are discrete or limited in range. Use of binary response multilevel mixed-effects models has gained popularity in recent years (e.g., Barbosa and Goldstein 2000, Biggeri et al. 2001, Yang and Goldstein 2000). Binary response models typically use the probit, logit, or complementary log–log (CLL) link functions (e.g., Biggeri et al. 2001, Hedeker et al. 2000). According to Allison (1999), there are several important features related to the probit, logit, and CLL models. Results from the probit and logit models typically differ little and rarely lead to conflicting conclusions. In addition, it generally takes a very large sample to distinguish between the probit and logit models and both are symmetrical around the probability 0.50. In contrast, the CLL model is asymmetrical and approaches zero much slower than it approaches one and is invariant to

Charles E. Rose Jr., Mathematical Statistician, Anthrax Vaccine Safety Team NIP/CDC, 1600 Clifton Rd., Mailstop E-61, Atlanta, GA 30333—Phone: (404) 639-3028; Fax: (404) 639-2483; cvr7@cdc.gov. Daniel B. Hall, Associate Professor, University of Georgia, Department of Statistics, Athens, GA 30602-1952—Phone: (706) 542-3302; dhall@stat.uga.edu. Barry D. Shiver, Professor, University of Georgia, Daniel B. Warnell School of Forest Resources, Athens, GA 30602—Phone: (706) 542-3009; shiver@smokey.forestry.uga.edu. Michael L. Clutter, Professor, University of Georgia, Daniel B. Warnell School of Forest Resources, Athens, GA 30602—Phone:(706) 542-5448; mclutter@smokey.forestry.uga.edu. Bruce Borders, Professor, University of Georgia, Daniel B. Warnell School of Forest Resources, Athens, GA 30602—Phone: (706) 542-6205; bborders@arches.uga.edu.

the length of the time interval. Note that, because of the CLL asymmetrical property, it is important to define the outcome, death or survival, correctly. In addition, the CLL is closely related to the Cox proportional hazards model and both models estimate the same underlying coefficients, and therefore have the same interpretation.

Recent studies have demonstrated the importance of considering multiple sources of heterogeneity for a binary response model. Here, we adopt a Cox proportional hazards model for modeling interval-censored individual tree binary response survival data and incorporate random effects to account for multiple sources of variability.

Survival Analysis Synopsis

Survival analysis focuses on a group or groups of individuals that have a well-defined failure event (e.g., death). The survival function can be defined by letting a random variable T represent the time until event of interest. Let the cumulative distribution function of T be $F(t) = \Pr[T \leq t]$, and let $f(t)$ be the corresponding probability density function. Then the survival function is defined as $S(t) = \Pr(T > t) = 1 - F(t)$, which is the proportion of the population still alive at time t. The hazard function is the instantaneous rate of mortality assuming the individual has survived to time t (Collet 1994) and is defined as

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}. \tag{1}$$

There are several reasons why the hazard function (Equation 1) is important. First, the immediate risk can be determined for an individual known to be alive at age t. Second, it may be easier to compare groups of individuals using the hazard function. Last, hazard function models are convenient when there is censoring or several types of failure. The cumulative hazard function is defined as

$$H(t) \int_0^t h(x)\, dx. \tag{2}$$

Equation 2 can be thought of as the cumulative risk, i.e., the sum of all risks faced by an individual during the period 0 to t. The survival function is related to the cumulative hazard function as

$$S(t) = \exp\left(-\int_0^t h(x)\, dx\right). \tag{3}$$

Equation 3 implies that the probability of surviving to t is a function of the hazard at all durations up to t. These relationships indicate that the survival and hazard functions, Equations 3 and 1, provide alternative characterizations of the distribution of T.

Survival analysis usually assumes a hazard function that arises from one of three general families of survival models: the proportional hazards, accelerated life, or proportional odds models. Here we adopt the Cox proportional hazard

model, which assumes that $h_j(t) = \exp(\eta_j)h_0(t)$, where η_j is a function of explanatory variables for the jth subject and $h_0(t)$ is an unspecified baseline hazard function. The corresponding survival function for the jth subject is given by $S_j(t) = [S_0(t)]^\phi$, where $\phi = e^{\eta_j}$ and $S_0(t)$ is the baseline survivor function.

Individual Tree Survival Model Development

Our survival model formulation assumes the data are from a permanent plot plantation study in which measurement occasions are nested within trees and trees are nested within plots. All trees enter the study at $t_0 = 0$; however, the calendar time corresponding to t_0 may vary by plot. Each tree is followed to time t_m, where $m = n_{jk}$ and n_{jk} is the number of measurement occasions for the jth tree on the kth plot. Note that the time in which a tree enters the study, t_0, is not a measurement occasion regardless of the age of the tree at time t_0. In addition, $t_{m+1} = \infty$ for trees that are still alive at their last respective measurement occasion. We begin by defining the probability of a tree dying during the ith time interval (t_{i-1}, t_i), $i = 1, 2, \ldots, m$. A tree's mortality noted at time t_i had an actual death time of t, where $t_{i-1} < t \leq t_i$. Let p_{ijk} be the probability of mortality occurring for the jth tree on plot k during the ith time interval, i.e., $p_{ijk} = \Pr(t_{i-1} < T_{jk} \leq t_i)$. Here T_{jk} is a random variable associated with the survival time for the jth tree on plot k. The conditional probability of mortality occurring during the ith time interval given that the death occurs after t_{i-1} is $\pi_{ijk} = \Pr(t_{i-1} < T_{jk} \leq t_i \mid T_{jk} > t_{i-1})$, where $i = 1, 2, \ldots, m + 1$. Now we define p_{ijk} in terms of π_{ijk} as $p_{ijk} = (1 - \pi_{1jk})(1 - \pi_{2jk}) \cdots (1 - \pi_{i-1,jk}) \pi_{ijk}$, where $i = 2, 3, \ldots, m + 1$ and $p_{1jk} = \pi_{1jk}$. Complementary probabilities are associated with a tree not dying during an interval and p_{1jk} is the probability it died during the first interval.

Now, the likelihood function is developed using the conditional probabilities. Let $\delta_{ijk} = 1$ if the jth tree on the kth plot dies in the interval from t_{i-1} to t_i and zero if it is alive. Hence, the likelihood function for the total number of observations for all j trees on all k plots is

$$L = \prod_{k=1}^{n} \prod_{j=1}^{n_k} \prod_{i=1}^{m+1} p_{ijk}^{\delta_{ijk}}. \tag{4}$$

Now substitute for p_{ijk} in Equation 4. In addition, because our endpoint is mortality the probability of death in the (m, ∞) interval is one, i.e., $\pi_{m+1,jk}^{\delta_{m+1,jk}} = 1$. Hence, the likelihood function can be expressed in terms of the π_{ijk} values as

$$L = \prod_{k=1}^{n} \prod_{j=1}^{n_k} \prod_{i=1}^{n_{jk}} \pi_{ijk}^{\delta_{ijk}} (1 - \pi_{ijk})^{1-\delta_{ijk}}. \tag{5}$$

Note that Equation 5 is the likelihood function for a binomial distribution with parameters 1 and π_{ijk} corresponding to $N = \sum_{k=1}^{n} \sum_{j=1}^{n_k} n_{jk}$ independent Bernoulli trials. Here N is the total number of measurement occasions for all trees in the data set.

Model formulation proceeds by noting that the survival function gives the probability that the jth tree on the kth plot dies after time i. Therefore, we can define the conditional survival function for the jth tree on the kth plot as

$$1 - \pi_{ijk} = \Pr(T_{jk} \geq t_i | T_{jk} \geq t_{i-1}) = \frac{S_{jk}(t_1)}{S_{jk}(t_{i-1})}. \quad (6)$$

Here we use Equation 6 and adopt a Cox proportional hazards model, which can be expressed in terms of the hazard rate at time t_i for the jth tree on plot k as $h_{jk}(t_i) = \exp(\eta_{jk})h_0(t_i)$. Where $\eta_{jk} = \beta_1 x_{1jk} + \beta_2 x_{2jk} + \cdots + \beta_p x_{pjk}$ and $h_0(t_i)$ is the baseline hazard function. This model assumes the hazards are proportional only at the specified times, which is a relaxation of the usual proportional hazards model, which always assumes proportionality (Collet 1994). Expressing the Cox proportional hazard model in terms of the survival function Equation 6 yields

$$1 - \pi_{ijk} = \left[\frac{S_0(t_1)}{S_0(t_{i-1})} \right]^{\exp(\eta_{jk})}. \quad (7)$$

After taking logs of Equation 7, we have the grouped time version of the continuous time proportional hazards model (McCullagh 1980), i.e.,

$$\log\{-\log[1 - \pi_{ijk}]\} = \eta_{jk} + \log\left[-\log\left(\frac{S_0(t_i)}{S_0(t_{i-1})} \right) \right]$$

$$= \eta_{jk} + \kappa_i. \quad (8)$$

Hence, Equation 8 is a linear model for the CLL transformation of π_{ijk} in which the parameters κ_i are associated with the ith time interval. The parameters κ_i yield estimates of the baseline log–log survival function. Coefficients of Equation 8 can be interpreted just as in a proportional hazards model (Allison 1995), i.e., Equation 8 expresses the covariate effects on the log of the integrated hazard function, defined as $H(t) = -\log(1 - \pi_{ijk}) = \exp(\eta_{jk} + \kappa_i)$. Thus for a covariate x, it expresses the $100(e^{\beta x} - 1)$ percent change in the hazard of death for a one-unit increase in x. Incorporating time-dependent covariates into the model necessitates relaxing the assumption of proportional hazards at the cut points. The CLL function that includes time-varying covariates would likely be sufficient for modeling forest mortality if the data structure were not hierarchical. However, data from permanent plot forestry studies inherently have a multilevel structure.

