Research in Higher Education, Vol. 40, No. 3, 1999

PREDICTING THE RETENTION OF UNIVERSITY STUDENTS

Paul A. Murtaugh, Leslie D. Burns, and Jill Schuster

Survival analysis was used to model the retention of 8,867 undergraduate students at Oregon State University between 1991 and 1996. Attrition was found to increase with age, and decrease with increasing high school GPA and first-quarter GPA. Nonresidents had higher attrition rates than did resident and international students, and students taking the Freshman Orientation Course appeared to be at reduced risk of dropping out. Statistically significant associations of retention with ethnicity/race and college at first enrollment were also noted. A proportional hazards regression model was developed to predict a student's probability of leaving school based on these demographic and academic variables. These analyses are helping to guide the university's efforts to improve retention through marketing, recruitment, and the development of orientation and other programs.

INTRODUCTION

Retention of students at colleges and universities has long been a concern for educators. Published work on student retention has focused on several themes: (1) examining the relationship between precollege characteristics of freshman students (e.g., high school GPA, SAT scores) and their success at a college or university (Anonymous, 1997); (2) examining the causes of student attrition, with recommendations to colleges and universities for interventions to reduce the rate at which students leave school before graduating (Astin, 1993; Tinto, 1993); (3) describing and evaluating specific campus programs that were designed and implemented to improve retention of students in general (Boudreau and Kromrey, 1994; Glass and Garrett, 1995; Reyes, 1997) and for specific groups of students (Dodd et al., 1995; Johnson, 1996; Person and Christensen, 1996); and (4) exploring the relationship between innovative or improved

Paul A. Murtaugh, Department of Statistics, Oregon State University; Leslie D. Burns, Undergraduate Academic Programs, Oregon State University; Jill Schuster, University Marketing, Oregon State University. Address correspondence to: Paul A. Murtaugh, Department of Statistics, Oregon State University, Corvallis, Oregon 97331.

0361-0365/99/0600- 0355$16.00/0

teaching techniques and student retention (Dougherty et al., 1995; Moore and Miller, 1996).

According to this body of work, precollege characteristics can be useful predictors of student retention, and the results of these studies have had direct implications for the recruitment of students. However, precollege characteristics do not explain all of the variation in attrition rates of students. Students are more likely to stay in school when they are actively involved in campus activities and feel a sense of community in the institution (Astin, 1993; Tinto, 1993; Naretto, 1995). Consequently, schools have implemented a number of support programs to increase students' feeling of connection to the institution. Because the greatest attrition tends to occur between the freshman and sophomore years, these programs have tended to focus on first-year students.

The literature includes descriptions of strategies that specific colleges and universities have found helpful in retaining students (e.g., see Bedford and Durkee, 1989; Hyman, 1995). The extent to which these strategies work at other institutions and with other groups of students is quite variable. Crucial to guiding the development and implementation of measures to improve student retention at an institution is an understanding of the factors that influence retention at that institution.

Longitudinal studies, in which individual students are followed through time, have often been used to identify risk factors associated with leaving before graduation. Various techniques have been used to analyze or model the data accruing from such studies, including simple cross-tabulation (Avakian et al., 1982), two-sample comparisons (Naretto, 1995), logistic regression and its variants (Dey and Astin, 1993), structural equation modeling (Cabrera et al., 1993), and Markov processes (Heiberger, 1993).

Here we use survival analysis to model student retention at Oregon State University (OSU) between 1991 and 1996. This methodology considers the timing as well as the occurrence of students' leaving, and it accommodates "censored" individuals, i.e., students who were still enrolled or who had graduated before study closure. The purpose of our paper is twofold: (1) to illustrate the use and advantages of survival analysis with retention data, and (2) to elucidate some of the factors associated with student retention at Oregon State University.

DATA

Information was obtained from a computerized database called the OSU Student Data Warehouse for 8,867 students who enrolled as first-time freshmen in the fall quarters of 1991 through 1995, inclusive. Throughout the paper, we use the term *withdrawal* to refer to the act of leaving school before graduation, whether or not the student actually formally withdrew from the University. Our dependent variable was the maximum observation time for each student, i.e., the time to withdrawal, graduation, or the closing date of our study (fall 1996), if the student was still enrolled at that time. Table 1 gives information on the status of the 8,867 students as of fall 1996.

As potential predictors of withdrawal, we focused on demographic and academic variables that were available as of each student's first quarter at the University. Table 2 includes a listing of these variables.

While we were limited by the availability of information in the OSU database, we believe the variables in Table 2 summarize many of the important influences on student retention. The model included students' pre-college characteristics (e.g., high school GPA, SAT score), involvement in campus programs (e.g., Freshman Orientation Course, Educational Opportunities Program), and relevant demographic characteristics (e.g., ethnicity/race, sex). The quality of a student's experience at a university will be influenced by how well he or she is prepared academically, as reflected by high school GPA and SAT score. Previous workers have identified differences in retention attributable to ethnicity/race and sex (e.g., see Avakian et al. 1982), and others have emphasized the importance of freshman orientation courses (Bedford and Durkee 1989). Finally, age and geographic origin may influence a student's quality of life at the university, and retention rates may differ among colleges, because of differences in the characteristics of students initially attracted to different academic programs and/or in their experiences and satisfaction with these programs.

