Published in final edited form as:

Stat Med. 2009 April 15; 28(8): 1301–1314. doi:10.1002/sim.3543.

A Finite Mixture Survival Model to Characterize Risk Groups of Neuroblastoma

Sally Hunsberger,

Biometric Research Branch, National Cancer Institute, 6130 Executive Boulevard, Rm 8120, Rockville MD, 20852

Paul S. Albert, and

Biometric Research Branch, National Cancer Institute, 6130 Executive Boulevard, Rm 8120, Rockville MD, 20852

Wendy B. London

Research Associate Professor & Assoc Program Director, Children's Oncology Group (COG), University of Florida, 104 N. Main St, #600, Gainesville, FL 32601

Sally Hunsberger: sallyh@ctep.nci.nih.gov; Paul S. Albert: albertp@ctep.nci.nih.gov; Wendy B. London: wendy@cog.ufl.edu

SUMMARY

Neuroblastoma is a childhood cancer with patients experiencing heterogeneous survival outcomes despite aggressive treatment. Disease outcomes range from early death to spontaneous regression of the tumor followed by cure. Due to this heterogeneity, it is of interest to identify patients with similar types of neuroblastoma so that specific types of treatment can be developed. Oncologists are especially interested in identifying patients who will be cured so that the minimum amount of a potentially toxic treatment can be given to this group of patients. We analyze a large cohort of neuroblastoma patients and develop a finite mixture model that uses covariates to predict the probability of being in a cure group or other (one or more) risk groups. A prediction method is developed that uses the estimated probabilities to assign a patient to different risk groups. The robustness of the model and the prediction method is examined via simulation by looking at misclassification rates under misspecified models.

Keywords

parametric cure model; classification

1. Introduction

Neuroblastoma is a childhood cancer that accounts for approximately 15% of all pediatric oncology deaths [1]. In neuroblastoma, malignant (cancer) cells form in nerve tissue of the adrenal gland, neck, chest, or spinal cord. Patients with this cancer have heterogeneous outcomes, suggesting there may be distinct, unknown subgroups of neuroblastoma. In a Children's Oncology Group (COG) study [2] some children's tumors spontaneously regressed and the children were disease free 12 years after diagnosis, while other children's tumors progressed within the first year of diagnosis despite aggressive treatment. Because of

the disease heterogeneity it is of interest to identify patients with similar types of neuroblastoma so that specific treatment strategies can be developed. Current treatment strategies include any or all of the following; watchful waiting, surgery, mild to aggressive chemotherapy, radiation and bone marrow transplants. Since neuroblastoma is a disease that most often strikes young children, treating patients with aggressive therapy is a concern because of the potential for long term health implications (from heart disease to second malignancies). Some children with neuroblastoma can be cured, and for these children, oncologists must try to give the minimum treatment possible while preserving cure. Although a fraction of patients are cured with current treatments, approximately 40% [1] will die of disease; for these patients improved treatment options are imperative. Identifying similar subgroups of neuroblastoma patients is an initial step in developing specific treatment strategies.

Various authors have tried to explain the heterogeneity in prognosis with multiple factors including age at diagnosis, tumor stage, and biologic variables. Age at diagnosis has been identified by many as an important risk stratification covariate with younger patients at lower risk for poor outcome and older patients at higher risk for poor outcome [2–6]. Tumor stage (International Neuroblastoma Staging System, (INSS)) [7] is a score usually based on the size of the tumor, the amount of tumor that has been removed during surgery, whether lymph nodes contain cancer, and whether the cancer has spread from the original site to other parts of the body. Stage 1 patients have the best prognosis and stage 4 patients have the worst prognosis. A biologic variable *MYCN* status [8, 9] refers to the number of copies of the N-myc gene and poor outcome is associated with amplification. In this paper we focus on using age at diagnosis, tumor stage and *MYCN* status to identify and distinguish between different subgroups of neuroblastoma. Since oncologists are particularly interested in identifying what types of patients have a high likelihood of being cured, we focus on a prediction method that will be able to predict cured patients with high probability. We analyze a large cohort of neuroblastoma patients and develop a model that uses the covariates to predict patients in the cure group and patients in risk groups with poor prognosis.

The organization of the paper is as follows. Section 2 reviews the literature on modeling survival data in which cure is an outcome and introduces our model. Section 3 presents the data set analysis and model selection. Section 4 describes a prediction method for assigning patients to a risk group and Section 5 examines the robustness of the prediction method. We close with a discussion in Section 6.

