Models to Predict Outcome from Childhood Neuroblastoma: The Role of Serum Ferritin and Tumor Histology

Jeffrey H. Silber,1 Audrey E. Evans, and Moshe Fridman

Division of Pediatric Oncology, The Children's Hospital of Philadelphia, and the Department of Pediatrics, The University of Pennsylvania School of Medicine [J. H. S., A. E. E.], and Department of Statistics, The University of Pennsylvania [M. F.], Philadelphia, Pennsylvania 19104

ABSTRACT

We report on the development of three multiple logistic regression models to predict death from childhood neuroblastoma in patients treated without bone marrow transplantation. The models have been developed using a data set of 125 patients for whom age, stage, serum ferritin, and/or histology were available from diagnosis. Seventy-seven patients had all four variables recorded at diagnosis, 34 had age, stage, and serum ferritin, and 14 had age, stage, and histology. Minimum time from diagnosis for all patients was 3 years. The four-variable (full) model showed a predictive value positive rate (or 1 — the false positive rate) of 91.3% and a predictive value negative rate (or 1 — the false negative rate) of 94.9%. Survival curves, based on derived “good” and “poor” prognosis, were constructed for the full model of 77 patients and for the same patients using subset models either without ferritin or without histology. Correcting for prognostic factors noted at diagnosis, no time trend could be identified over the study period. Point estimates for the probability of death in all three models are displayed in graphical form. The results suggest that serum ferritin and tumor histology at diagnosis have independent prognostic significance and that patient outcome in neuroblastoma can be very accurately predicted with a four-variable model. Such information will help sort patients into good and poor prognosis for bone marrow transplant and intensive chemotherapy protocol triage and will help evaluate the efficacy of future therapeutic innovations.

INTRODUCTION

Accurate models for prediction of outcome are crucial for pediatric oncology patients, yet are rarely available. Treatment toxicity and the development of tumor resistance make the proper choice of initial treatment essential for long term survival and optimal quality of life. Childhood neuroblastoma patients present an extreme example where treatment decisions, made at diagnosis, have tremendous and irrevocable implications. Since bone marrow transplant has become the only current method of treatment which has succeeded in changing overall survival in poor prognosis patients (1), the decision as to which patients will require transplant and which will not is vitally important. To date, numerous univariate methods have been used to sort good and poor prognosis patients; these include age, stage, serum ferritin, neuron-specific enolase, histology, ploidy, and N-myc amplification. All methods, in the end, must be evaluated on their predictive value positive or negative rates and their specificity and sensitivity rates. However, little work has been done to define these overall rates, nor have multiple regression models been produced to aid in prediction.

The patient experience from the Children’s Hospital of Phil-

Received 7/25/90; accepted 12/17/90.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 This research was supported by the American Cancer Society and NIH Grant KO4CA01480. Point estimates for the derived predictive models, along with 95% confidence intervals, are available upon request.

2 To whom requests for reprints should be addressed, at The Children’s Hospital of Philadelphia, Division of Oncology, Room 9093, 34th & Civic Center Blvd., Philadelphia, PA 19104.

Philadelphia presented an opportunity to construct predictive models. In previous work we have described a set of patients in whom serum ferritin and/or tumor histology were available (2). These two factors, together with age and stage, proved to be the best predictors of outcome of eight predictive variables tested. The present analysis reports an enlarged sample with significantly longer follow-up, thereby enabling us to develop more clinically useful statistical models and more rigorously test selected hypotheses. We have been collecting banked serum from patients at diagnosis since 1972. Using serum ferritin measures and tumor histology when available, we have constructed a set of multiple logistic regression models to predict death in neuroblastoma. The aim of this study, therefore, was to develop useful predictive models for outcome in childhood neuroblastoma, which could be applied to both research and clinical settings.