Multilevel Individual Tree Survival Model

Multilevel models assume multiple sources of heterogeneity and recognize units at one level as grouped (nested) in the next higher level. Here we adopt the nomenclature in which the error term is identified with the lowest level. Hence, we have measurement occasions nested within trees (level 1), trees nested within plots (level 2), and plots (level 3). According to Goldstein (1991), explicitly modeling the manner in which subjects are grouped allows the analyst to obtain statistically efficient estimates of the regression co-

efficients. In addition, if the relationship between the response and the covariates is nonlinear (e.g., logistic regression), then ignoring grouping structures can result in large biases in the parameter estimates (Rodriguez and Goldman 2001).

Our multilevel model formulation considers tree and plot levels by defining y_{ijk} as one if the jth tree on the kth plot dies during the ith interval, zero otherwise. Given plot-specific random effects $\mathbf{b}_{(3)k}$ and tree within plot-specific random effects $\mathbf{b}_{(2)jk}$, we assume that the y_{ijk} values are independent Bernoulli random variables with conditional expectation π_{ijk}. Here π_{ijk} is defined as the conditional probability that the jth tree on the kth plot succumbs during the ith interval given that it has survived the previous intervals and given $\mathbf{b}_{(3)k}$ and $\mathbf{b}_{(2)jk}$. Our multilevel individual tree survival model can be expressed as

$$y_{ijk} | \mathbf{b}_{(2)jk}, \mathbf{b}_{(3)k} \sim \text{Binomial}(1, \pi_{ijk})$$

$$\log\{-\log(1 - \pi_{ijk})\} = \kappa_i + \boldsymbol{\beta}' \mathbf{x}_{ijk} + \mathbf{b}'_{(2)jk} \mathbf{z}_{(2)ijk} + \mathbf{b}'_{(3)k} \mathbf{z}_{(3)ijk}$$

$$= \kappa_i + \sum_{l=1}^{p} \beta_l x_{ijkl} + \sum_{l=1}^{q_2} b_{(2)jkl} z_{(3)ijkl}$$

$$+ \sum_{l=1}^{q_3} b_{(3)kl} z_{(3)ijkl} \quad (9)$$

$$\mathbf{b}_{(2)jk} \sim N_{q_2}(\mathbf{0}, \mathbf{D}_{(2)}) \qquad \mathbf{b}_{(3)k} \sim N_{q_3}(\mathbf{0}, \mathbf{D}_{(3)})$$

where $\mathbf{x}_{ijk}$ is the $p \times 1$ covariate vector associated with the $p \times 1$ vector of fixed effects β, p is the number of fixed effects parameters in Model 9 corresponding to the $\mathbf{x}_{ijk}$ covariates, $\mathbf{z}_{(2)ijk}$ is the level 2 covariate vector associated with random effects vector $\mathbf{b}_{(2)jk}$ (both of dimension $q_2 \times 1$), $\mathbf{z}_{(3)ijk}$ is the level 3 covariate vector associated with random effects vector $\mathbf{b}_{(3)j}$ (both of dimension $q_3 \times 1$), and κ_i are the baseline hazards associated with the ith interval. Here, q_m is the number of random effects at level m in Model 9, and it is assumed that random effects are independent across different levels, but possibly dependent within a given level with variance–covariance matrices $\text{var}(\mathbf{b}_{(2)jk}) = \mathbf{D}_{(2)}$ and $\text{var}(\mathbf{b}_{(3)k}) = \mathbf{D}_{(3)}$.

Our previous likelihood function, Equation 5, is still appropriate for the multilevel model, but it is now conditional on the random effects. Define $\boldsymbol{\delta}$ as the binary response vector pattern for the n plots having n_k trees observed at n_{jk} measurement occasions. Hence, $\boldsymbol{\delta}$ has elements $\delta_{ijk} = y_{ijk}$, $i = 1, \ldots, n_{jk}$, $j = 1, \ldots, n_k$, $k = 1, \ldots, n$. Assuming independence of the response vector conditional on the random effects, the conditional likelihood function can be expressed as (Gibbons and Hedeker 1997)

$$L(\delta, \boldsymbol{\beta} | \mathbf{b}_{(3)k}, \mathbf{b}_{(2)jk}) = \prod_{k=1}^{n} \prod_{j=1}^{n_k} \prod_{i=1}^{n_{jk}} (\pi_{ijk})^{\delta_{ijk}} (1 - \pi_{ijk})^{1 - \delta_{ijk}}.$$

$$(10)$$

The marginal distribution of δ for the jth tree on the kth plot

can be obtained from Equation 10 by integrating over F(**b**), the joint distribution of the combined random effects vector **b**, which is multivariate normal; i.e.,

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{n} \int \prod_{j=1}^{n_k} \prod_{i=1}^{n_{jk}} (\pi_{ijk})^{\delta_{ijk}} (1 - \pi_{ijk})^{1-\delta_{ijk}} \, dF(\mathbf{b}). \quad (11)$$

However, the integral of Equation 11 is difficult to evaluate because there is no closed-form solution when assuming a multivariate normal distribution for the random effects. Because of this difficulty in parameter estimation, the use of multilevel models to analyze binary response variables is relatively recent (e.g., Goldstein 1995, Hedeker et al. 2000, Biggeri et al. 2001, Rodriguez and Goldman 2001).

Parameter estimation challenges are mainly related to the need to integrate over the random effects distribution and the fact that this integral is intractable for a nonlinear model with non-normal distribution. In our study, it was necessary to restrict attention to methods of estimation that were computationally feasible, i.e., methods that converged for the considered models. A common multilevel parameter estimation technique for a binary response variable, and the one we adopt is the first-order marginal quasi-likelihood (MQL-1) method (Breslow and Clayton 1993). This method is motivated by using a linearization of the multilevel nonlinear model. MQL-1 approximates $L(\boldsymbol{\beta})$ by using a first-order Taylor series expansion of the model function around the true values of the fixed effects $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and random effects $\mathbf{b} = \mathbf{0}$. The MLwiN (Rasbash et al. 1999) software package was used to estimate our multilevel survival model parameters. We chose the MLwiN software over software packages such as SAS or S-Plus for two primary reasons. First, MLwiN was developed as a multilevel model fitting software and offers more methods for fitting complex multilevel models than most alternative software. Second, MLwiN can fit multilevel models with more than two levels for a binary response, whereas S-Plus could not (at the time this study was conducted) and SAS is inefficient for fitting these models with large data sets.

Data

Permanent-plot loblolly pine data were obtained from the Consortium for Accelerated Pine Production Studies (CAPPS), which is overseen by the Warnell School of Forestry at the University of Georgia. Loblolly pine plantations were established throughout Georgia at Athens, Dawsonville, Eatonton, Thompson, Tifton, and Waycross. The study called for two complete blocks to be established at each location with each block containing four 0.15-ha treatment plots. These were established at each location using bareroot seedlings planted with a 2.44-m × 2.44-m spacing. A 0.05-ha measurement plot was centered within each of the treatment plots. Four cultural treatments were randomly assigned to the blocks at each location. These cultural treatments are

1. Herbicide (H): Spray plot with nonsoil active herbicide as needed to maintain complete control of woody and herbaceous vegetation.

2. Fertilization (F): Apply recommended rates of fertilizer to achieve and maintain accelerated growth rates.

3. Herbicide–Fertilization (HF): Apply both herbicide and fertilization treatments.

4. Control (C): No cultural treatment.

The original study called for a replication of all treatment plots every 2 years for the first 10 years of the study. This protocol would have resulted in five complete sets of experimental plots at all installations. However, the actual study varies from the protocol because of limitations that required the number of experimental plots at installations to be repeated at different intervals for different locations. The plots have been measured annually beginning at age one. The CAPPS data contain 112,365 total observations for 11,956 trees on 146 plots. Study survival data are summarized by plot distribution (Table 1) and plot and tree attributes (Table 2).