2. LITERATURE REVIEW AND MODEL

Many authors have studied cure mixture models with survival data. Work with parametric models of populations with a cured fraction began with [10] and [11] with many others building on this work. Some authors use parametric mixture models based on the standard failure time densities (log-normal, exponential, Weibull, Gompertz, etc.) to model a cured proportion and the death rate [12–16].

[14], [17], and [18] extend the Cox proportional hazards regression model to allow for a cure component, however, their models only allow for 2 groups, a cured group and a group that is not cured. [19] use a parametric cure model to analyze a data set where the patients who are not cured are modeled according to a Weibull distribution. Our approach is similar to that of Greenhouse and Silliman, but we potentially allow for more than two mixture components corresponding to more than two types of neuroblastoma. As in Greenhouse and Silliman we use a Weibull distribution to model the survival in the non-cured groups. The Weibull distribution is described by two parameters, thus allowing more flexibility than an exponential. The Weibull can describe survival curves that drop off rapidly or have very long slow decreases in survival. We propose a maximum-likelihood approach to estimating the number of mixture components. Alternative Bayesian approaches are also possible. In particular, Reversible-jump Monte-Carlo methods have been proposed for this problem [20]. Alternatively, a Markov-Chain Monte-Carlo (MCMC) method which treats the parameters of the model as a marked Poisson process have been proposed [21].

We now introduce our notation and models. Let y_i be the true failure time for individual i , let c_i be an individual's potential censoring time and d_i be the corresponding censoring indicator. Observations then consist of $t_i = \min(y_i, c_i)$ with $d_i = 1$ if $y_i < c_i$ (uncensored) and $d_i = 0$ if $y_i > c_i$ (censored).

We introduce a finite mixture model with m groups where the probability of being in a particular group $j = 1, 2, \dots, m$ can be allowed to depend on subject-specific covariates where x_{ik} denotes covariates $k = 1, 2, \dots, K$ measured on subject i . The following polychotomous logistic regression parameterization is used

$$\theta_{ij} = \theta_j(\beta_j, \mathbf{x}_i) = \exp(\beta_{j0} + \sum_{k=1}^K x_{ik}\beta_{jk}) / (1 + \sum_{j=2}^m \exp(\beta_{j0} + \sum_{k=1}^K x_{ik}\beta_{jk})) \quad (1)$$

with $\theta_{i1} = 1 - \sum_{j=2}^m \theta_{ij}$, $\beta_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jK})'$ and $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iK})'$. The vector β_j contains parameters which characterize the effect of covariates $\mathbf{x}_i$ on the probability of being in the j^{th} group.

The survival function for the i^{th} patient is

$$S(t|\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{p}) = \theta_{i1}F_1(t|\lambda_1, p_1) + \theta_{i2}F_2(t|\lambda_2, p_2) + \dots + \theta_{im}F_m(t|\lambda_m, p_m), \quad (2)$$

where $F_j(t|\lambda_j, p_j)$ is a Weibull survival distribution indexed by parameters λ_j, p_j also $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_m)'$ and $\mathbf{p} = (p_1, p_2, \dots, p_m)'$. Alternatively, a cure component can replace a Weibull distribution in this formulation. For example a cure component can replace the Weibull distribution for group 1 by replacing $F_1(t|\lambda_1, p_1)$ with $F_1(t) = 1$ in equation 2. Models are identifiable up to a permutation of the group labels. Thus, we order the groups in terms of the median survival in each group.

We use maximum likelihood to estimate the model parameters. As an example, the likelihood, $L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{p})$, for the model with three groups, two Weibull and a cure component is

$$L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{p}) = \prod_{i=1}^n [\theta_{i2} f_2(t_i | \lambda_2, p_2) + \theta_{i3} f_3(t_i | \lambda_3, p_3)]^{d_i} [\theta_{i1} + \theta_{i2} F_2(t_i | \lambda_2, p_2) + \theta_{i3} F_3(t_i | \lambda_3, p_3)]^{(1-d_i)}$$

where $f_2(t|\lambda_2, p_2)$ and $f_3(t|\lambda_3, p_3)$ are different Weibull distributions and $F_2(t|\lambda_2, p_2)$ and $F_3(t|\lambda_3, p_3)$ are the respective cumulative distribution functions. The maximum likelihood estimates are found by maximizing $\log(L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{p}))$ with respect to the set of unknown parameters. The parameter estimates are obtained using a quasi Newton-Raphson algorithm in the software package gauss [22]. Hypothesis testing for assessing the effect of covariates on the probability of being in a particular group is conducted with likelihood ratio tests.