PATIENTS AND METHODS

Patient Selection. Criteria for entry into this study were determined prior to data analysis, as follows. Patients included in development of the predictive models must have: (a) initially been diagnosed and treated at the Children’s Hospital of Philadelphia from January 1, 1972 to October 15, 1987 (3 years from the date of analysis of this study); (b) not undergone bone marrow transplantation (although analysis with and without bone marrow transplantation patients will be presented); and (c) had serum ferritin and/or tumor histology determined at diagnosis prior to chemotherapy or radiotherapy. There were a total of 152 patients treated between 1972 and October 15, 1987. One hundred forty-six patients had either serum ferritin or histology available for analysis along with age and stage. Of these, 126 never received bone marrow transplant. One patient was considered an outlier and removed from model development. This was a patient noted to have a tumor while still in utero, 2 days prior to emergency cesarean section secondary to fetal distress. Apgar scores were 2 at 1 min and 6 at 5 min. Diagnosis of neuroblastoma stage IVs was made at the third day of life, after a surgical resection of the retroperitoneal tumor. The patient developed multiple complications, including respiratory distress requiring intubation at birth and subsequently requiring ventilatory support until death, grade IV intraventricular hemorrhage, a severe neonatal coagulopathy, obstructive jaundice of unknown etiology, mechanical ventilatory obstruction, and severe gastrointestinal bleeding. The patient died at 15 weeks of age. All of the remaining 125 patients had age and stage available at diagnosis; 77 had both ferritin level and histology available for analysis, 34 had ferritin level but no histology, and 14 had histology but no ferritin level.

Patient Treatment. Treatment for advanced disease patients gradually intensified during the study period, by addition to the basic three agents cyclophosphamide, vincristine, and imidazole used in 1972. In collaboration with other Children’s Cancer Study Group institutions, randomized studies were conducted comparing the additions of doxorubicin, VM-26, and cisplatin. In the late 1970s, a pilot study was carried out using low dose total-body irradiation as a “systemic” agent added to chemotherapy. None of the various agents improved the 2-year survival rate. During the same period, treatment for patients with favorable factors was already reduced, so the majority were treated by surgery alone. For the construction of the following models, we will be assuming that there was no improvement in conventional treatment.
over the years of this study. The time trend analysis shown in “Results” also supports this assumption.

Statistical Methods. Multiple logistic regression and Cox regression models were developed using age, stage, histology, and ferritin level (the “full” model), as well as models for age, stage, and ferritin level (the “ferritin” model) and age, stage, and histology (the “histology” model). Model comparisons were performed using the full data set of 77 patients. Interaction terms were explored between these four variables. Optimal subset models were constructed with the largest possible cut point of 6 years was used, because it produced the best overall coefficients, using both leverage and Cook statistics (3). For the logistic regression models, cut points for splitting “poor” from “good” prognostic status in the logistic regression models was October 15, 1987, or 3 years of age follow-up time for all patients was 3 years and the median was 8.6 years for those censored alive, the use of survival analysis was not critical to guard against bias from unequal follow-up times in censored patients. Therefore, survival analysis was not used, and the reference point for status in the logistic regression models was October 15, 1987, or 3 years from the date of this analysis. Diagnostics were performed, testing for patients that exerted excessive influence on the model’s estimated coefficients, using both leverage and Cook statistics (3). For the logistic regression models, cut points for splitting “poor” from “good” prognosis were based on a probability estimate of 0.5 or greater in the logistic model (4) or 6 years or less estimated survival in the Cox regression model. Receiver operator characteristic analysis showed no advantage to manipulating these cut points to values greater or less than 0.5 for the logistic regression. For the Cox regression models, a cut point of 6 years was used, because it produced the best overall discrimination. Comparisons between the full model and subset models were made by noting the difference in the log likelihood statistics with variables excluded from the full models. Comparisons were also made between survival curves produced by defining good and poor prognosis patients using either the full model or each subset. Differences in survival curves were then compared using Wilcoxon and log-rank test statistics.

The coding of variables for the regression models used the following numeric weights for Evans stage: stage I = 41, II = 51.3, IVa = 53.2, III = 80.6, and IV = 95. These weights reflect the United Kingdom’s experience in childhood neuroblastoma mortality, as reported by Kinney Wilson and Draper (5) during a time period prior to that spanned by this analysis. We also tested a set of categorical variables coding for stages. By coding stages I, II, and IVa as a single group and comparing this group to stage III and stage IV (coded separately), only stage IV was noted to be significantly different from the first group. The resulting model, however, using data-derived categorical weights, produced the same accuracy, specificity, and sensitivity as the modeling using the United Kingdom weights. We chose to report our model with the United Kingdom weights, because they reflected a very large sample that was independent of the Children’s Hospital experience.