Preliminary Analysis

We conducted preliminary analysis to detect survival and hazards trends and to investigate trends in potential time-varying covariates to be included in the linear component of the CLL function. Survival and hazards by age were estimated using the nonparametric Kaplan–Meier (KM) product limit survival estimator (Kaplan and Meier 1958). Kaplan–Meier survival estimates for the plots reveal that the underlying hazard function is bathtub-shaped (Figure 1). In addition, the KM survival and hazard function estimates by treatment (Figure 2) reveal that, using the C treatment as the baseline, there is an acceleration in mortality for the F treatment in the early years and again at about age 12. The HF treatment hazards appear constant relative to the C treatment until about age 8, after which begins acceleration in mortality. The H treatment has a deceleration in mortality relative to the C treatment. An age by treatment interaction is evident for the hazards, which suggests nonproportional hazards at the cut points. The empirical complementary log–log was computed by grouping observations into quintiles for all time-varying covariates. The basal area per hectare (BA/ha) and relative spacing (RS) were the only

Table 1. The CAPPS study plot distribution by year planted for the spectrum of plots and by treatment

Year planted	Plot age	Plots	Plots by treatment			
			Control	Fertilized	Herbicide	HF[1]
1986	14	26	8	4	8	6
1987	13	28	8	6	8	6
1988	12	36	10	8	10	8
1989	11	24	8	4	8	4
1992	8	20	6	4	6	4
1994	6	12	4	2	4	2
Total		146	44	28	44	30

[1] HF is the fertilizer and herbicide treatment.

time-varying covariates to exhibit quadratic trends. Therefore, we considered these attributes as both linear and quadratic terms in Model 9.

Model-Fitting Procedure

Fang and Bailey (2001) noted that the determination of which parameters are purely fixed and which are both fixed and random is frequently data-dependent. Pinheiro and Bates (2000) recommend using an "inside-out" model-building strategy to determine which coefficients are likely to be common to all groups and which appear to be group-specific. We start with a model that allows for plot and tree random effects at the cut-points of Model 9 and has linear terms for the considered covariates. A cut-point is analogous to an intercept for each interval and treatment combination. Let $\log(-\log(1 - \pi_{ijk}))$ denote the CLL function for the ith interval of the jth tree on the kth plot. There is no evidence to suggest that the variance components can be adequately modeled using one variance component for all cultural treatments. In addition, preliminary analysis suggests a treatment by age interaction with respect to the hazards (Figure 2) and the necessity of time-varying covariates. Therefore, our baseline mixed effects survival model is

$$y_{ijk}|\mathbf{b}_{(2)jk}, \mathbf{b}_{(3)k} \sim \text{Binomial}(1, \pi_{ijk})$$

$$\log\{-\log(1 - \pi_{ijk})\} = \kappa_i + \kappa_{i,F}x_{ijk,F} + \kappa_{i,H}x_{ijk,H}$$
$$+ \kappa_{i,HF}x_{ijk,HF} + \boldsymbol{\beta}'\mathbf{x}_{ijk} + \mathbf{b}'_{(2)jk}\mathbf{z}_{(2)ijk}$$
$$+ \mathbf{b}'_{(3)k}\mathbf{z}_{(3)ijk}$$

$$\boldsymbol{\beta}'\mathbf{x}_{ijk} = \beta_1 \text{BA/ha}_{ijk} + \beta_2 \text{TPH}_{ijk} + \beta_3 \text{RS}_{ijk}$$
$$+ \beta_4 H^{\text{Plot}}_{ijk} + \beta_5 \text{dbh}^{\text{Plot}}_{ijk} + \beta_6 \text{dbh}_{ijk} + \beta_7 H_{ijk}$$

$$\mathbf{b}_{(2)jk} = (b_{(2)jk,C}, b_{(2)jk,F}, b_{(2)jk,H}, b_{(2)jk,HF})' \quad (12)$$

$$\mathbf{b}_{(3)k} = (b_{(3)k,C}, b_{(3)k,F}, b_{(3)k,H}, b_{(3)k,HF})'$$

$$\mathbf{z}_{(2)ijk} = \mathbf{z}_{(3)ijk} = (z_{ijk,C}, z_{ijk,F}, z_{ijk,H}, z_{ijk,HF})'$$

$$\mathbf{b}_{(2)jk} \sim N_4(\mathbf{0}, \mathbf{D}_{(2)}), \qquad \mathbf{b}_{(3)k} \sim N_4(\mathbf{0}, \mathbf{D}_{(3)})$$

$$\mathbf{D}_{(m)} = \begin{bmatrix} \sigma^2_{(m),C} & 0 & 0 & 0 \\ 0 & \sigma^2_{(m),F} & 0 & 0 \\ 0 & 0 & \sigma^2_{(m),H} & 0 \\ 0 & 0 & 0 & \sigma^2_{(m),HF} \end{bmatrix}, \quad m = 2, 3$$

Where y_{ijk} is a binary response that equals one if the jth tree on the kth plot dies during the ith interval and zero otherwise, $\mathbf{x}_{ijk}$ are tree- and plot-level observations for an interval, and z is an indicator variable that equals one if the response belongs to the cultural treatment and zero otherwise (C = control, F = fertilizer, H = herbicide, and HF = herbicide and fertilizer). In addition, κ_i is the cut-point for the ith interval, $\mathbf{b}_{(2)jk}$ and $\mathbf{b}_{(3)k}$ are random effects for trees within plots and plots, respectively, and $\boldsymbol{\beta}$ are the fixed effects (parameters), where $i = 1, \ldots, n_{jk}, j = 1, \ldots, n_k$, $k = 1, \ldots, n$. Intervals are defined as $1, 2, \ldots, 13$, which

Table 2. The CAPPS study summary statistics ($N = 112365$) for dbh, tree height (H_t), trees per hectare (TPH), basal area per hectare (BA/ha), quadratic mean diameter (D_q), and relative spacing (RS) across the age range (1–13)

Attribute	Mean	Minimum	Maximum	Standard deviation
Dbh (cm)	8.45	0.00	33.78	6.49
H_t (m)	6.95	0.03	22.86	4.93
TPH/100	14.98	5.53	18.78	1.75
BA/ha (m^2)	13.06	0.00	46.41	12.35
D_q (cm)	8.64	0.00	24.08	6.29
RS	0.92	0.12	6.24	1.23

correspond to ages 1–2, 2–3, ..., 13–14, respectively. Random effects vectors $\mathbf{b}_{(2)jk}$ and $\mathbf{b}_{(3)k}$ are assumed multivariate normally distributed (each 4-dimensional) with mean zero and with variance–covariance matrix $\mathbf{D}_{(2)}$ at level 2 (trees within plots) and $\mathbf{D}_{(3)}$ at level 3 (plots), respectively. Off-diagonal elements of the random effects covariance matrices are assumed to equal zero, i.e., it is assumed that the tree-specific and plot-specific treatment effects are independent. Standardized time-dependent covariates are basal area per hectare (BA/ha), trees per hectare (TPH), relative spacing (RS), mean plot height (H^{Plot}), mean plot dbh (dbh$^{\text{Plot}}$), tree dbh (dbh), and tree height (H). Plot-level covariates were standardized by subtracting the mean plot value and dividing by the SD, whereas individual tree-level covariates were standardized for each plot. Our baseline (Model 12) includes 13 intervals, 39 treatment by interval interactions (F, H, and HF), 7 time-varying covariates, and 8 variance components.

A likelihood ratio test is typically used to determine the necessity of fixed and random effects for a mixed effects model that assumes the lowest level variance is normally distributed (Pinheiro and Bates 2000). However, for multi-level binary response models using quasi-likelihood to estimate the parameters, there is no true likelihood available. Hence, the likelihood ratio test is not available. Instead, the Wald test is preferred for testing fixed effects (Goldstein 1995). Our first hypothesis tested the age by treatment interaction using a joint Wald χ^2 test; the test statistic is 231.08 with 39 degrees of freedom (P-value < 0.0001). Because our joint test for the interactions is significant, we did not remove insignificant individual interaction terms. However, time-varying covariates were removed from our chosen model (Model12) using a stepwise procedure. Time-varying quadratic covariates BA/ha^2 and RS^2 were not significant. Interval, treatment by age interactions for intervals 1 and 2 (for the sake of brevity), time-varying covariates, and variance components for Model 12 are presented in Table 3.