3. DATA ANALYSIS AND MODEL SELECTION

The goal of this paper is to develop a method to predict who will fall into different neuroblastoma groups or types based on age at diagnosis, INSS stage, and *MYCN* status (unamplified versus amplified) using data from the COG. INSS stage is a score of 1, 2A, 2B, 3, 4, or 4S. The 4S patients are similar to stage 4 patients in that they have disseminated tumor, where the dissemination is limited to the skin, liver, or bone marrow. The 4S children are also younger than 1 year and the stage 4 children are typically older. Observationally, the 4S patient's outcomes are much better than stage 4 patients and are most similar to stage 2A or 2B. Therefore, in this analysis (as is typical) we group stage 2A, 2B and 4S patients into the category of Stage 2. The data set includes all patients consented and enrolled in biology and therapeutic studies from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG) (now the groups are combined into the COG) from 19862001. Despite differences in treatments across studies, this data set has been used by the COG to determine which patients should go onto low risk, intermediate risk and high risk studies [2]. The studies enrolled 3666 patients. After excluding patients who did not have data on age, stage or *MYCN* status the data from 2558 patients were used for this analysis.

The primary endpoint used to assess outcome in neuroblastoma is event-free survival (EFS) time, which is defined as the time from study enrollment until the first occurrence of relapse, progression, second malignancy, or death. The median follow-up was 5.74 years (using the reverse Kaplan-Meier estimate of [23]). In the data set, 30% of the patients were followed for greater than six years and of the 30% (714 patients) only seven patients had events after six years of follow-up. This censoring pattern was consistent across all subgroups. The Kaplan-Meier estimate of EFS at seven years was 61% (the median EFS could not be estimated). Figure 1 shows the cumulative distribution of patients across ages by stage and *MYCN* status. The dashed lines designate *MYCN* unamplified and the solid line designates *MYCN* amplified. It can be seen that a majority of the patients are *MYCN* unamplified, with the majority of *MYCN* amplified patients being stage 4. Over 50% of all patients are less than two years of age.

Our interests are two-fold. We are interested in determining how many groups are supported by the data and also in whether a cure group should be included in the model. We consider models with one to three groups since from a clinical perspective identifying one, two, or three groups of patients corresponds to how current patients are treated and how new

treatments are developed. We also consider models with and without cure. Although it can be difficult to distinguish between a cure group and survival functions with long tails [24], [25] showed that the cure fraction could be estimated well if follow-up is longer than the median survival. In this data set, the median follow-up is 5.74 years, only 9 of the 894 events (1%) occur after 5.74 years, and 13% of patients had longer follow-up than the last event which occurred at 7.8 years. The longest follow-up was 12 years.

When considering a model with two or three groups we use equations (1) and (2) to allow the covariates to influence the probability of group membership. A model with one group is also considered. This model includes covariates in the parameterization of the hazard so that

$$S(t|\mathbf{x}_i, \boldsymbol{\eta}, p) = \exp(-(\lambda t)^p), \quad (3)$$

where $\boldsymbol{\eta} = (\eta_0, \eta_1, \dots, \eta_K)'$ and $\lambda = \lambda(\boldsymbol{\eta}, \mathbf{x}_i) = \exp(\eta_0 + \sum_{k=1}^K x_{ik} \eta_k)$. The covariates we examine are age, stage, and *MCYN* status. Each model includes age as a continuous covariate, *MCYN* status as an indicator variable, and three indicator variables for INSS stage with stage 1 being the baseline stage. As suggested by [19], we use the Bayesian Information Criteria (BIC) to select the model that best fits the data. The BIC is a penalized likelihood technique in which smaller BIC values indicate a better fit of the data. The BIC is defined as $\text{BIC} = -2 \log(L(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{p})) + \gamma \log(n)$ where n is the number of observations and γ is the number of model parameters.

Table 1 gives the BIC values for different mixture models fit to the neuroblastoma data. The number of estimated parameters is also given in the table, showing how the complexity of the models increases as groups are added to the model. In the table, we give the number of Weibulls (abbreviating Weibull with W) and we denote a model with + cure if there is a cure component. There is definite evidence of heterogeneity or different types of neuroblastoma since the BIC for the single Weibull model (1W) is much larger than for all other models. There is also support for including a cure component in the model as can be seen by comparing BIC's for models with the same number of groups where one model contains a cure component and the other doesn't. The model with the cure component always has the smallest BIC (see for example 2W+cure BIC=4248.6503 and 3W BIC=4261.888). It appears that the 2W+cure model fits the data the best and so we will explore this model further.