Age was coded as the natural logarithm of age, in units of months. Ages 0–6 months were equally weighted as equal to 6 months in the model. Logarithm of age was used rather than a linear measure of age because of the clinical suspicion that increasing age, beyond a certain limit, would not continue to increase risk. It should be pointed out that there was not statistical or clinical difference between using age or logarithm of age in the predictive models. Clinical suspicion led us to use logarithm of age on the belief that such a choice may be superior for older neuroblastoma patients. Histology was coded 1 for unfavorable prognosis, using the Shimada classification, and 0 for good prognosis (6). Ferritin was coded 0 if the patient’s value was not elevated (age corrected) and 1 if elevated (7, 8). Elevated values were as follows: greater than 500 ng/ml for ages 0–2 months, greater than 400 ng/ml for ages 2–4 months, and greater than 150 ng/ml for ages 4 months or more.

We chose our outcome variable in construction of the following models to be death rather than progressive disease, because of some ambiguity in the reporting of progressive disease. However, we note that, as of the time of submission of this manuscript, no patients who are currently alive more than 3 years from diagnosis (the minimum follow-up time for this study) have progressive disease.

Finally, we performed a “bootstrap” analysis of our results using the methodology of Gong (9), based on the work of Efron (10). By randomly sampling our data with replacement, we constructed an estimate of the model “excess error” over the “apparent error” obtained from the derived multiple logistic regression model based on the observed data set. The simulation created 400 separate random samples, each with a sample size of 77. New logistic regression models were estimated for each sample and then compared to the reported logistic regression model, using the newly constructed population. Comparisons were made using the false positive rate, false negative rate, and overall accuracy rate, based on a decision rule cut point probability of 0.5. Using the terminology of Efron and that of Gong, we report the “apparent error,” the “expected excess error,” and the “bias-corrected estimates of error,” as well as the 95% confidence interval for the difference between these estimates (based on 400 simulations) and theoretical estimates based on any larger number of simulations.

RESULTS

Description of Population. Prior to the selection criterion exclusions in this study, there were 146 patients with either serum ferritin or histology obtained prior to treatment. Of these 146 patients, 55% were less than 2 years of age and 58% had stage III or IV disease. After the exclusions described in “Methods,” 59% of 125 patients were less than 2 years of age and 53% had stage III or IV disease. This compares to 51% under 2 years of age and 61% having stage III and IV disease, as reported by Breslow and McCann (11) using Children’s Hospital data from 1947 to 1967. An analysis of the effects on the model of the exclusion of transplant patients is discussed later in this section.

A total of 125 patients were used to construct the three predictive models. Table 1 describes the distribution of age, stage, and status (number alive and dead) for these patients.

There were 77 patients for whom age, stage, histology, and ferritin level were available for analysis (the full model). The median follow-up time in the 53 censored patients was 111 months (range, 40 to 216). Table 2 describes the distribution of ferritin level and histology by stage and status for each group in the full model.
MODELS TO PREDICT OUTCOME IN NEUROBLASTOMA

Table 2 Distribution of ferritin level and histology by stage in full model, with accompanying mortality status (n = 77)

<table>
<thead>
<tr>
<th>Stage</th>
<th>Ferritin favorable and histology favorable</th>
<th>Ferritin unfavorable and histology favorable</th>
<th>Ferritin favorable and histology unfavorable</th>
<th>Ferritin unfavorable and histology unfavorable</th>
<th>Alive</th>
<th>Dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>9 0</td>
<td>14 1</td>
<td>0 0</td>
<td>0 0</td>
<td>7 0</td>
<td>5 1</td>
</tr>
<tr>
<td>II</td>
<td>2 0</td>
<td>2 1</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>2 1</td>
</tr>
<tr>
<td>III</td>
<td>0 0</td>
<td>0 2</td>
<td>0 0</td>
<td>0 0</td>
<td>0 0</td>
<td>3 0</td>
</tr>
<tr>
<td>IV</td>
<td>1 0</td>
<td>4 4</td>
<td>1 2</td>
<td>6 0</td>
<td>0 0</td>
<td>15 0</td>
</tr>
<tr>
<td>Total</td>
<td>11 0</td>
<td>21 4</td>
<td>7 0</td>
<td>7 5</td>
<td>5 3</td>
<td>24 0</td>
</tr>
</tbody>
</table>

Age, stage, and ferritin level were available on 111 patients. The median time of follow-up for the 61 censored patients was 110 months (range, 40 to 216). The age, stage, and histology group consisted of 91 patients. The median time of survival in the 62 censored patients was 115 months (range, 40 to 216).