Several diagnostic tools were used to assess Model 12 adequacy. Level 3 residual graphs and Q–Q plots revealed no evidence of serious heterogeneity or departures from normality. For example, plot level rank residuals and Q–Q plots by treatment illustrate no serious outliers or departures from the assumption of normality (Figure 3). Plots by treatment and rank (Figure 3), where plots by treatment are
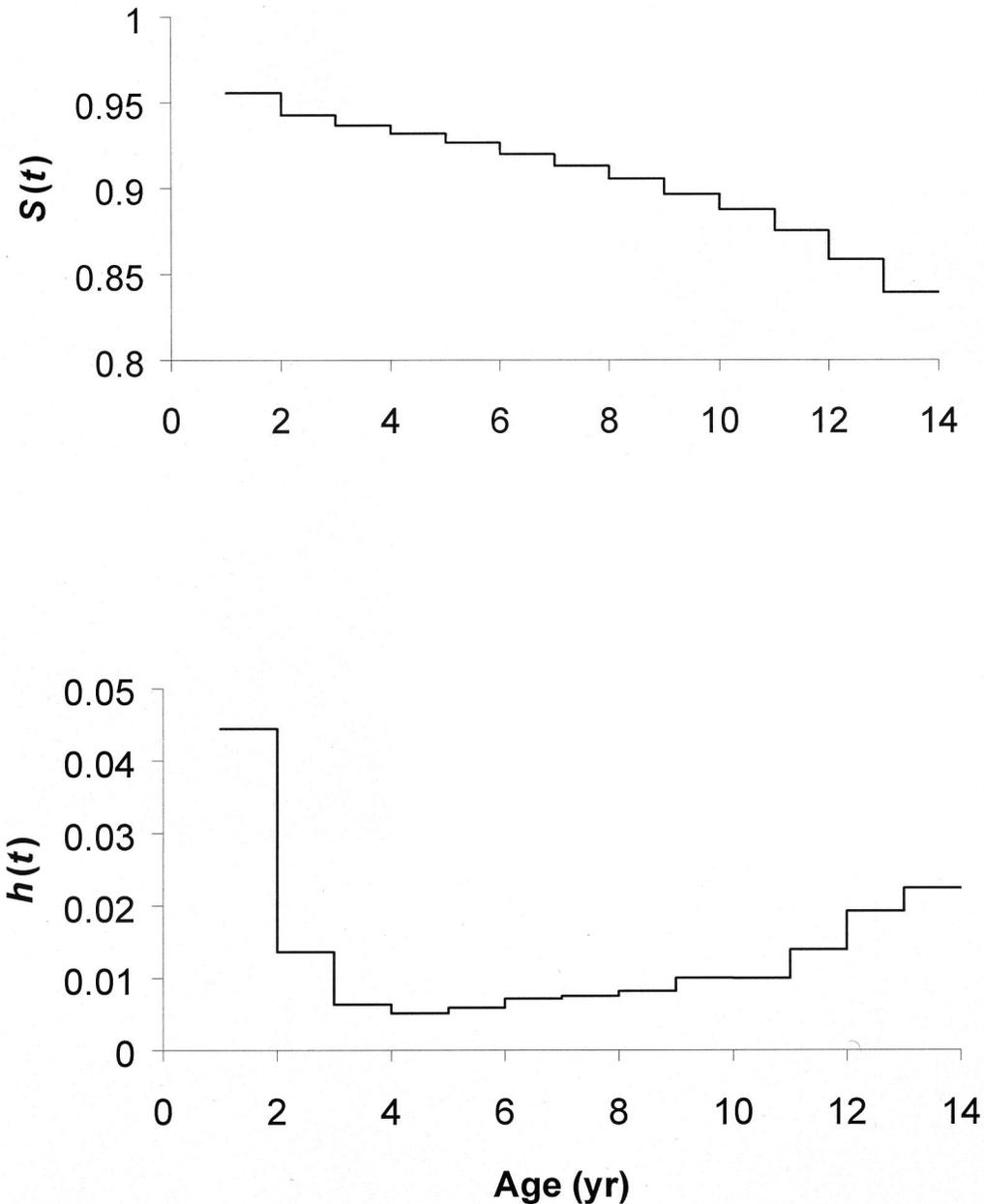
Figure 1. The Kaplan–Meier product limit survival (S(t)) and hazard functions (h(t)) estimates for the CAPPS study.

ranked from smallest to largest residual, illustrate that H and C have the least and most variability, respectively. We compared our mixed effects, or subject-specific (SS) Model 12 with the marginal raw mortality probabilities. There is strong agreement between the marginal and predicted probabilities (Figure 4). We compared several SS fit statistics with a population-averaged (PA) model. The PA model was obtained by re-fitting our final Model 12 excluding the random effects. We used the PA and SS models to compute the Pearson χ^2 statistic for survival by plot and age. The SS and PA models have 6 and 10 plots, respectively, which are significantly different from the actual stand tables. In addition, the SS model fits the stand tables better relative to the PA model for 115 of the 146 plots. Hence, it provides a

substantial improvement in fit for over three-quarters of the plots.

Parameters of Model 12 are interpreted as for a proportional hazards model. A covariate's effect on the hazard can be expressed as $-\log(1 - \pi_{ijk}) = e^{\eta_{ijk}}e^{\beta x_{ijk}}$, where the left-hand side of this equation is the cumulative hazard function for a given interval and the effect of the covariate on the hazard as a percentage is expressed as $100(e^{\beta x_{ijk}} - 1)$. For example, the estimated parameter for the F treatment and interval 1 interaction is 0.2921. This means, holding all other covariates constant, that there is a $100(e^{0.2921} - 1) \cong$ 33.92% increase in the hazard of mortality for a fertilized tree during interval 1 relative to the C treatment. Conversely, for the H treatment there is a $100(e^{-1.4053} - 1) \cong$
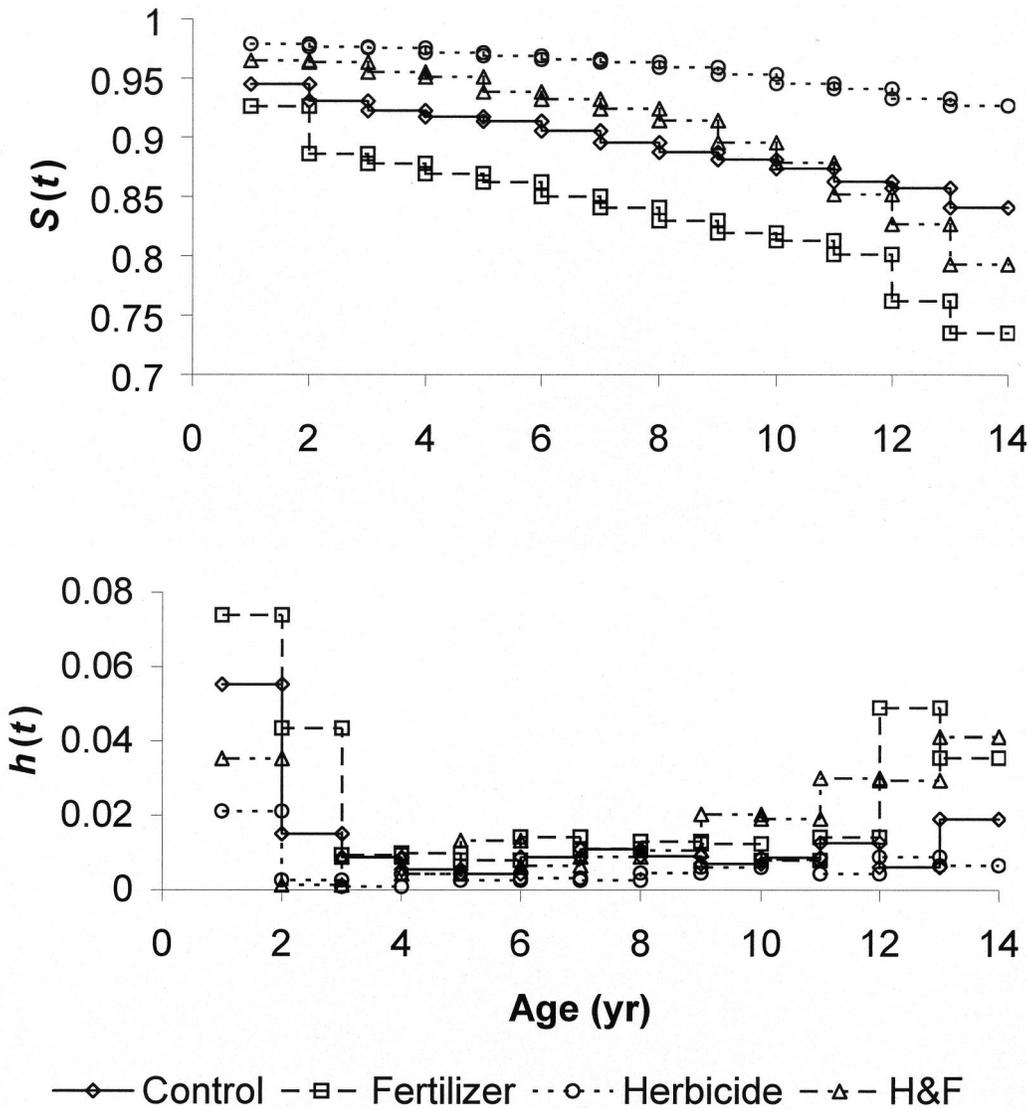
Figure 2. CAPPS study Kaplan–Meier (Kaplan and Meier 1958) product limit estimates for the survival (S(t)) and hazard functions (h(t)) by treatment.