Although we use the BIC to select the model, various authors [26] have raised concerns about the theoretical justification of this approach with mixture models. Therefore, we performed three simulations to examine how well the BIC performed when selecting the number of groups in this mixture model setting. In each simulation we generated data under a model with different numbers of groups (the data were generated with parameter values estimated from the neuroblastoma data when one, two, and three groups were assumed). For each simulation we fit models with one, two, and three groups to the data and used the BIC to select the model with the best fit. The two and three group models included a cure group and were parameterized as in (2) and (1), while the one group model followed (3). In each simulation 10,000 data sets were generated and in each case the BIC always choose the

correct model. Thus, the results of this simulation study support the use of BIC for model selection for the neuroblastoma analysis.

We used likelihood ratio tests to examine the significance of the covariates in predicting group membership in the 2W + cure model. The test for each covariate was a simultaneous test of the parameters associated with the covariate in each θ_{ik} . Tests for all covariates were all highly significant ($p < .001$), indicating age, MYCN, and INSS stage are all important in discriminating between types of neuroblastoma.

We computed point estimates and 95% bootstrap confidence intervals (using the percentile interval method) [27] for the estimates of λ , p and the median EFS for the groups following the Weibull distributions. To calculate the bootstrap confidence intervals, 1000 bootstrap samples were generated and for each sample parameters were estimated for the mixture model with a cure group and two Weibull distributions. Starting values for estimation of the parameters in each sample were the estimated values from the original data set. (We found that the parameter estimates for each sample were insensitive to the choice of starting values). The parameter estimates for the two parameter Weibull distributions and 95% bootstrap confidence intervals for the high risk group are $\lambda_3 = 1.34$, (1.05, 2.83) and $p_3 = 1.35$, (1.14, 1.58) and for the intermediate risk are $\lambda_2 = .49$, (.41, .61), $p_2 = 1.30$, (1.18, 1.41). The median EFS and 95% bootstrap confidence intervals for the high and intermediate risk groups are .57, (.27, .72) years and 1.54, (1.22, 1.81) years respectively. For both curves the confidence interval for the scale parameter p_j ($j = 2$ and 3) did not include 1. This indicates the Weibull model is a better fit than the one parameter exponential distribution (a Weibull distribution with $p = 1$ is equivalent to an exponential).

Figure 2a shows the probability of being in the cure group as a function of stage, age and MYCN status. The dashed lines show the probability for each stage with the MYCN unamplified and the solid lines show the probability for each stage with MYCN amplified. As can be seen from the figure, the probability of being in the cure group is higher with MYCN unamplified across all stages. Stage 1 children who have MYCN unamplified have the highest probability of being in the cure group. Figure 2b shows the probability of being in the high risk group. Across all stages and ages, children with MYCN amplified have the highest probability of being in the high risk group. Figure 2c shows the probability of being in the intermediate risk group. As age increases, the probability of being in the intermediate group increases across all stages and either MYCN status. The prediction coefficients for the equations for a given group are provided in table A. in the column labeled 2 Weibulls+C. The range of the curves in the figure reflect the range of ages in each subgroup.

Goodness-of-fit of the selected model was examined by comparing empirical Kaplan-Meier survival curves to average of subject-specific model-based fitted curves from the 2W+C model. The model-based estimated survival for each patient is

$$S(t|x_i, \beta, \lambda, p) = \hat{\theta}_i + \hat{\theta}_2 F_2(t|\hat{\lambda}_2, \hat{p}_2) + \hat{\theta}_3 F_3(t|\hat{\lambda}_3, \hat{p}_3). \quad (4)$$

For different covariate levels we compare average predicted survival from the model to empirical survival using Kaplan-Meier curves. Average predicted survival is obtained from

point-wise averages of the survival curves over patients with a particular covariate level. Figures 3a–c shows the average predicted empirical curves and the average predicted model-based curves. Figure 3a shows the curves by *MYCN* status, Figure 3b shows the curves by the different stages and Figure 3c shows the curves by age groups (less than one year, one to five years, and greater than five years). The curves show that the model fits well for *MYCN* status and stage. The model fits fairly well for younger ages but not as well for older ages. Modeling age using a different functional form may produce a slightly better model fit, but we believe it will have little impact on our final inferences

4. PREDICTION

The 2W + cure model gives the probability that a patient is in one of three groups: a cure group, a group that dies rapidly (high risk) and a group that dies less rapidly (intermediate risk). Given a set of covariates for an individual patient i , the estimated probability of being in each group $\hat{\theta}_{ij}$ can be calculated (where $j=1,2$, and 3 corresponds to the cure, intermediate, and high risk group respectively). In practice, one might use the three estimated probabilities of group membership to assign a patient to a unique risk group so that a patient could be treated with individualized therapy. A key issue is how to use the three continuous predicted probabilities of group membership to categorize patients. That is, how large should $\hat{\theta}_{i1}$ be in order to predict that child i is in the cure group and what covariate combinations correspond to these values. Similarly, a criteria for $\hat{\theta}_{i2}$ and $\hat{\theta}_{i3}$ should be identified that could be used to assign patients to the intermediate and high risk groups.