The pattern of survival for all 125 patients is presented in Fig. 1. Here it can be seen that, by 3 years (the minimum follow-up for the censored alive patients in this study), over 95% of patients who eventually died did so prior to the minimum follow-up for this analysis.

Predictive models were developed using the three groupings described above. Each model combines a set of prognostic variables using multiple logistic regression. The probability of death, given the prognostic variables, was then estimated for each patient. As can be seen in Table 3, for the full model there is an independent predictive contribution from each of the four variables. For each regression there is also an associated contingency table based on the decision cut point of 0.5. If the probability of death was greater than 0.5, we labeled the patient as having poor prognosis; otherwise, we labeled the patient as having good prognosis. Predictive value positive and predicted value negative rates and specificity and sensitivity rates are shown in Table 3, along with an overall accuracy. The c statistic of Hanley and McNeil (12) was used to compare the predictive ability of each model. A c value of 0.5 implies no predictive ability of the model over chance, and a value of 1.0 implies perfect discrimination. As can be seen, the overall accuracy for the model is 93.5%, with a PV¹ positive of 91.3% and a PV negative of 94.4%. If we had included the one patient whom we excluded from analysis due to multiple neonatal complications, our overall accuracy would have fallen to 92.3%, with a PV negative 92.7%. After analysis of leverage and Cook influence statistics, no justification was found for omitting other observations. Interaction terms between the four variables were explored and found to be statistically insignificant, when such terms were added to the four-variable model.

Similar best models were constructed with age, stage, and ferritin level and age, stage, and histology. These models are also shown, with associated contingency tables, in Tables 4 and 5, respectively.

Model Comparisons. To compare models and determine the relative importance of each prognostic variable, two methods were used. First, the full model using four variables was examined by deleting either ferritin level, histology, or both ferritin level and histology from the model. The resulting change in the log likelihood statistic was then compared. In each case, the full model versus either of the three variable models resulted in a significant change in the log likelihood statistic and the associated χ² tests were significant beyond the 0.01 level. Comparison of the three- and four-variable models with a model using age and stage alone yielded similarly significant results, implying that the ability of the models to predict outcome is significantly improved with addition of each prognostic variable. Another method of viewing the importance of the contri-

---

Table 3 Logistic regression full model to predict death (n = 77)

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(age)</td>
<td>1.201</td>
<td>0.575</td>
<td>0.037</td>
</tr>
<tr>
<td>Stage</td>
<td>0.069</td>
<td>0.026</td>
<td>0.009</td>
</tr>
<tr>
<td>Histology</td>
<td>3.077</td>
<td>1.019</td>
<td>0.003</td>
</tr>
<tr>
<td>Ferritin</td>
<td>3.312</td>
<td>1.162</td>
<td>0.004</td>
</tr>
<tr>
<td>Intercept</td>
<td>-11.902</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c Statistic</td>
<td>0.963</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Observed | Model-predicted prognosis
---|---|---|
Alive | Good | Poor
51 | 2 |
Dead | 3 | 21 |

* Predicted value positive = 91.3%; predicted value negative = 94.4%; specificity = 96.2%; sensitivity = 87.5%; accuracy = 93.5%.

---

Fig. 1. The overall Kaplan-Meier survival curve for all 125 patients analyzed in the study, is shown (patients treated with bone marrow transplantation have been excluded). Note that the shortest follow-up time for censored alive patients was beyond 3 years.
bution of both serum ferritin and histology, over and above age and stage, is to compare model-derived survival curves using these same 77 patients. Fig. 2 displays the survival of the 77 good or poor prognosis patients as labeled by the full model (G4,P4), the ferritin model (G3,P3), and a model using only age and stage (G2,P2). Each survival plot is derived from using either the four-variable model in Table 3, a three-variable ferritin model based on the same 77 patients’ age, or a similarly derived model using only age and stage as prognostic factors. As can be seen in Fig. 2, when individual prognostic group survival curves were compared, there was a trend for improved separation for both the good and poor prognosis groupings between the two- and four-variable models. However, there was little difference between the full model and the ferritin model when comparisons were made inside prognostic groupings. The discrepancy between the full model and the ferritin model when comparisons were made together (using the above-mentioned log likelihood approach) there is a definite significant difference, showing an improved model with four variables over just three. Finally, when comparing G4 versus P4, G3 versus P3, or G2 versus P2, all survival curves are significantly different ($P < 0.001$ level), using the log-rank test.