75.47% reduction in the mortality risk for a herbicide tree during the first interval relative to treatment C. The HF treatment decreases the risk for an individual tree during the first interval by 69.75% relative to treatment C. Time-dependent covariates are interpreted similarly, i.e., if a plot is one standard deviation below the mean for BA/ha, then there is a $100(e^{-2.6893} - 1) \cong 93.21\%$ decrease in the mortality risk, holding all other covariates constant. Parameter estimates reveal that BA/ha and dbh have the largest and RS has the smallest impact on survival, respectively. In addition, note that the signs for the plot average and individual tree parameter estimates for H_t are negative. This means that, as a tree or plot increases in height relative to the mean, its hazard decreases. The estimated parameter for TPH is negative, which implies that, as TPH increases, the probability of mortality decreases. This seems counterintuitive, but our study plots were all planted at the same density, and plots that have more TPH as time increases are those

plots with a lower rate of mortality. Therefore, for our data it is logical that the TPH estimated parameter is negative.

Variance component parameter estimates of Model 12 are indicative of the survival variability for trees within a plot and plots of a given treatment (Table 3). The C treatment is the only treatment that has a larger estimated variance at the plot level than for trees within a plot. Treatments H and F have the least variability at the plot and trees within a plot levels, respectively. The H treatment has the largest level 2 to level 3 ratio of the variability, i.e., there is 15.4 times more variability at the tree level relative to the plot level.

Mortality Predictions

We can obtain individual tree mortality predictions at all levels and for new plots, assuming different resolutions of information are available. Here we demonstrate two scenarios for individual tree mortality predictions using a new

Table 3. Estimated fixed and random parameter estimates for the multilevel complementary log–log individual tree survival model. Only the first 2 of the 13 estimated parameters for treatment by interval[1] interaction are presented

Parameter	Estimate	Standard Error	P-value
Interval 1	−3.6298	0.3195	<0.0001
Interval 2	−4.9038	0.2692	<0.0001
Interval 3	−5.4412	0.2653	<0.0001
Interval 4	−5.3547	0.2880	<0.0001
Interval 5	−5.2564	0.3082	<0.0001
Interval 6	−4.6412	0.2527	<0.0001
Interval 7	−4.8289	0.2570	<0.0001
Interval 8	−5.0276	0.2763	<0.0001
Interval 9	−5.4391	0.3091	<0.0001
Interval 10	−5.3685	0.3062	<0.0001
Interval 11	−5.2006	0.3419	<0.0001
Interval 12	−5.9219	0.4515	<0.0001
Interval 13	−5.6449	0.4779	<0.0001
Fertilizer*I1	0.2921	0.2084	0.1610
Fertilizer*I2	0.1813	0.2416	0.4530
Herbicide*I1	−1.4053	0.2198	<0.0001
Herbicide*I2	−2.3563	0.3981	<0.0001
HF*I1	−1.1958	0.2712	<0.0001
HF*I2	−3.2856	0.6465	<0.0001
BA/ha	2.6893	0.1601	<0.0001
TPH	−1.0020	0.0558	<0.0001
dbh	−2.5903	0.1514	<0.0001
H(plot)	−1.7569	0.5005	0.0004
RS	−0.1505	0.0768	0.0500
dbh (plot)	1.3448	0.5954	0.0239
Height	−0.5418	0.1736	<0.0001
Variance components			
Level 2			
Control	0.5219	0.1294	
Fertilizer	0.4588	0.0985	
Herbicide	0.4930	0.2686	
HF	1.1454	0.2083	
Level 3			
Control	0.5872	0.1586	
Fertilizer	0.3903	0.1317	
Herbicide	0.0321	0.0513	
H&F	0.4063	0.1487	

[1] Intervals 1, 2, . . . , 13 correspond to ages 1–2, 2–3, . . . , 13–14, respectively.

plot, because it is common to make predictions for a plot not included in the original study.

Case I: Typical Response Mortality Predictions for a New Plot Associated with the HF Treatment Having No Prior Measurements

Suppose mortality predictions are desired for trees on a new plot that has no previous records but has the HF cultural treatment. Furthermore, suppose we are not interested in estimating the random effects but desire an estimate for the typical responses. Let the age 3 standard deviations for the mean of this plot for BA/ha, TPH, RS, H^{Plot}, and dbh^{Plot} be −0.25, −0.50, −2.0, −0.25, and −0.25, respectively. For simplicity, we will consider four trees on the plot that have standardized dbh and H_t values of −1.0, −0.5, 0.5, and 1.0. The estimated mortality probabilities for interval 3 (age 3 to 4) are obtained using Model 12 and setting

the random effects to zero, i.e.,

$$\log\{-\log(1 - \pi_{ijk})\} = \kappa_3 + \kappa_{3,HF} + \beta_1 BA/ha_{ijk}$$

$$+ \beta_2 TPH_{ijk} + \beta_3 RS_{ijk} + \beta_4 H^{Plot}_{ijk}$$

$$+ \beta_5 dbh^{Plot}_{ijk} + \beta_6 dbh_{ijk} + \beta_7 H_{ijk}.$$

Here all terms are defined previously. The substitution of plot attributes and their respective parameter estimates yields

$$\log\{-\log(1 - \hat{\pi}_{ijk})\}$$

$$= -5.2296 - 2.5903\, dbh_{ijk} - 0.5418\, H_{ijk}.$$

Mortality predictions are obtained by substituting the standardized dbh and H_t values for the four trees. The matrix of dbh and H_t values for these four trees is given by

$$\begin{bmatrix} -1.0 & -0.5 & 0.5 & 1.0 \\ -1.0 & -0.5 & 0.5 & 1.0 \end{bmatrix}^T,$$

where the first row represents dbh and the second row represents H_t. The mortality predictions for the four trees are 0.115527, 0.025315, 0.001118, and 0.000234. As expected, the probability of mortality decreases as the standardized variables increase. These mortality predictions are typical responses for a new plot that has the HF cultural treatment. However, we could estimate the plot-level random effect for this plot because it is associated with the HF cultural treatment.

Case II: Mortality Predictions Using the Estimated Plot Level Random Effect for a New Plot Associated with the HF Treatment Having No Prior Measurements

Suppose for Case I we want to estimate the plot-level random effect. Because Model 12 is nonlinear and has a binomial response, we present some background to motivate the method used to estimate the random effects. Goldstein's (1991) MQL-1 parameter estimation method is used to estimate the random effects, which is motivated by a generalized linear mixed model. A linear mixed-effects model can be expressed as

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \mathbf{z}\mathbf{b} + \boldsymbol{\epsilon}, \tag{13}$$

where $\mathbf{z}$ and $\mathbf{x}$ are the design matrices corresponding to the random and fixed effects, $\mathbf{b}$ and $\boldsymbol{\beta}$ are the respective parameters, and $\boldsymbol{\epsilon}$ is usually assumed to be normally distributed with mean zero. We can estimate the random effects of Equation 13 using the best linear unbiased predictor (BLUP) (Davidian and Giltinan 1995), which is often referred to as an empirical Bayes (EB) or shrinkage-type estimator and is defined as

$$\hat{\mathbf{b}} = \mathbf{D}\mathbf{Z}^T(\mathbf{Z}\mathbf{D}\mathbf{Z}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}), \tag{14}$$