We propose a two tier system for assigning patients to the different risk groups. In the first tier, we wish to be very certain (have a high probability) that a child predicted to be in a given risk group is truly a member of that risk group. Let l = cure, intermediate or high risk group, let G_i denote true group assignment, and $\hat{G}_i$ be predicted group assignment. Then, the criterion is set so that $P[G_i = l | \hat{G}_i = l] > .9$. Under this criterion there is a probability of .9 that the predicted group membership is correct for an individual child. We form our criteria by conditioning on the predicted group rather than the true group since we are interested in the clinical relevance rather than the inherent accuracy of the prediction [28].

Because we set a very high criterion for group assignment in the first tier, there are some patients whose group membership cannot be assigned in the first tier. In the second tier, the patients not meeting the 90% criterion are assigned to a risk group based on the highest value of $\hat{\theta}_{i1}$, $\hat{\theta}_{i2}$, or $\hat{\theta}_{i3}$. In the first tier the probability of being assigned to the correct group will be 90% by design; however, this probability will be much lower and unspecified in the second tier.

In the first tier for each group j , a value R_j is chosen such that individual i , is predicted to be in group $G_i = l$ if $\hat{\theta}_{ij} > R_j$. The value R_j gives $P[G_i = l | \hat{G}_i = l] > 90\%$. The following method is used to determine R_j . Five thousand data sets were generated using the estimated parameter values ($\hat{\beta}$, $\hat{\lambda}$ and $\hat{p}$) and covariates from the original data set. For each of the data sets the 2W+cure model was fit. Then for each observation the true and predicted group memberships were recorded for each value r' where $r' = .51$ to $.99$ by $.01$. The quantities R_j are computed as $R_j = \min(r')$ where

$$\frac{\# \text{ patients with } G_i=l \text{ and } (\hat{G}_i=l \text{ when } \hat{\theta}_{ij} > r_{r'})}{\# \text{ people with } (\hat{G}_i=l \text{ when } \hat{\theta}_{ij} > r_{r'})} > .9,$$

for each group j . Note that R_j is required to be larger than .5 which ensures that a patient will be predicted to be in only one group.

Based on the 5000 data sets with 2558 observations, the following cutpoints satisfy the criteria

$\hat{\theta}_{i1} > .85$, then assign patient i to cure group
 $\hat{\theta}_{i2} > .76$, then assign patient i to intermediate risk group
 $\hat{\theta}_{i3} > .88$, then assign patient i to high risk group
 otherwise, unclassified.

In the second tier of predicting patients risk groups, patients who are unclassified in tier 1 are predicted to be in a risk group based on the maximum $(\hat{\theta}_{i1}, \hat{\theta}_{i2}, \hat{\theta}_{i3})$. This prediction will have much more misclassification since the largest two θ s could be very similar, yet the largest value will determine a patient's risk group. Through simulations we look at the misclassification rates for each tier separately.

Figure 2 has the tier 1 criterion for predicted group membership indicated on each plot. From the figure the covariate values that are important in determining group membership for the patients in the 90% criteria group can be identified. Note that in the second tier a given value of θ_{ij} may result in two different risk group assignments for two different patients. For example, consider two patients with different covariate values resulting in two sets of θ . It would be possible for patient one to have $\hat{\theta}_{i1} = .3$, $\hat{\theta}_{i2} = .37$, $\hat{\theta}_{i3} = .33$ and patient two to have $\hat{\theta}_{i1} = .40$, $\hat{\theta}_{i2} = .37$, $\hat{\theta}_{i3} = .23$. In this example, despite both patients having $\hat{\theta}_{i2} = .37$, patient one would be assigned to the intermediate risk group and patient two would be assigned to the cure group.

According to the first tier prediction, only patients *MYCN* unamplified are in the cure group. For stage 1, 2, and 3 the age requirements for being in the cure group are age less than 8.65, 3.57, and 2.01 years, respectively. The $R_3 > .88$ criteria predicts no one will be in the high risk group. Older ages predict being in the intermediate risk group with the age requirement increasing as stage decreases.