Some potentially important variables were not retrievable from the OSU database. For example, we do not have information on the financial well-being of students, their living arrangements, or the extent of their personal interactions with other students and professors at the University. This prevents us from making inferences about such variables, but it is hoped that models based on the available variables will still lead to reliable predictions of student retention.

STATISTICAL METHODOLOGY

We modeled student retention using the statistical methodology known as survival or failure-time analysis (e.g., see Kalbfleisch and Prentice, 1980;

TABLE 1. Status of the 8,867 Students as of Fall 1996

Date Enrolled	Total Students	Status as of Fall 1996 Withdrawn	Continuing	Graduated
Fall 1991	1609	592	145	872
Fall 1992	1710	676	590	444
Fall 1993	1637	623	1005	9
Fall 1994	1886	652	1234	0
Fall 1995	2025	496	1529	0

Lawless, 1982; Lee, 1992; Collett, 1994). These methods are appropriate for responses that are times to the occurrence of some event, e.g., withdrawal from the university. A subject for whom the event has not occurred by the end of the study is said to be *censored*. Survival analysis makes efficient use of the information available from all subjects, censored and uncensored, in identifying associations of survival probabilities with independent variables of interest.

The starting time in all our analyses was the date of the student's enrollment as a freshman at OSU. If a student withdrew from the university before graduating, the event time was the date of withdrawal. All other students were censored at their times of last known enrollment—either the date of graduation, or the date of study closure (fall 1996), if the student was still enrolled at that time.

During the time period investigated, between 25% and 35% of the students whose enrollment was interrupted returned to the university within a year. For example, of the students enrolled during fall 1995 who did not return in winter 1996, 26.1% had returned to the university by winter 1997 for at least one term: 13.5% stopped out for one term; 10.8% stopped out for two terms; and 1.8% stopped out for three terms. According to university regulations, students who have stopped out for more than three terms are required to go through a readmission process. For this analysis, students who had stopped out for more than three terms were considered as withdrawals and then as new students when they were readmitted.

For univariate analyses (i.e., looking at one predictor variable at a time), we estimated retention probabilities using the Kaplan–Meier method (Kaplan and Meier, 1958). This method provides a nonparametric estimate of retention probability over time, with appropriate adjustment for the varying numbers of students at risk of withdrawing, as students are censored due to study closure or graduation. It allows us to make full use of the data at hand, including that for the later-enrolling students for whom only limited follow-up is available. The log-rank test (Peto and Peto, 1972) was used to test for significant differences in retention probabilities among categories of a predictor.

For multiple-variable analyses (i.e., looking at two or more predictors simultaneously in a single model), we used the Cox proportional hazards regression model (Cox, 1972). In the Cox model, each student is given a risk score:

$$R = \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{1}$$

where $X_1, X_2, \ldots, X_k$ are the values of k independent variables (risk factors) and $\beta_1, \ldots, \beta_k$ are regression coefficients, estimated from the data by the method of maximum likelihood. High risk scores correspond to poor retention.

A key quantity in survival analysis is the hazard function, or instantaneous failure rate, in units of reciprocal time. If R_0 is the risk score for a hypothetical reference individual, and $\lambda_0(t) \equiv \lambda(t, R_0)$ is that individual's hazard at time t, then the hazard for an individual with risk score R can be written

$$\lambda(t, R) = \lambda_0(t) \cdot \exp(R - R_0) \tag{2}$$

This equation implies that the hazard ratio for individuals with risk scores R_1 and R_2 is $\lambda(t, R_2)/\lambda(t, R_1) = \exp(R_2 - R_1)$. Therefore, recalling Equation (1), we see that each regression coefficient β_i has the interpretation that a unit increase in the i^{th} covariate, X_i, increases the hazard of withdrawal by the multiplicative factor $\exp(\beta_i)$.

In most applications, we would like to estimate the survivor function, $S(t, R)$, i.e., the probability that a student with risk score R will not have withdrawn after t years. The survivor function and the hazard function are related by $S(t, R) = \exp[-\int_0^t \lambda(u, R)\, du]$. For the hypothetical reference individual having risk score R_0, we can estimate $S(t, R_0)$ using a method suggested by Breslow (1974). It then follows from the proportional hazards assumption that

$$S(t, R) = \{S_0(t)\}^{\exp(R - R_0)} \tag{3}$$

where $S_0(t) \equiv S(t, R_0)$ is referred to as the baseline survivor function. We defined our reference, or baseline, student as a 19-year-old white resident, enrolled in the College of Science, with a 3.4 high school GPA and 2.7 first-quarter GPA, who did not attend the Freshman Orientation Course.

The use of Equations (1) and (2) to calculate risk scores and retention probabilities is illustrated in detail in the Appendix.