where $\mathbf{D}$ and $\mathbf{R}$ are the variance–covariance matrices of the random effects and errors, respectively. Using Goldstein's (1991) method, we can write our multilevel binary response model as (Rodriguez and Goldman 1995)
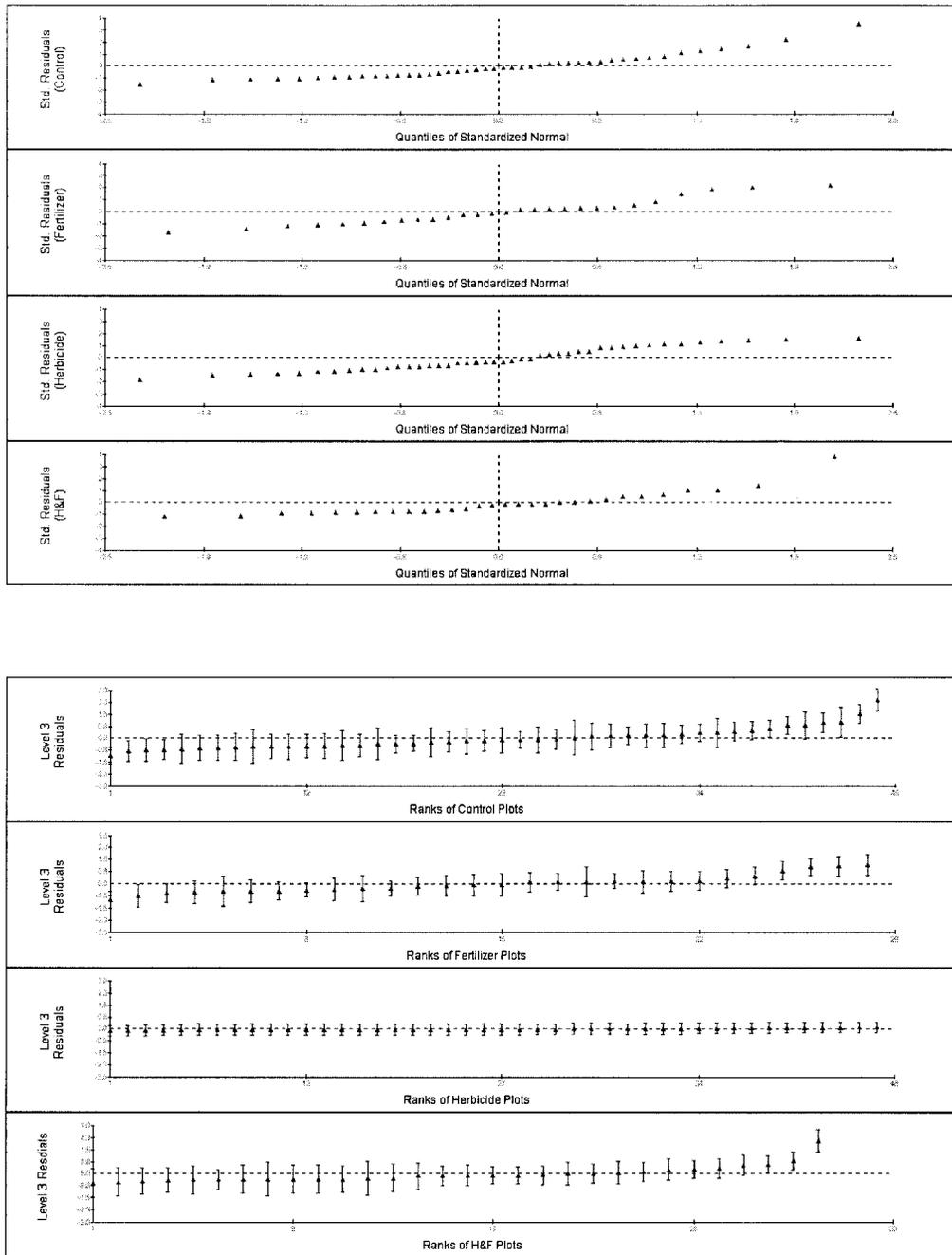
Figure 3. Standardized residuals by plot rank and quantiles of the standardized normal distribution for the plot-level random effects.

$$\mathbf{y} = \mathbf{g}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(2)}\mathbf{b}_{(2)} + \mathbf{Z}_{(3)}\mathbf{b}_{(3)}) + \boldsymbol{\epsilon}, \qquad (15)$$

where $\boldsymbol{\epsilon}$ has a mean of zero and a variance that depends on the mean $\mathbf{g}$. The inverse link $\mathbf{g}(\cdot)$ of Equation 15 is approximated using a first-order Taylor series expansion, around $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ and $\mathbf{b}_{(2)} = \mathbf{b}_{(3)} = \mathbf{0}$, as

$$\mathbf{y} \approx \mathbf{g}(\boldsymbol{\eta}_0) + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\eta}_0} \mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$$

$$+ \frac{\partial \mathbf{g}}{\partial \boldsymbol{\eta}_0} \mathbf{Z}_{(2)}\mathbf{b}_{(2)} + \frac{\partial \mathbf{g}}{\partial \boldsymbol{\eta}_0} \mathbf{Z}_{(3)}\mathbf{b}_{(3)} + \boldsymbol{\epsilon}, \qquad (16)$$

where $\partial \mathbf{g}/\partial \boldsymbol{\eta}_0$ is a diagonal matrix of derivatives of the mean with respect to the conditional linear predictor evaluated at $\boldsymbol{\eta} = \boldsymbol{\eta}_0$. For our CLL Model 12, the derivative (Λ) is

$$\frac{\partial g}{\partial \eta} = e^{\eta} e^{-e^{\eta}} \qquad (17)$$

The model that approximates the nonlinear Equation 15 can be expressed as

$$\mathbf{y}^* \approx \mathbf{X}^*\boldsymbol{\beta} + \mathbf{Z}_{(2)}^*\mathbf{b}_{(2)} + \mathbf{Z}_{(3)}^*\mathbf{b}_{(3)} + \boldsymbol{\epsilon}, \qquad (18)$$
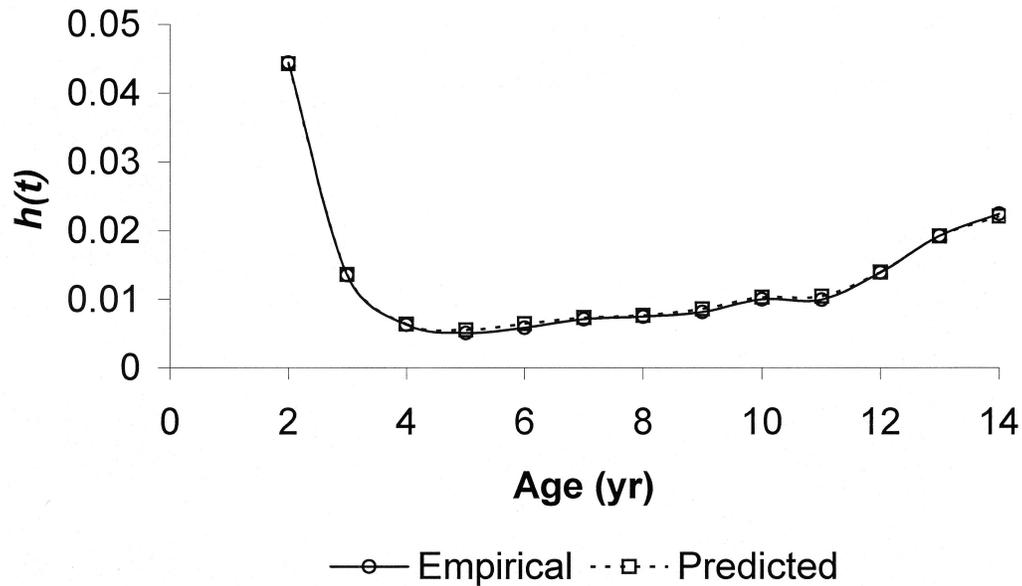
Figure 4. Predicted and empirical hazards by age for the CAPPS study data.

which has the same form as the multilevel linear equation (13); where $\mathbf{y}^* = \mathbf{y} - \mathbf{g}_0 + \mathbf{X}^*\boldsymbol{\beta}_0$, $\mathbf{X}^* = \boldsymbol{\Lambda}\mathbf{X}$, and $\mathbf{Z}^* = \boldsymbol{\Lambda}\mathbf{Z}$. Note that $\mathbf{E}(\mathbf{y}^*) = \mathbf{X}^*\boldsymbol{\beta}$ and $\boldsymbol{\Lambda}$ is an estimate of the variance. The BLUP of the random effects for Equation 18 at a given level is

$$\hat{\mathbf{b}}_{(i)} = \mathbf{D}_{(i)}\mathbf{Z}_{(i)}^{*\mathrm{T}}(\mathbf{Z}_{(2)}^*\mathbf{D}_{(2)}\mathbf{Z}_{(2)}^{*\mathrm{T}} + \mathbf{Z}_{(3)}^*\mathbf{D}_{(3)}\mathbf{Z}_{(3)}^{*\mathrm{T}}$$
$$+ \boldsymbol{\Lambda})^{-1}(\hat{\mathbf{y}}^* - \hat{\mathbf{X}}^*\boldsymbol{\beta}), \quad (19)$$

where $i = 2$ (tree level) or 3 (plot level) for our Model 12. At convergence, $\boldsymbol{\eta} = \boldsymbol{\eta}_0$, hence we can write the BLUP for the plot-level random effects as

$$\hat{\mathbf{b}}_{(3)} = \mathbf{D}_{(3)}\mathbf{Z}_{(3)}^*(\mathbf{Z}_{(2)}^*\mathbf{D}_{(2)}\mathbf{Z}_{(2)}^{*\mathrm{T}} + \mathbf{Z}_{(3)}^*\mathbf{D}_{(3)}\mathbf{Z}_{(3)}^{*\mathrm{T}} + \boldsymbol{\Lambda})^{-1}(\mathbf{y} - \hat{\mathbf{g}}).$$
$$(20)$$