In order to investigate the stability of the combination of covariates that predict patient cure in the first tier, we evaluate 80% intervals around the criteria that predict for cure using the bootstrap [27]. We computed 1000 bootstrap data sets from the original data set and, for each data set, estimated the parameters from the three group mixture model. We then calculated the θ for all combinations of stage and *MYCN* status with age ranging from zero to fifteen years. The only covariate combinations that predicted cure in the first tier were *MYCN* unamplified and stages 1–3. For these covariate combinations the 10th and 90th percentiles of the age criteria that predicted cure were recorded. For stage 1, the 80% bootstrap confidence interval for the minimum age to be in the cure group is zero (i.e., both

the 10th and 90th percentiles are zero) and the maximum age interval is (6.90,10.52) years. For stage 2 the confidence interval around the minimum age is (0,.51) years and the maximum age is (1.66,4.96) years. For stage 3 the minimum age is zero and the maximum confidence interval is (1.66,2.60) years.

The criteria for predicting cure, essentially places all stage 1 patients in the cure group since very few stage 1 patients are greater than 10 years old and very few are *MYCN* amplified. The age criteria for stage 2 and 3 *MYCN* unamplified are around two to three years. These criteria are in agreement with guidelines presented in [2].

5. Robustness

From the data analysis, a three group model was chosen and then a two tier procedure for assigning patients to risk groups was developed. We performed three simulations in order to evaluate the robustness of the group prediction procedure. In all simulations, we analyze the data using the three group model (cure and two Weibulls) and use the prediction procedure to predict group membership. We generated data according to the correctly specified three group model and two misspecified models. For the two misspecified models, simulations were performed where the underlying data are generated according to a two group model and a four group model. All models include a cure group with the other groups following Weibull distributions. The parameters that are needed to generate data in each simulation are given in the appendix in Table A and are obtained from fitting the two, three, or four group mixture models to the original data set. In each simulation, the covariates and maximum potential follow-up for each patient are the same as in the original data set. The maximum potential follow-up for each person was calculated as the difference between the date when a subject entered a study to the time the data were frozen to perform this analysis (in these COG studies patients are still being followed). Each simulation consisted of 5000 replications.

In all three simulations, 90% of patients predicted to be cured in tier 1 were actually cured confirming that the group prediction rule for the cure group is robust to model misspecification. For tier 2 with the misspecified model this percent dropped to 68–72%. The fact that these probabilities are within 4% suggest that the rule is robust to model misspecification.

Recall that for patients who are assigned to a risk group in tier one, we identify the combinations of covariates that predict cure. We are interested in how well these covariate combinations are correctly identified with our approach and how robust the model is. Table 2 displays the simulation results under the correctly specified and misspecified models. Despite the misspecified analysis models (generate two and four groups but then analyze according to three groups), the covariate criteria that predict cure are very similar to the truth. (The true values are calculated by applying the prediction criteria to the parameter values that are used to generate the data). The 80% confidence intervals always included the true value. When four curves are generated, stage 1 *MCYN* amplified patients meet the criteria for cure but the confidence intervals around age both included zero and are overlapping. This indicates that there is little support for including this group as cured.

These simulations show that even if the two or four group model is the correct one, analyzing the data according to the three group model (two Weibulls and a cure) and then making group predictions using the two tier method predicts cure well.

6. Discussion

Neuroblastoma is a heterogeneous childhood cancer. The goal of this paper is to identify patients who are cured and to identify patients with similar types of neuroblastoma. At an individual level, we predict the probability of being in a particular neuroblastoma group. We then develop a method for using these probabilities to assign individual patients to neuroblastoma risk groups.

The biological basis for believing that there are different types of neuroblastoma is strong [1] and our model reflects this belief by assuming a different Weibull survival function with hazard parameter λ_j and shape parameter p_j for each group j . This model could be extended to allow λ_j to depend on covariates. These covariates could be different or the same as those used to predict group membership.

We develop a two-tier method for predicting group membership. In the first tier of the prediction method, we set the criterion so the rate of correctly classifying a patient is 90%. The criterion was set so that we could feel very confident of predicting the correct group. The drawback was that in over 40% of patients we were unable to classify a patient into a risk group. We are able to classify all remaining patients in the second tier, but unlike the first tier, we cannot rigorously control the misclassification rate.

The COG dataset includes patients treated (or observed) on a variety of therapeutic or biologic studies from 1986–2001. Since patients are currently treated with these therapeutic options, the inferences based on this data set are relevant to clinicians. Treatment assignments in this group of patients were based on age, stage and *MYCN* status. Therefore, with the covariate information and prediction model developed here, new treatment strategies that build on the current treatment strategies can be developed. For example, currently stage 4 patients older than 18 months (both *MYCN* amplified and unamplified) are classified as high-risk and undergo surgery, receive intensive chemotherapy, stem cell transplant, radiation, and maintenance therapy. Despite having received this intensive treatment regimen, our model predicts that stage 4 patients over four years of age should be in the intermediate risk group (with predicted median EFS of 1.5 years) and require more alternative treatment options. As another example, there is currently controversy regarding the treatment of stage 1 and 2 *MYCN* amplified patients. Like the stage 4 older patients, some of these patients are treated with aggressive therapy, including stem cell transplant. Recent proposals in COG consider therapy reductions for these patients. Our model does not predict these patients to have a high probability of cure, therefore clinicians should think carefully before reducing therapy.