Model estimation and assessment were done in the S+ language (Statistical Sciences Inc., 1993). To identify a subset of variables that adequately predicted retention, we used stepwise subset-selection methods. The statistical criterion for adding and retaining variables in the model was $P < 0.01$. Students with missing values of one or more predictors were omitted from the stepwise procedure.

RESULTS AND DISCUSSION

Univariate Analyses

Figure 1 shows a Kaplan–Meier curve of the estimated retention probabilities over time for the entire cohort of 8,867 students. The drops in the curve after each quarter represent withdrawals of students, scaled according to how many were at risk of withdrawing at the time. At the end of the summer quarter following the fourth academic year (rightmost point on Fig. 1), the estimated retention probability is 0.593. That is, Oregon State University has a roughly 40% attrition rate over four years. We chose to truncate our graphs at four years

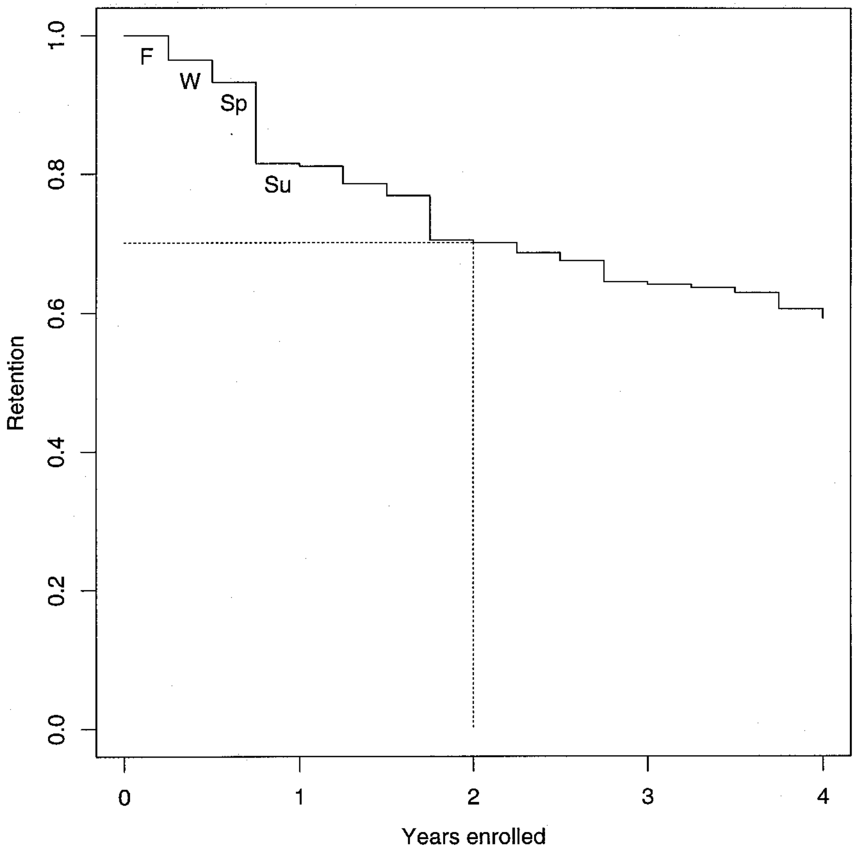

FIG. 1. Kaplan–Meier curve for the entire cohort of 8,867 students. Time zero is the beginning of each student's freshman fall quarter, and the curve drops after every quarter (fall, winter, spring, summer, as shown for the first year) to reflect attrition. The dotted line shows how to read from the graph the estimated retention probability (0.702) for the beginning of the fall quarter of the third year.

because of the large number of graduating seniors who are censored at or near that time, even though the data set includes follow-up on 1,325 students who were at the university for more than four years. A separate analysis could focus on characteristics of these "extenders," as they have been called (Volkwein and Lorang, 1996).

Figure 1 illustrates an advantage of survival analysis over some other methods of quantifying retention. Unlike logistic regression and other techniques based just on the occurrence of some event ("yes" or "no"), survival analysis makes use of the *timing* of events. In our application, the Kaplan–Meier curves give a detailed look at the dynamics of withdrawal through time.

For example, notice in Fig. 1 that withdrawals tend to occur in pulses at the end of each school year, with an especially precipitous decline at the end of the students' first spring quarter. Examination of the forms of these curves could give clues about what factors are triggering the withdrawals.

Table 2 examines differences in retention among categories of a variety of independent variables, and Figure 2 presents graphical summaries for two of these variables (residency and first-quarter GPA). All of the predictors except sex show statistically significant associations with retention. Because P-values decrease with increasing sample size, the large number of students in this study means that even weak associations may be judged statistically significant. Attention should therefore focus on the magnitudes of the differences shown in Table 2, rather than their statistical significance.

The associations in this observational study do not imply cause-and-effect relationships. For example, consider the results for participation in the Educational Opportunities Program, which provides assistance to nontraditional students. The four-year probability of withdrawal is 56% for students participating in the program and 39% for nonparticipants. But there are many other differences between these groups of students besides their involvement in the program—e.g., the mean high school GPA is 2.88 for participants and 3.40 for nonparticipants. Thus, the unfavorable retention profile does not necessarily imply that the program is ineffective.