Since there was a nonrandom selection of patients in the four-variable full model, (i.e., 28 stage IV patients with positive bone marrow aspirations did not have tumor histology determined at diagnosis), we explored these stage IV patients for possible bias. The death rate in stage IV patients in the full 4-variable group was 15 of 22 or 68%. The death rate in the stage IV patients with ferritin levels only in the ferritin group was 25 of 28 or 89%. This may point to a potential bias. However, the effect is probably small, since the unequal death rates can be explained largely by a much higher unfavorable ferritin rate in the 28 patients of the ferritin only model. There were 13 of 22 patients or 59% of stage IV patients with elevated ferritin levels in the ferritin model (for those 28 patients not in the full model).

Time Trends. Since the data set used in this analysis spanned 16 years, we looked for any time trends in survival. To accomplish this, we added to the full model various measures related to year of diagnosis. We first used the continuous variable year of diagnosis, and this showed no trend. Next, in separate regressions, we added dummy variables for different year groupings; years 1972–1979, 1980–1987; and 1972–1977, 1978–1981, 1982–1985, and 1986–1987. No clinically important or statistically significant trends could be seen.

Cox versus Logistic Regression. The bulk of the multiple regression models performed in this analysis used logistic regression. This technique was utilized because the median follow-up in the censored patients was well into the plateau phase for neuroblastoma, so we did not anticipate a need for survival methods such as Cox regression. Since our motivation was to determine a long term probability of death, logistic

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(age)</td>
<td>1.067</td>
<td>0.370</td>
<td>0.004</td>
</tr>
<tr>
<td>Stage</td>
<td>0.068</td>
<td>0.017</td>
<td>0.000</td>
</tr>
<tr>
<td>Ferritin</td>
<td>3.200</td>
<td>0.752</td>
<td>0.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>-9.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c Statistic</td>
<td>0.945</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-predicted prognosisa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>58</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(age)</td>
<td>0.811</td>
<td>0.410</td>
<td>0.048</td>
</tr>
<tr>
<td>Stage</td>
<td>0.074</td>
<td>0.023</td>
<td>0.001</td>
</tr>
<tr>
<td>Histology</td>
<td>3.602</td>
<td>0.844</td>
<td>0.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>-10.242</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c Statistic</td>
<td>0.958</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model-predicted prognosisa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>59</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Dead</td>
<td>6</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

* Predicted value positive = 93.0%; predicted value negative = 85.3%; specificity = 95.1%; sensitivity = 80.0%; accuracy = 88.3%.

** Table 4 Logistic regression ferritin model to predict death (n = 111)**

** Table 5 Logistic regression histology model to predict death (n = 91)**
regression appeared more convenient. However, to be certain that our estimates were similar to those obtained using Cox regression, we compared the logistic regression probability estimates to those of Cox regression. These comparisons were made in two ways. First, contingency tables, exactly as those derived in Table 3, were constructed for the Cox regression models, using two methods. Method 1 used estimated survival time in the Cox regression model of greater than 6 years as the definition of good prognosis and less than 6 years as the definition of bad prognosis. The 6-year survival time was derived using receiver operator characteristic analysis. This time period was well beyond the longest failure in the full model. The resulting contingency table yielded the same overall accuracy as the logistic regression method and had almost exactly the same predicted value positive rate, predicted value negative rate, specificity, and sensitivity as the logistic model. No improvement could be made in the overall accuracy of the Cox regression by manipulating the 6-year cut point. Method 2 used the Cox regression model-derived probabilities of survival at 6 years (13), as described above. Using the same decision rule cut point of 0.5, equivalent overall accuracy for the Cox and logistic models was obtained. Finally, we estimated the probability of death using logistic regression and compared these estimates to the probability of death before 6 years using the Cox model (14). The overall Pearson correlation between the Cox regression model- and the logistic regression model-estimated patient probabilities of death was 0.96 ($P < 0.0001$).