Neuroblastoma is a childhood cancer that is very heterogeneous. Treatments that cure children with cancer can have very long term health implications. Long term treatment implications must be weighed differently in children than in adults and must be considered heavily when developing treatment regimens. In this paper we provided a methodology that

could be used in the future to group patients with similar types of neuroblastoma (who have had similar amounts of treatment). After patients with similar types of neuroblastoma have been identified targeted treatments could be developed.

Acknowledgments

The authors would like to thank Dr. Malcolm Smith for his thoughtful comments and discussion of the manuscript. We would also like to thank the reviewers for their thorough and insightful review of the manuscript. We thank the Center for Information Technology, National Institutes of Health, for providing access to the high performance computational capabilities of the Beowulf cluster computer system.

References

1. Maris J, Hogarty M, Bagatell R, SLC. Neuroblastoma. *The Lancet*. 2007; 369:2106–2120.
2. London W, Castleberry R, Matthay K, Look A, Seeger R, Shimanda H, Thorner P, Brodeur G, Maris J, CPR, et al. Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the children's oncology group. *Journal of Clinical Oncology*. 2005;6459–6465. [PubMed: 16116153]
3. Breslow N, McCann B. Statistical estimation of prognosis for children with neuroblastoma. *Cancer Research*. 1971; 31:2098–103. [PubMed: 5120301]
4. Berthold F, Kassenbohmer R, Zieschang J. Multivariate evaluation of prognostic factors in localized neuroblastoma. *American Journal of Pediatric Hematologic Oncology*. 1994; 16:107–115.
5. Evans A. Staging and treatment of neuroblastoma. *Cancer*. 1980; 45:1799–1802. [PubMed: 7370930]
6. Saito T, Tsunematsu Y, Saeki M, Honna T, Masaki E, Kojima Y, Miyauchi J. Trends of survival in neuroblastoma and independent risk factors for survival at a single institution. *Medical and Pediatric Oncology*. 1997; 29:197–205. [PubMed: 9212844]
7. Brodeur B, Pritchard J, Berthold F, Carlsen N, Castel V, Castelberry R, De Bernardi B, Evans A, Favrot M, Hedborg F. Revisions of the international criteria for neuroblastoma diagnosis, staging, and response to treatment. *Journal of Clinical Oncology*. 1993; 11:1466–77. [PubMed: 8336186]
8. Brodeur G, Seeger R, Schwab M, Varmus H, Bishop J. Amplification of n-myc in untreated human neuroblastomas correlates with advanced disease stage. *Science*. 1984; 224:1121–24. [PubMed: 6719137]
9. Seeger R, Brodeur G, Sather H, Dalton A, Siegel S, Wong K, Hammond D. Association of multiple copies of the n-myc oncogene with rapid progression of neuroblastomas. *New England Journal of Medicine*. 1985; 313:1111–16. [PubMed: 4047115]
10. Boag J. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society*. 1949; 11:15–44.
11. Berkson J, Gage RP. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*. 1952; 47:501–515.
12. Farewell V. The use of mixture models for the analysis of survival data with long term survivors. *Biometrics*. 1982; 38:1041–46. [PubMed: 7168793]
13. Gordon N. Application of the theory of finite mixtures for the estimation of cure rates of treated cancer patients. *Statistics in Medicine*. 1990; 9:397–407. [PubMed: 2194263]
14. Kuk AYC, Chen CH. A mixture model combining logistic regression with proportional hazards regression. *Biometrika*. 1992; 79:531–541.
15. Maller RA, Zhou S. Estimating the proportion of immunes in a censored sample. *Biometrika*. 1992; 79:731–739.
16. Gamel JW, Vogel RL, Valagussa P, Bonadonna G. Parametric survival analysis of adjuvant therapy for stage ii breast cancer. *Cancer*. 1994; 74:2483–2490. [PubMed: 7923004]
17. Sy J, Taylor J. Estimation in a cox proportional hazards cure model. *Biometrics*. 2000; 56:227–236. [PubMed: 10783800]