Multiple-Variable Analyses

Multiple-variable analyses allow us to identify important subsets of predictors in the face of potentially strong correlations among the predictors. Stepwise variable selection based on Cox proportional hazards regression led to a model containing the following independent variables, listed along with the P-values for their inclusion in the model: first-quarter GPA ($P < 0.0001$), Freshman Orientation Course (<0.0001), high school GPA (<0.0001), residency (<0.0001), college (0.0065), ethnicity/race (0.0066), and age (0.0075). The final model was based on 8,293 students (574 observations were omitted due to missing values of one or more of the predictors). The Appendix gives the estimated regression coefficients and illustrates in detail how this model can be used to estimate retention probabilities.

Table 3 shows hazard ratios, i.e., factors by which a student's hazard of withdrawal is multiplied by a unit increase in the predictor, for both univariate models (in which each predictor was entered separately into a Cox model) and the final multiple-variable model. For example, the multiple-variable model predicts that the risk of withdrawal for a student with a first-quarter GPA of 3.5 is 49% of that for a student with a GPA of 2.5, and the risk for a liberal arts student is 15% higher than that for a College of Science student, assuming equality of other risk factors.

TABLE 2. Estimated Probabilities of Retention after 1, 2 and 4 Years

Variable (and categories)	Number of Students	Retention 1 Year	Retention 2 Years	Retention 4 Years
Age at first enrollment (2 missing)				
12.1–18	814	0.810	0.680	0.572
18–20	7721	0.814	0.709	0.602
20–25	270	0.759	0.571	0.451
25–44.6	60	0.750	0.650	0.341
Sex				
Female	4061	0.811	0.698	0.594
Male	4806	0.812	0.704	0.590
Ethnicity/race (208 missing)				
Asian	831	0.848	0.733	0.627
Black	183	0.809	0.623	0.416
Hispanic	345	0.765	0.617	0.492
American Indian	135	0.704	0.595	0.537
Other	40	0.825	0.640	NA
Pacific Islander	35	0.686	0.610	NA
White	7090	0.812	0.707	0.601
Residency				
International	168	0.821	0.740	0.598
Nonresident	1602	0.763	0.611	0.502
Resident	7097	0.822	0.721	0.613
College at first enrollment (1 missing)				
Agricultural Sciences	474	0.823	0.718	0.606
Business	1420	0.815	0.706	0.595
Liberal Arts	1091	0.777	0.661	0.528
Forestry	226	0.739	0.634	0.548
Home Economics & Education	376	0.816	0.700	0.581
Health & Human Performance	446	0.841	0.721	0.592
Pharmacy	260	0.838	0.738	0.659
Pre-Engineering	2002	0.849	0.749	0.641
Science	1616	0.803	0.698	0.598
University Exploratory Studies	955	0.768	0.635	0.534
High school GPA (386 missing)				
Less than 2.0	41	0.659	0.316	0.175
2.0–2.7	417	0.751	0.563	0.350
2.7–3.3	3198	0.751	0.615	0.502
3.3–4.0	4825	0.862	0.780	0.689

TABLE 2. (*Continued*)

Variable (and categories)	Number of Students	Retention 1 Year	2 Years	4 Years
SAT score (630 missing)				
Less than 1000	2842	0.777	0.645	0.521
1000–1200	3878	0.823	0.726	0.632
1200–1600	1517	0.873	0.789	0.676
First quarter GPA				
0.0–2.0	1611	0.572	0.434	0.330
2.0–2.7	2583	0.818	0.688	0.563
2.7–3.3	2579	0.876	0.772	0.659
3.3–4.0	2094	0.907	0.840	0.759
Participation in Educational Opportunities Program				
No	8201	0.814	0.709	0.606
Yes	666	0.776	0.620	0.444
Enrollment in Freshman Orientation Course				
No	3651	0.776	0.654	0.546
Yes	5216	0.836	0.735	0.624

Probabilities are estimated by the Kaplan–Meier method. Differences in retention among categories are statistically significant for all variables except sex ($P = 0.89$).

It is evident from Table 3 that our view of the nature of an association may well depend on what other variables are in the model. An interesting example is shown in Figure 3, which plots hazard ratios for different races. Univariate analyses suggest that blacks, Hispanics, and American Indians are at higher risk of withdrawing than are whites, but, in the multiple-variable analysis, the differences for Hispanics and American Indians disappear, and blacks actually appear to have reduced risk, compared to whites. In other words, the average black is more likely to withdraw than is the average white, but, if we compare students of similar age, GPA, residency, etc., the black student is actually less likely to withdraw. Since some of the predictors are strongly correlated (e.g., the two GPA variables), however, it may not be sensible to envision varying one while holding the other constant. Therefore, the multiple-variable model is probably most useful as a tool to predict overall risk of withdrawal, rather than as a means of elucidating the nature of the associations with individual predictors.