Continuing with our example, using the CLL values from Case I, the matrix of evaluated differentials is

$$\boldsymbol{\Lambda} = \begin{bmatrix} 0.108571 & 0 & 0 & 0 \\ 0 & 0.024991 & 0 & 0 \\ 0 & 0 & 0.001117 & 0 \\ 0 & 0 & 0 & 0.000234 \end{bmatrix}.$$

To simplify the computations for our example and because the covariances of the plot level variance components are zero for Model 12, we can let $\mathbf{D} = 0.4063$, which is the estimated variance component for the plot-level HF treatment. In addition, $\mathbf{Z}$ is a vector of ones corresponding to the four trees and let $\mathbf{V} = \mathbf{Z}_{(2)}^*\mathbf{D}_{(2)}\mathbf{Z}_{(2)}^{*\mathrm{T}} + \mathbf{Z}_{(3)}^*\mathbf{D}(3)\mathbf{Z}_{(3)}^{*\mathrm{T}} + \boldsymbol{\Lambda}$. The level 2 and 3 design matrices are identical for our Model 12,

hence

$$\mathbf{Z}_{(2)}^* = \mathbf{Z}_{(3)}^* = \boldsymbol{\Lambda}\mathbf{Z}_{(3)}$$

$$= \begin{bmatrix} 0.108571 & 0 & 0 & 0 \\ 0 & 0.024991 & 0 & 0 \\ 0 & 0 & 0.001117 & 0 \\ 0 & 0 & 0 & 0.000234 \end{bmatrix}$$

$$\cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.108571 \\ 0.024991 \\ 0.001117 \\ 0.000234 \end{bmatrix}.$$

Then estimate $\mathbf{V}$ using

$$\mathbf{Z}_{(2)}^*\mathbf{D}_{(2)}\mathbf{Z}_{(2)}^{*\mathrm{T}} =$$

$$\begin{bmatrix} 0.0135016 & 0.0031078 & 0.0001389 & 0.0000291 \\ 0.0031078 & 0.0007154 & 0.0000320 & 6.6982E-6 \\ 0.0001389 & 0.0000320 & 1.4291E-6 & 2.9938E-7 \\ 0.0000291 & 6.6982E-6 & 2.9938E-7 & 6.2718E-8 \end{bmatrix}$$

and

$$\mathbf{Z}_{(3)}^*\mathbf{D}_{(3)}\mathbf{Z}_{(3)}^{*\mathrm{T}} =$$

$$\begin{bmatrix} 0.0047893 & 0.0011024 & 0.0000493 & 0.0000103 \\ 0.0011024 & 0.0002538 & 0.0000113 & 2.3760E-6 \\ 0.0000493 & 0.0000113 & 5.0694E-7 & 1.0620E-7 \\ 0.0000103 & 2.376E-6 & 1.0620E-7 & 2.2247E-8 \end{bmatrix}.$$

Hence,

$$\mathbf{V} = \mathbf{Z}_{(2)}^*\mathbf{D}_{(2)}\mathbf{Z}_{(2)}^{*\mathrm{T}} + \mathbf{Z}_{(3)}^*\mathbf{D}_{(3)}\mathbf{Z}_{(3)}^{*\mathrm{T}} + \boldsymbol{\Lambda} =$$

$$\begin{bmatrix} 0.1268619 & 0.0042102 & 0.0001882 & 0.0000394 \\ 0.0042102 & 0.0259601 & 0.0000433 & 9.0742E-6 \\ 0.0001882 & 0.0000433 & 0.0011189 & 4.0558E-7 \\ 0.0000394 & 9.0742E-6 & 4.0558E-7 & 0.0002341 \end{bmatrix}.$$

Now supposing that tree 2 had recently died, our corresponding vector of residuals is then defined as

$$\mathbf{y} - \hat{\mathbf{g}} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.115527 \\ 0.025315 \\ 0.001118 \\ 0.000234 \end{bmatrix} = \begin{bmatrix} -0.115527 \\ 0.974685 \\ -0.001118 \\ -0.000234 \end{bmatrix}.$$

Thus, the plot-level random effect using Equation 20 is

$$\hat{b}_{(3)} = [0.4063] \begin{bmatrix} 0.108571 \\ 0.024991 \\ 0.001117 \\ 0.000234 \end{bmatrix}^{\mathrm{T}} V^{-1} \begin{bmatrix} -0.115527 \\ 0.974685 \\ -0.001118 \\ -0.000234 \end{bmatrix}$$

$$= 0.2881946.$$

Now we use the predicted random effect for this plot and add it to the linear predictor obtained for Case I. Hence, the age 3-to-4 mortality probabilities for the three living trees is estimated using

$$\log\{-\log(1 - \pi_{ijk})\} = -5.2296 - 2.5903\, \text{dbh}_{ijk}$$

$$-0.5418\, \text{H}_{ijk} + 0.2881946.$$

Predicted mortality probabilities for the three surviving trees are 0.151061, 0.001491, and 0.000312. Including the plot-level random effect for Model 12 increased the predicted probability of mortality, which is expected because we are using only four trees in our example and assumed one tree died. It may be of interest to assess the plot effect in terms of the mortality risk. This plot has about a $100(e^{0.2881946} - 1) = 33.40\%$ increase in the risk of mortality relative to the typical HF treatment plot in which the random effect is zero. Here we focused on estimating a plot-level random effect for Model 12. However, the same method is used to estimate tree-level random effects if we have prior information for trees within a plot.

Discussion

Permanent-plot forest inventory data naturally have a hierarchical structure, i.e., measurement occasions (repeated measurements) are nested within trees and trees are nested within plots. Consideration of the multilevel structure allows the analyst to obtain statistically efficient estimates of the regression coefficients, and incorporating the heterogeneity from the groups provides the correct standard errors, confidence intervals, and tests of significance (Rodriguez and Goldman 1995). Conceptually, our multilevel survival Model 12 is straightforward. However, multilevel binary response models present formidable challenges to parameter estimation. This is due to the necessity of integrating out the random effects in a nonlinear, non-normal model. There have been several methods proposed for multilevel binary response model parameter estimation. These parameter estimation methods include maximum marginal likelihood (MML), first- and second-order marginal quasi-likelihood (MQL-1 and MQL-2, respectively), and first- and second-order penalized quasi-likelihood (PQL-1 and PQL-2, re-

spectively). Currently, there is no consensus on which parameter estimation method is preferable under a given set of conditions. However, it is clear from the literature that some of these methods are deficient in certain respects and that some methods are better than others. Unfortunately, the best methods are not currently feasible computationally for our data set.

Our Model 12 considered and estimated the variance components by treatment at the plot and trees-within-a-plot levels. There may be questions about our failure to consider or model the correlation over time within a tree. It is well-known that ignoring serial correlation may result in underestimating the parameter standard errors. However, we modeled a nonrepeatable binary response. Hence there is no theoretical reason to consider the correlation over time for our repeated measurements. The likelihood function Equation 5 was developed using the conditional probabilities, i.e., the probability of mortality during the ith interval is conditional on mortality not occurring during the previous intervals. Hence, if the jth tree on the kth plot succumbs during the second interval, we can factor the likelihood function for this tree as $(1 - \pi_{1jk})\pi_{2jk}$. Therefore, each of these terms for this tree may be treated as though it came from a distinct independent observation (Allison 1995).

Our Model 12 demonstrates the importance of considering the heterogeneity that may occur at different levels and, for our data, most treatments have significant survival variability at the plot and trees-within-a-plot levels. Differences among the treatment variance components can be used to assess the impact of the treatment on survival. For example, fertilization and herbicide usually accelerate stand development, and the impact of these treatments on stand development and survival can be obtained using our Model 12. We estimated the ages that each treatment obtained its respective mean BA/ha, which are C = age 8–9, F = age 6–7, H = age 6–7, and HF = age 5–6. Hence, the HF treatment obtains the mean BA/ha first and it occurs about 5 years 7 months. Comparing the other treatments to the HF at this age, we can determine roughly how many standard deviations from the mean the other treatments are at this time. Calculating the standard deviations for these treatments and then obtaining the corresponding relative risks, we find that the C, F, and H treatments result in a 52.95, 41.96, and 22.46% decrease in the hazard relative to the HF treatment, holding all other variables constant. Corresponding mean BA/ha for the C, F, and H treatments are 71.4, 51.7, and 24.4% less than the BA/ha of the HF treatment (13.1 m²). It is important to note that, when the HF treatment achieved the mean BA/ha, its survival was substantially greater than the C and F treatments.