18. Peng Y. Fitting semiparametric cure models. *Computational Statistics & Data Analysis*. 2003; 41:481–490.
19. Greenhouse J, Silliman N. Applications of a mixture survival model with covariates to the analysis of a depression prevention trial. *Statistics in Medicine*. 1996; 15:2077–2094. [PubMed: 8896141]
20. Richardson S, Green P. On bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*. 1997; 59:731–792.
21. Stephens M. Bayesian analysis of mixture models with an unknown number of components- an alternative to reversible jump methods. *The Annals of Statistics*. 2000; 28:40–74.
22. Aptec Systems Version 3.2 kent. washington: Gauss Systems; 1995.
23. Schemper M, Smith T. A note on quantifying follow-up in studies of failure time. *Controlled Clinical Trials*. 1996; 17:343–346. [PubMed: 8889347]
24. Li C, Taylor J, Sy J. Identifiability of cure models. *Statistics and Probability Letters*. 2001; 54:389–395.
25. Yu B, Tiwari R, Cronin K, Feuer E. Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*. 2004; 23:1733–1747. [PubMed: 15160405]
26. Titterton, D.; Smith, A.; Makov, U. *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons; 1985.
27. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*. Chapman & Hall; 1993.
28. Pepe, M. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; 2003.

7. Appendix

Table A

Maximum-likelihood estimates for parameters from three models fit to the neuroblastoma data. The estimates were used to generate data in the robustness simulations. Each column represents the parameters associated with a Weibull distribution.

	1 Weibull+C		2 Weibulls+C		3 Weibulls+C	
λ	.66	1.34	.49	3.48	.44	.99
p	1.50	1.35	1.30	1.64	1.19	1.87
β_0	-3.08	-2.47	-7.94	-1.94	-8.12	-4.25
β_{age}	.36	-.13	.70	-1.77	.71	.31
β_{MYCN}	2.02	3.26	1.02	2.40	.83	2.96
β_{stage2}	.92	.62	2.89	.49	3.18	.94
β_{stage3}	.74	-.55	4.56	-3.93	4.32	1.06
β_{stage4}	2.36	.49	6.37	-8.78	6.50	2.12
median EFS*	1.10	.57	1.54	.23	.83	1.66

* Although not a parameter, the median EFS provides insight into the prognostic differences between the various risk groups.

Figure 1.

Cumulative proportion of patients across ages by stage and *MYCN* status. The solid lines designate *MYCN* amplified and the dashed line designates *MYCN* unamplified. The total number of patients in the study is 2558.

Figure 2.

Figure 2a,b,c. The estimated probability of being in each risk group as a function of age (in years). The solid lines designate *MYCN* amplified and the dashed lines designate *MYCN* unamplified. The probability of being in each group is calculated according to the parameter estimates provided in Table A in the appendix. Panel 2a plots the probability of being in the cure group, panel 2b plots the probability of being in the high risk group, and panel 2c plots the probability of being in the intermediate risk group. The tier one prediction rule is plotted in each panel.

Figure 3.

Figure 3a,b,c. Goodness-of-fit curves by MYCN status, stage, and age. The solid lines are the Kaplan-Meier estimates of EFS for each subgroup and the dashed lines are the 2W+cure model based estimates of EFS.

Table 1

BIC values comparing models

Model [*]	BIC	Number of Parameters
1 W	4818.4633	7
1 W + cure	4382.7216	8
2 W	4396.7500	10
2 W+cure	4248.6503	16
3 W	4261.888	18

^{*} The covariates in the all models are age (continuous), indicator for *MYCN* (amplified=1), indicators for stage with stage 1 as the baseline.

Table 2

Covariate combinations that predict cure

	80% Confidence Intervals		True Criteria	
	Minimum Age	Maximum Age	Minimum Age	Maximum Age
2 groups				
Stage 1, <i>MYCN</i> =0	0*	(3.13–4.63)	0	3.72
Stage 2, <i>MYCN</i> =0	0	(.74–1.63)	0	1.17
Stage 3, <i>MYCN</i> =0	0	(1.11–2.22)	0	1.65
3 groups				
Stage 1, <i>MYCN</i> =0	0	(7.26–10.39)	0	8.60
Stage 2, <i>MYCN</i> =0	(0–.57)	(2.26–4.63)	0	3.60
Stage 3, <i>MYCN</i> =0	0	(1.63–2.41)	0	2.00
4 groups				
Stage 1, <i>MYCN</i> =0	(0–.11)	(5.61–8.07)	0	7.16
Stage 2, <i>MYCN</i> =0	(.065–.54)	(2.74–3.94)	0	3.31
Stage 3, <i>MYCN</i> =0	0	(1.61–2.33)	0	2.09
Stage 4, <i>MYCN</i> =1	(0–2.15)	(0–3.87)	** _	–

* Both the 10th and 90th percentiles are 0.

** In the true model this group was not in the cure group.