To obtain a rough index of the goodness-of-fit of the final model, we can use it to “predict” the retention of the students in our data set, and then compare these predictions to the observed retention. Figure 4 shows the results of such

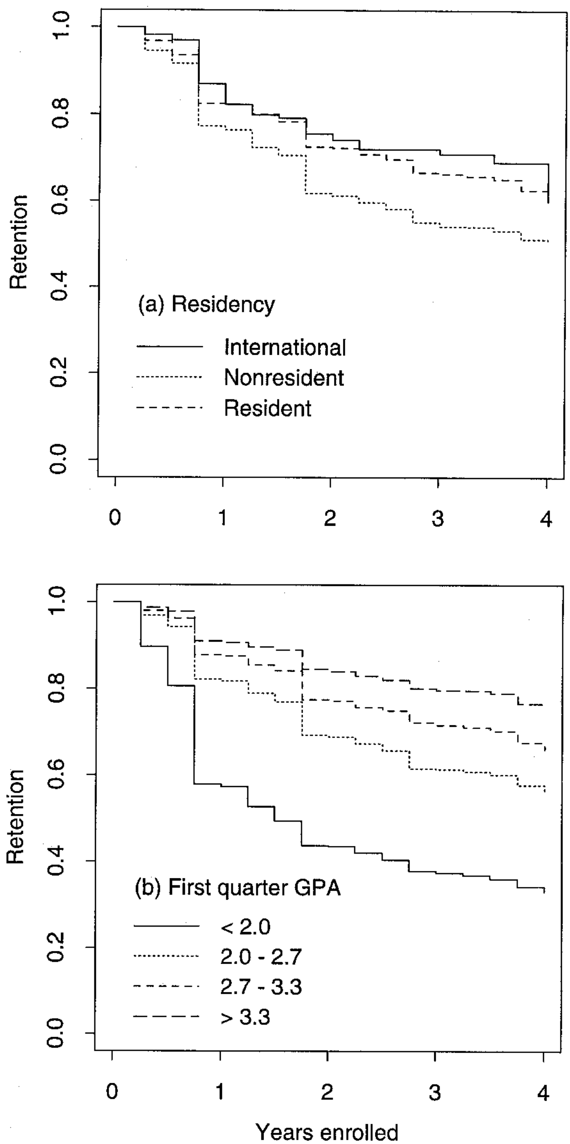

FIG. 2. Kaplan–Meier curves for different categories of (a) residency and (b) first-quarter GPA.

TABLE 3. Relative Hazards of Withdrawal, Based on Univariate and Multiple-Variable Analyses

Variable (and categories)	Hazard Ratio	
	Univariate	Multiple-variable
Age	1.04*	1.05*
Ethnicity/race		
Asian	0.88	0.83*
Black	1.38*	0.68*
Hispanic	1.37*	0.95
American Indian	1.45*	1.14
Other	1.07	0.91
Pacific Islander	1.47	0.96
Residency		
International	0.89*	0.49
Nonresident	1.43*	1.29*
College		
Agricultural Sciences	0.94	0.93
Business	1.00	1.06
Liberal Arts	1.19*	1.15*
Forestry	1.32*	1.16
Home Economics & Education	1.04	1.09
Health & Human Performance	0.93	0.87
Pharmacy	0.79	0.82
Pre-Engineering	0.82*	0.94
University Exploratory Study	1.25*	1.15*
High school GPA	0.41*	0.73*
First-quarter GPA	0.46*	0.49*
Freshman Orientation	0.76*	0.79*

For age and the GPA variables, the hazard ratio is the factor by which a student's risk of withdrawal is multiplied by a unit increase in the variable. For the other variables, the ratios are relative to the appropriate reference categories: white, resident, College of Science, and nonparticipant in Freshman Orientation. Asterisks denote statistical significance of regression coefficients, or comparisons to the reference categories.

an exercise for students divided into four groups based on risk score. Agreement between observed and predicted retention is good, with perhaps a tendency for slight underprediction of retention for the two highest-risk groups. Since we are predicting retention for students who were used in the construction of the model, this good agreement is not surprising. A more informative