Analysis of Bone Marrow Transplant Experience. The decision to remove patients who had received bone marrow transplant from the development of the predictive models was made for two reasons. First, a major motivation for this research was to develop a model to help determine who should receive a transplant. Consequently, including the results from transplant patients would not contribute to the determination of survival in the alternative, non-transplant therapy group. Second, since we suspected that transplantation had changed the survival in some poor prognosis patients, we did not want these results to confuse the prognostic information which was obtained from age, stage, histology, and ferritin levels in those patients who had not received transplants. However, we still wished to determine if any bias in our estimates had occurred due to the exclusion of the transplant patients. In an attempt to address this question, we studied those transplant patients for whom the four variables were available for analysis. There were 12 patients available. Of these, we removed 4 patients who received transplants in first remission. This left 8 patients who received transplants due to progression of disease. Since these patients represented failures, we coded these patients as "deaths," added these patients to the 77 full model group, and re-estimated the model coefficients. What resulted was a model showing coefficients which were almost identical to those of the original full model. Furthermore, this expanded model was then used to construct a prediction contingency table for the original 77 patients, as was shown in Table 3, with exactly the same accuracy, 93.5%, and identical sensitivity and specificity. For completeness we also constructed a model, using the full four-variable model, to predict death in all patients including those who underwent bone marrow transplantation. There was a total of 89 patients. The resulting model was very similar to the original full model excluding transplantation. The model including transplant-treated patients showed an overall accuracy of 92.1% for the 89 patients studied and 94.8% accuracy for the original 77 patients used in the full model.

Bootstrap Analysis of Error. Ideally, we would have preferred to validate our model with either a split-sample approach or an external validation. Since this was not possible, we chose to perform a bootstrap analysis of the reported error rates for this model. Efron and Gong have pointed out that the reported “apparent” error rates determined by any model are an overly optimistic representation of the accuracy of such a model when it is applied to another population not used in the construction of the model. In an attempt to correct for this, using the bootstrap methodology, we determined that the expected excess error of the false positive rate was 1.3%, the excess false negative rate was 2.7%, and the excess inaccuracy was 2.4%. Adding these rates to the multiple logistic regression model results in Table 3 gave a bias-corrected estimate of error of the false positive rate equal to 10%, a corrected false negative rate of 8.3%, and a bias-corrected overall accuracy of 91.1%. These estimates were based on the construction of 400 random samples of size 77. Increasing the number of simulation samples would have given little improvement in these estimates. A 95% confidence interval for the bias-corrected overall accuracy rate using any number of simulations greater than 400 would be included within the range (0.911 ± 0.003). Consequently, 400 simulations were considered adequate.

Use of the Model. Point estimates for the probability of death using each of the developed models are shown in Figs. 3 through 10. For the full model ferritin level and histology status must be specified, and for the ferritin and histology models only ferritin level or histology, respectively, needs be specified. A plot was constructed for each combination of ferritin level and/or histology, with age at diagnosis on the horizontal axis and probability of death on the vertical axis. Each stage has an associated plot, corresponding to one of the five curves on each graph. Note that much caution must be taken when using these graphs, since they represent extrapolations from the model. In particular, there are far fewer observations comprising the graphs where ferritin level and histology diverge than those graphs where ferritin level and histology are coincident. Consequently, results where ferritin level and histology diverge are more suspect than those where ferritin level and histology are coincident. Points estimates for the full variable model are found in Figs. 3–6. Figs. 7 and 8 correspond to the ferritin model, and Figs. 9 and 10 correspond to the histology model. These figures provide a visual method of using the results of the regression models developed in Tables 3, 4, and 5 and also provide a visual method of using the results of the regression models developed in Tables 3, 4, and 5. For completeness we also constructed a model, using the full four-variable model, to predict death in all patients including those who underwent bone marrow transplantation. There was a total of 89 patients. The resulting model was very similar to the original full model excluding transplantation. The model including transplant-treated patients showed an overall accuracy of 92.1% for the 89 patients studied and 94.8% accuracy for the original 77 patients used in the full model.
Figs. 3–6. Point estimates for the probability of death given the four specified variables used in the full model. These estimates are derived from the logistic regression results of Table 3. Note that Figs. 3–6 represent each possible histology and serum ferritin combination. For each combination, stage and age at diagnosis are used to estimate probability of death.

**DISCUSSION**

We have presented a number of multiple logistic regression models to predict death in neuroblastoma. It is clear that use of the four-variable model described here provides an accurate method for determining prognosis with conventional therapy. Since the model is correct in over 90% of cases, it will be difficult to improve significantly upon these estimates. It remains to be seen whether information of N-myc amplification will add any predictive improvement to the model. Every model attempting to compile disparate clinical information makes assumptions which lead to limitations. The models reported here are no exception. The major limitation to clinical applications using the full model and subsets derived in this paper is in the cases where ferritin levels and histology diverge. Since there were only 19 cases of 77 in which this happened, reliable point estimates of probabilities become more difficult to obtain. When ferritin level and histology diverge, age and stage become more important, as can be seen from the graphs in Figs. 5 and 6. Interestingly, by using a decision rule cut point of 0.5, the
model is generally correct. In fact, of the 19 patients that had divergent ferritin levels and histology, only 3 patients were incorrectly predicted by the model. The two other prediction errors in the full model involved patients with favorable ferritin levels and histology, and both of these patients died.