Our estimated variance components for Model 12 suggest that the F, H, and HF treatments will reduce the plot-level survival variability relative to the C treatment. Moreover, the estimated variance components at the tree level reveal that the survival variability for the C, F, and H treatments are not substantially different. Estimated variance components for the HF treatment reveal that it has substantially more survival variability at the tree level.

Estimated variance components are larger for the F, H, and HF treatments at the tree level, and the converse is true for the C treatment.

Conclusion

We demonstrated through model formulation that the CLL link function is the natural choice for permanent-plot binary response data because it is derived directly from the likelihood function, which accounts for interval censoring. Moreover, parameters can be interpreted as for a Cox proportional hazards model. In addition, we demonstrated that our Model 12 can relax the assumption of proportional hazards at the cut points and include random effects at each level.

Multilevel models have become increasingly common in forestry, which is likely to continue as more efficient software becomes available for estimation of large complex data sets. It is common for multilevel binary response models to use first- and second-order MQL and PQL to estimate the parameters. It has been suggested that MQL-1 tends to underestimate the variance components and that PQL-2 is the preferred method. However, in many instances, especially for large complex data sets such as ours, the only method that may converge is MQL-1. Moreover, it is usually better to allow for a multilevel structure and use MQL-1 than to ignore the multilevel structure (Rodriguez and Goldman 1995). Additional studies are necessary to assess the effects of different parameter estimation techniques on the fixed and random effects.

We demonstrated that plot- and tree-level random effects generally result in more precise subject-specific predictions. However, if the goal is to obtain the marginal probabilities, then these can be computed. Hence, the multilevel model is theoretically correct for obtaining the marginal probabilities, but the converse is not true for the PA model. If fixed-effects hypothesis testing is the study focus, then the SS and PA approaches may result in similar inferences. However, the PA approach will not provide any information about the heterogeneity that may exist at the different levels. If predictions are our primary purpose, then the SS approach is usually preferred (Ten Have and Uttal 1994). Individual tree plantation survival models naturally focus on prediction. Therefore, the SS approach is the more natural method.

Literature Cited

ALLISON, P.D. 1995. Survival analysis using the SAS system: A practical guide. SAS Institute Inc., Cary, NC. 292 p.

ALLISON, P.D. 1999. Logistic regression using the SAS system. SAS Institute Inc., Cary, NC. 288 p.

BARBOSA, M.F., AND H. GOLDSTEIN. 2000. Discrete response multilevel models for repeated measures; an application to voting intentions data. Quality and Quantity 34(3):323–330.

BIGGERI, L., M. BINI, AND L. GRILLI. 2001. The transition from university to work: A multilevel approach to the analysis of the time to obtain the first job. J. R. Statist. Soc. A. 293–305.

BRESLOW, N.R., AND D.G. CLAYTON. 1993. Approximate inference in generalized linear mixed models. JASA. 88:9–25.

BUFORD, M.A., AND W.L. HAFLEY. 1985. Modeling the probability of individual tree mortality. For. Sci. 31:331–341.

COLLET, D. 1994. Modelling survival data in medical research. Chapman & Hall, London. 347 p.

DAVIDIAN, M., AND D.M. GILTINAN. 1995. Nonlinear models for repeated measurement data. Chapman and Hall, New York. 359 p.

DOBBERTIN, M., AND G.S. BIGING. 1998. Using the non-parametric classifier CART to model forest tree mortality. For. Sci. 44(4):507–516.

FANG, Z., AND R.L. BAILEY. 2001. Nonlinear mixed effects modeling for slash pine dominant height growth following intensive silvicultural treatments. For. Sci. 47(3):287–300.

GIBBONS, R.D., AND D. HEDEKER. 1997. Random effects probit and logistic regression models for three-level data. Biometrics 53:1527–1537.

GOLDSTEIN, H. 1991. Nonlinear multilevel models, with an application to discrete response data. Biometrika 78:45–51.

GOLDSTEIN, H. 1995. Multilevel statistical models, 2nd ed. Halstead Press, New York. 178 p.

GREGOIRE, T.G., O. SCHABENBERGER, AND J.P. BARRET. 1995. Linear modeling of irregularly spaced, unbalanced, longitudinal data from permanent-plot measurements. Can. J. For. Res. 25:137–156.

GUAN, B.T., AND G.Z. GERTNER. 1991. Modeling red pine tree survival with an artificial neural network. For. Sci. 37(5):1429–1440.

HALL, D.B., AND R.L. BAILEY. 2001. Modeling and prediction of forest growth variables based on multilevel nonlinear mixed models. For. Sci. 47(3):311–321.

HAMILTON, D.A. 1974. Event probabilities estimated by regression. USDA For. Serv. Intermt. Res. Stn. Res. Pap. INT-152. 18 p.

HAMILTON, D.A. 1986. A logistic model of mortality in thinned and unthinned mixed conifer stands of northern Idaho. For. Sci. 32:989–1000.

HEDEKER, D., O. SIDDIQUI, AND F.B. HU. 2000. Random-effects regression analysis of correlated grouped-time survival data. Statistical Methods in Medical Research 9:161–179.

HUEBSCHMANN, M.M., L.R. GERING, T.B. LYNCH, O. BITOKI, AND P.A. MURPHY. 2000. An individual tree growth and yield prediction system for uneven-aged shortleaf pine stands. SJAF. 24(2):112–120.

KAPLAN, E.L., AND P. MEIER. 1958. Nonparametric estimation from incomplete observations. JASA. 53:457–481.

KOBE, R.K., AND K.D. COATES. 1997. Models of sapling mortality as a function of growth to characterize interspecific variation in shade tolerance of eight tree species of northwestern British Columbia. Can. J. For. Res. 27:227–236.

MONSERUD, R.A. 1976. Simulation of forest tree mortality. For. Sci. 22:438–444.

MONSERUD, R.A., AND H. STERBA. 1999. Modeling individual tree mortality for Austrian forest species. For. Ecol. Manage. 113:109–123.

MCCULLAGH, P. 1980. Regression models for ordinal data. J. R. Statist. Soc. Series B. 42:109–142.

PINHEIRO, J.C., AND D.M. BATES. 2000. Mixed-effects models in S and S-PLUS. Springer, New York. 528 p.

PREISLER, H.K., AND G.W. SLAUGHTER. 1997. A stochastic model for tree survival in stands affected by annosum root disease. For. Sci. 43(1):78–85.

RASBASH, J., W. BROWNE, H. GOLDSTEIN, M. YANG, I. PLEWIS, M. HEALY, G. WOODHOUSE, AND D. DRAPER. 1999. A user's guide to MlwiN, 2nd ed. Institute of Education, London. 278 p.

RODRIGUEZ, G., AND N. GOLDMAN. 1995. An assessment of estimation procedures for multilevel models with binary responses. J. R. Statist. Soc. A. 158(1):73–89.

RODRIGUEZ, G., AND N. GOLDMAN. 2001. Improved estimation procedures for multilevel models with binary response: A case study. J. R. Statist. Soc. A. 339–355.

SHEN, G., J.A. MOORE, AND C.R. HATCH. 2001. The effect of nitrogen fertilization, rock type, and habitat type on individual tree mortality. For. Sci. 47(2):203–213.

TEN HAVE, T.R., AND D.H. UTTAL. 1994. Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival profiles. Appl. Statist. 43(2):371–384.

VANCLAY, J.K. 1991. Mortality functions for north Queensland rain forests. J. Tropical For. Sci. 4(1):15–36.

VOLNEY, W.J.A. 1998. Ten-year tree mortality following a jack pine budworm outbreak in Saskatchewan. Can. J. For. Res. 28(12):1784–1793.

WYKOFF, W.R., N.L. CROOKSTON, AND A.R. STAGE. 1982. User's guide for the stand prognosis model—Version 5.0. For. Serv. Handbook 2409.21, Chapt. 300. USDA For. Serv. Region One, Missoula, Montana. 113 p.

YANG, M., AND H. GOLDSTEIN. 2000. Multilevel models for repeated binary outcomes: Attitudes and voting over the electoral cycle. J. R. Statist. Soc. A. 163(1):49–62.

YAO, X., S.J. TITUS, AND S.E. MACDONALD. 2001. A generalized logistic model of individual tree mortality for aspen, white spruce, and lodgepole pine in Alberta mixedwood forests. Can. J. For. Res. 31:283–291.