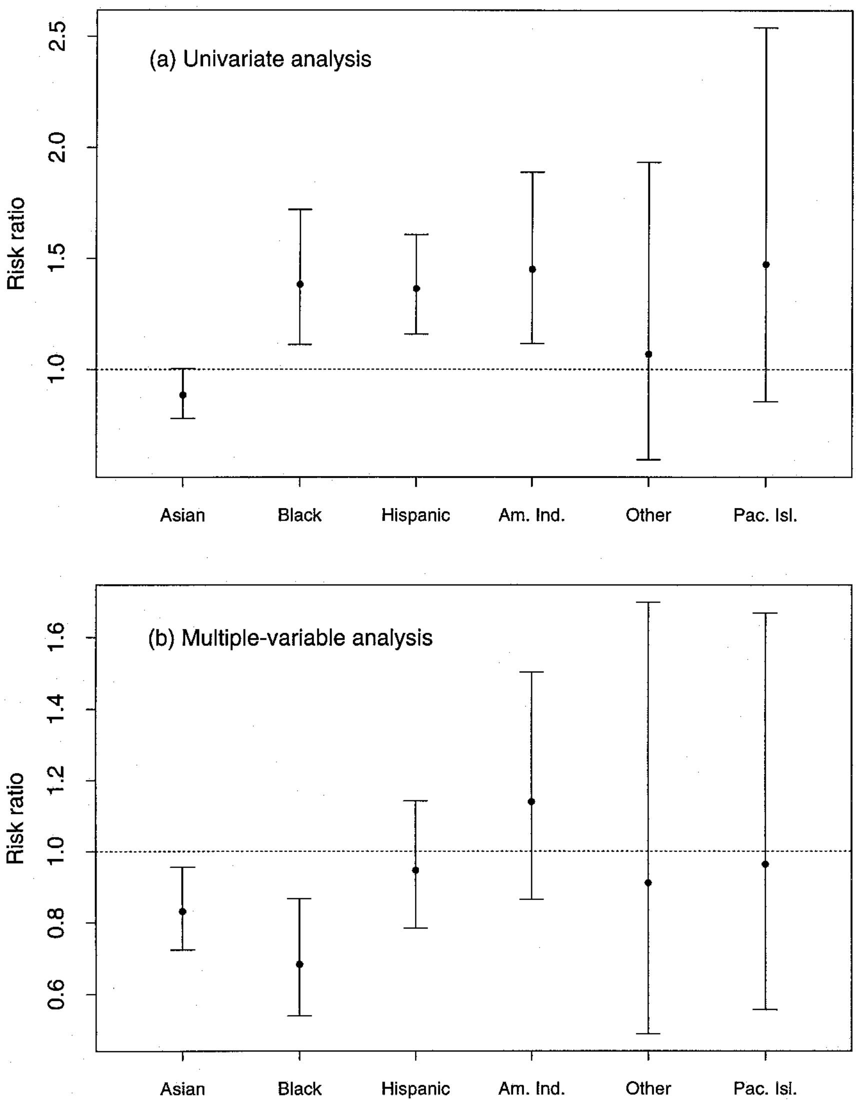

FIG. 3. Hazard of withdrawal for different races, relative to whites (for whom the hazard is defined to be 1), as estimated in (a) univariate analyses, and (b) the multiple-variable model. Bars give 95% confidence intervals.

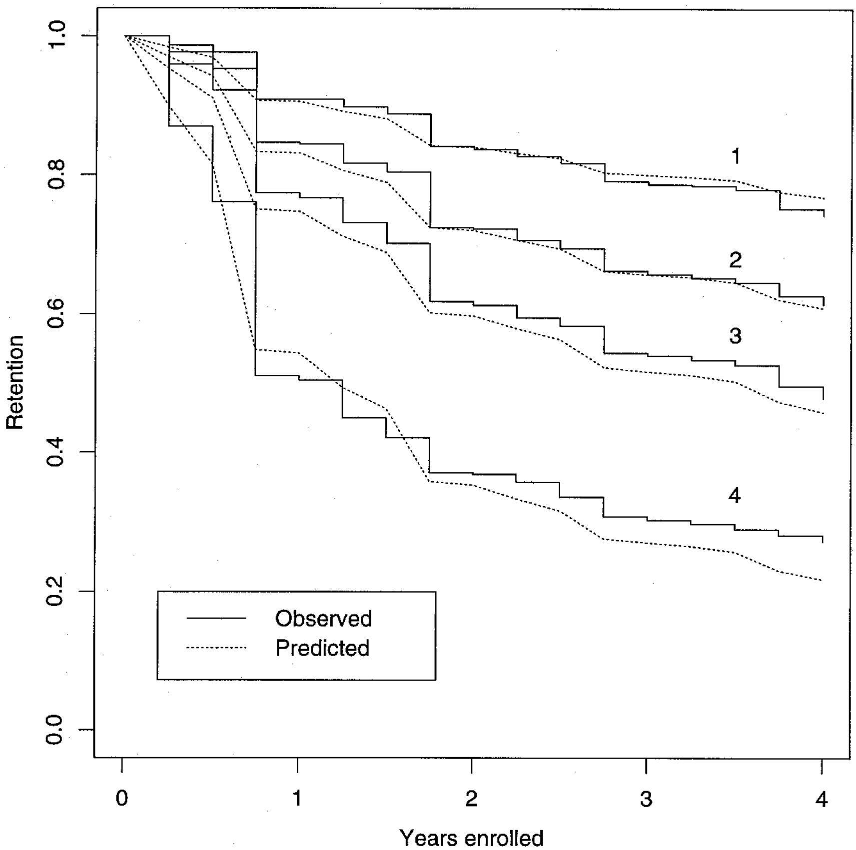

FIG. 4. Comparison of observed retention (Kaplan–Meier curves) to that predicted by the multiple-variable model, for four risk groups of students: 1 = lowest risk (3,496 students, 703 withdrawals, risk scores from −3.78 to −2.34); 2 = lower risk (2,158 students, 701 withdrawals, risk scores from −2.34 to −1.89); 3 = higher risk (1,587 students, 705 withdrawals, risk scores from −1.88 to −1.39); 4 = highest risk (1,052 students, 700 withdrawals, risk scores from −1.39 to 0.66).

view of the predictive ability of the model might be obtained with a data-splitting or bootstrapping approach (Harrell et al., 1996).