Despite the limitation of small sample size, as noted above, we believe there are a number of potential applications for this model. From a clinical perspective, we have shown that good and poor prognosis can be assigned to patients with, generally, a high degree of accuracy at diagnosis. Such probability estimates could be helpful in determining who would benefit from bone marrow transplant or increasingly intensive therapy (if poor prognosis was determined) and who should be treated with minimal therapy or surgical resection alone (for the good prognosis group). Still other uses can be made of this research. The four-variable model provides an excellent severity of illness correction for those interested in more precisely evaluating the success of bone marrow transplant or higher intensity chemotherapy protocols. A more precise estimate of the expected number of survivors in a transplant group can be made by developing an individual probability of death based on the full information known about that patient. Should a bone marrow

Figs. 7 and 8. Estimated probability of death using the ferritin model (without histology). Here it can be seen that there is less separation of high and low probability combinations, as compared to the full four-variable model.

Figs. 9 and 10. Estimated probability of death using the histology model (without ferritin). Here it can be seen that a similar separation is achieved between combinations of histology as between combinations of serum ferritin.
MODELS TO PREDICT OUTCOME IN NEUROBLASTOMA

transplant program result in improved group survival, after
correction for each individual patient’s characteristics, then a
decision to transplant in defined high risk patients may be
justified. Of course, the utility of aggressive chemotherapy
versus transplantation must also be taken into consideration,
and some increase in probability of survival above that for
the non-transplant protocol would need to be expected for trans-
plantation to be recommended. The advantage of using a model
to individually estimate survival without transplant is that
patient characteristics can be used to better define prognosis.
When these are aggregated across patients, a group-wide ex-
pected survival may be estimated and compared to the actual
survival with or without transplant.

In conclusion, we have presented models to predict survival
in neuroblastoma. These estimates need to be validated on a
larger external data set to establish their ability to be generalized
to other childhood neuroblastoma populations. This is because
apparent accuracy will tend to be overestimated when the same
data used to develop the model of interest are also used to test
for its accuracy. However, we believe that our estimates do use
large enough numbers of patients to contribute to a better
understanding of neuroblastoma and its outcome. Furthermore,
we suggest that, in the future, predictive models in neuro-
blastoma be assessed with estimates of the predictive value
positive and predictive value negative rates for such models.
Should other patient or tumor characteristics be included in
prognostic models, improvements in predictive ability should
be compared to the results shown here. Finally, we believe that
serum ferritin plays an important role in neuroblastoma out-
come prediction. Since serum ferritin has a strong prognostic
significance, independent of age, stage, and histology, the etiol-
ogy for this association may provide clues to tumor behavior
and treatment.

ACKNOWLEDGMENTS

We thank Drs. Giulio J. D’Angio, Anna T. Meadows, and Hie-Wan
Hann for their assistance in conducting this research. We also thank
Jack L. Elias for assistance with the preparation of this manuscript.

REFERENCES

neuroblastoma with supralethal chemotherapy, radiation and allogeneic or
2. Evans, A. E., D’Angio, G. J., Propert, K., Anderson, J., and Hann, H-W. L.
3. Aitkin, M., Anderson, D., Francis, B., and Hinde, J. Statistical Modelling in
5. Kinnier Wilson, L. M., and Draper, G. J. Neuroblastoma, its natural history
7. Saarinen, U. M., and Siimes, M. A. Serum ferritin in assessment of iron
9. Gong, G. Cross-validation, the jackknife, and the bootstrap: excess error
10. Efron, B. Estimating the error rate of a prediction rule: improvement on
12. Hanley, J. A., and McNeil, B. J. The meaning and use of the area under the
13. Crowley, J., and Hu, M. Covariance analysis of heart transplant survival
Models to Predict Outcome from Childhood Neuroblastoma: The Role of Serum Ferritin and Tumor Histology

Jeffrey H. Silber, Audrey E. Evans and Moshe Fridman


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/51/5/1426

**E-mail alerts**  Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.