Implications

These analyses have had important implications for current efforts to improve student retention at Oregon State University. Taken in conjunction with other qualitative and quantitative market research data on targeted student populations, the data point to both opportunities and areas for improvement at Oregon State.

One key opportunity identified by this analysis is the strengthening of efforts to orient new students to increase the likelihood of success during their first three terms. The data suggest the value of being more sensitive in our orientation process to the needs of out-of-state students and students who are members of ethnic minority groups. In fall 1997, OSU initiated a new week-long student orientation program, to be continued throughout the first year for new students, supplementing a revised Freshman Orientation Course.

This program was tested among prospective students and has been promoted in recruiting publications as a tool for increasing graduation rates. Over half of the incoming fall 1997 freshman class responded positively to these messages and participated in the orientation program.

Our analysis indicates that in-state students have lower attrition rates than nonresidents. Based on this result, marketing and recruiting efforts have been strengthened within the state of Oregon while we explore possible reasons for the relatively poor retention of out-of-state students. It appears that scholarship opportunities may be more limited and that our summer orientation programs may be inconvenient for out-of-state students. Programs are under way to build scholarship programs that include out-of-state students, and the New Student Orientation has been scheduled just prior to the start of classes for the convenience of new students.

An additional area of concern is the retention of black and other minority students. As a direct result of our analysis, a marketing plan aimed at high-achieving black students in Oregon was recently developed. It appears from the multiple-variable model that black students are more likely to graduate than are members of other ethnic groups, if they enter equally prepared. This message has been incorporated into marketing efforts aimed at black students. Not surprisingly, black students have expressed significant concerns about their ability to succeed academically at a predominantly white institution, as compared to other alternatives. Advertising, recruiting messages, and publications designed to address these concerns have been favorably received in qualitative testing.

A final area for improvement identified by our analysis centers around the age of incoming students. While qualitative and anecdotal information suggests that older students tend to have higher graduation rates, retention rates at OSU appear to *decrease* with increasing age. Based on other information, it appears that OSU may not have offered relevant courses at the times and places that are most convenient for older students. Consequently, OSU has restaged its Continuing Higher Education curriculum under the OSU-Statewide initiative program banner. This program has been designed with older-than-average students in mind, and it is expected to improve our ability to attract and retain these students.

SUMMARY

Survival analysis allows the efficient modeling of student retention based on data from recently enrolled students as well as from students who have already graduated or dropped out. Univariate techniques permit the study of one predictor variable at a time, while multiple-variable models illustrate the association of a variable with retention after adjustment for other predictors.

Our application of survival analysis to the OSU data indicates important, independent associations of student retention with age, residency, high school and first-quarter academic performance, college at first enrollment, ethnicity/race, and enrollment in the Freshman Orientation Course. Among our more surprising results are the superior predictive value of high school GPA over SAT score; the decrease in retention with increasing age at enrollment; and the difference between the univariate and multiple-variable views of the association between ethnicity/race and retention. These results should be of use in focusing recruitment efforts on the most promising students, developing programs to increase student retention, and identifying enrolled students who are at high risk of withdrawing before graduation.

Acknowledgments. We thank K. Lee and A. Mathany for expert help in extracting information from the database; G. Beach and D. Faulhaber for helpful input and discussions; and two anonymous reviewers for constructive suggestions on an earlier version of the manuscript.

APPENDIX

We illustrate the calculation of retention probabilities from the multiple-variable model fit to the OSU data. The first step is to calculate risk scores using regression coefficients (β's) obtained from computer output, as in Equation (1). For the categorical variables:

$$R_{\text{college}} = \left\{ \begin{array}{llll} -0.075 & \text{Agric. Sci} & -0.141 & \text{HHP} \\ 0.056 & \text{Business} & -0.195 & \text{Pharmacy} \\ 0.136 & \text{Lib. Arts} & -0.062 & \text{Pre-Engr.} \\ 0.150 & \text{Forestry} & 0.142 & \text{UES} \\ 0.084 & \text{HEE} & 0 & \text{Science} \end{array} \right.$$

$$R_{\text{race}} = \left\{ \begin{array}{llll} -0.184 & \text{Asian} & -0.093 & \text{Other} \\ -0.380 & \text{Black} & -0.037 & \text{Pacific Isl.} \\ -0.055 & \text{Hispanic} & 0 & \text{White} \\ 0.132 & \text{Am. Indian} & & \end{array} \right.$$

$$R_{\text{res}} = \left\{ \begin{array}{llll} -0.706 & \text{International} & 0 & \text{Resident} \\ 0.257 & \text{Nonresident} & & \end{array} \right.$$

These results are combined with the coefficients for the quantitative variables to calculate the overall risk score:

$$R = 0.048 \cdot \text{AGE} + R_{\text{college}} - 0.315 \cdot \text{HSGPA} - 0.234 \cdot \text{ORIENT} \\ - 0.709 \cdot \text{FIRSTGPA} + R_{\text{race}} + R_{\text{res}}$$

where ORIENT is 1 if the student enrolled in the Freshman Orientation Course, and 0 otherwise.

The risk score for our reference student (a 19-year-old white resident, enrolled in the College of Science, with a 3.4 high school GPA and 2.7 first-quarter GPA, who did not attend the Freshman Orientation Course) is

$$R_0 = 0.048(19) - 0.315(3.4) - 0.709(2.7) = -2.07$$

The estimated baseline survivor function for this individual, $S_0(t)$, has numerical values for each quarter-year increment in time; the values for 0, 1, 2, 3 and 4 years are 1, 0.824, 0.710, 0.645, and 0.595. We can then use Equation (3) to predict the retention probability for a student with any values of the independent variables, for any of these times.

REFERENCES

Anonymous (1997). Freshman-to-sophomore persistence rates, 1983–1997. *Postsecondary Education OPPORTUNITY*, Number 60, pp. 1–7. The Mortenson Research Seminar on Public Policy Analysis of Opportunity for Postsecondary Education, Iowa City, Iowa.

Astin, A. W. (1993). *What Matters in College: Four Critical Years Revisited*. San Francisco: Jossey-Bass.

Avakian, A. N., MacKinney, A. C., and Allen, G. R. (1982). Race and sex differences in student retention at an urban university. *College and University* 57: 160–165.

Bedford, M. H., and Durkee, P. E. (1989). Retention: Some more ideas. *NASPA Journal* 27: 168–171.

Boudreau, C. A., and Kromrey, J. D. (1994). A longitudinal study of the retention and academic performance of participants in freshmen orientation course. *Journal of College Student Development* 35: 444–449.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* 30: 89–99.

Cabrera, A. F., Nora, A., and Castañeda, M. B. 1993. College persistence: Structural equations modeling test of an integrated model of student retention. *Journal of Higher Education* 64: 123–139.

Collett, D. (1994). *Modelling Survival Data in Medical Research*. New York: Chapman and Hall.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B* 34: 187–220.

Dey, E. L., and Astin, A. W. (1993). Statistical alternatives for studying college student retention: A comparative analysis of logit, probit, and linear regression. *Research in Higher Education* 34: 569–581.

Dodd, J. M., Garcia, F. M., and Nelson, J. R. (1995). American Indian student retention. *NASPA Journal* 33(1): 72–78.

Dougherty, R. C., Bowen, C. W., Berger, T., Rees, W., Mellon, E. K., and Pulliam, E. (1995). Cooperative learning and enhanced communication: Effects on student performance, retention, and attitudes in general chemistry. *Journal of Chemical Education* 72(9): 793–797.

Glass, J. C., and Garrett, M. S. (1995). Student participation in a college orientation course, retention, and grade point average. *Community College Journal of Research and Practice* 19: 117–132.

Harrell, F. E., Jr., Lee, K. L., and Mark, D. B. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 15: 361–387.

Heiberger, R. M. (1993). Predicting next year's enrollment: Survival analysis of university student enrollment histories. *Proceedings of the American Statistical Association, Social Statistical Section*, pp. 143–148.

Hyman, R. E. (1995). Creating campus partnerships for student success. *College and University* 71: 2–8.

Johnson, R. (1996). The adult student: Motivation and retention. *The American Music Teacher* 46(2): 16–18, 60–61.

Kalbfleisch, J. D., and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kaplan, E. L., and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53: 457–481.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley.

Lee, E. T. (1992). *Statistical Methods for Survival Data Analysis.* New York: Wiley.

Moore, R., and Miller, I. (1996). How the use of multimedia affects student retention and learning. *Journal of College Science Teaching* 25(4): 289–293.

Naretto, J. A. (1995). Adult student retention: The influence of internal and external communities. *NASPA Journal* 32: 90–97.

Person, D. R., and Christensen, M. C. (1996). Understanding black student culture and black student retention. *NASPA Journal* 34(1): 47–56.

Peto, R., and Peto, J. (1972). Asymptotically efficient rank invariant procedures. *Journal of the Royal Statistical Society, Series A* 135: 185–207.

Reyes, N. (1997). Holding on to what they've got. *Black Issues in Higher Education* 13(26): 36–40.

Statistical Sciences Inc. (1993). *S-Plus Statistical Software, User's Manual,* Version 3.2. Seattle: Statistical Sciences Inc.

Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*. Chicago: The University of Chicago Press.

Volkwein, J. F., and Lorang, W. G. (1996). Characteristics of extenders: Full-time students who take light credit loads and graduate in more than four years. *Research in Higher Education* 37: 43–68.

Received on September 10, 1997.